Fluctuations for products of random matrices

Yi Sun

Columbia University

January 2020

I. Setting: Products of M random $N \times N$ matrices

II. Mathematical setup and results for fixed M

III. Main results: LLN and CLT with $N, M \to \infty$ jointly

IV. Method: Multivariate Bessel generating functions

V. Overview of statistical research areas

Products of random matrices

Consider M independent $N \times N$ random matrices $Y_1, \ldots, Y_M$ satisfying the rotational invariance in law

$$Y_k U \overset{d}{=} Y_k$$

for any unitary matrix U. Define the product

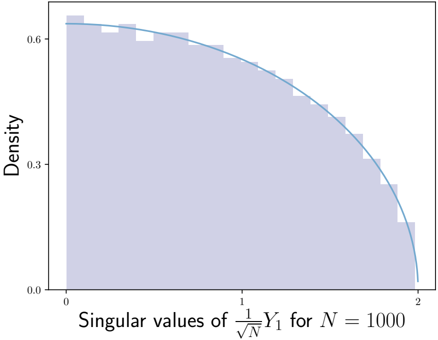
$$X = Y_1 Y_2 \cdots Y_M.$$

Question: How do singular values of X look as $N \to \infty$?

Example: “White” sample covariance ($M = 1$)

If $M = 1$, Y_1 with i.i.d. $\mathcal{N}_{\mathbb{C}}(0, 1/2)$ entries:

- $X = Y_1$ has i.i.d. multivariate Gaussian columns
- $XX^* =$ sample covariance for population covariance $\mathrm{Id}_{N\times N}$



- **Law of Large Numbers:** Quarter-circle law

Example: General sample covariance ($M = 2$)

If $M = 2$, Y_1 arbitrary and Y_2 with i.i.d. $\mathcal{N}_{\mathbb{C}}(0, 1/2)$ entries:

- $X = Y_1 Y_2$ has i.i.d. multivariate Gaussian columns
- $XX^* =$ sample covariance for population covariance $Y_1 Y_1^*$

Extensively studied in statistics and mathematics:

- Random matrix models: [Marchenko-Pastur '67, Jonsson '82, Bai-Silverstein '04]
- High-dimensional PCA: [Wachter '76, Johnstone '01, Baik-Silverstein '06, El Karoui '07, Paul '07, Nadler '08, Bai-Yao '08]
- Spectrum estimation: [El Karoui '08]
- Sphericity testing / signal detection [Ledoit-Wolf '02, Onatski-Moreira-Hallin '13, '14, Johnstone-Onatski '18]

Example: Separable sample covariance ($M = 3$)

The **separable covariance model** considers a data matrix

$$X = Y_1 \cdot Y_2 \cdot Y_3$$

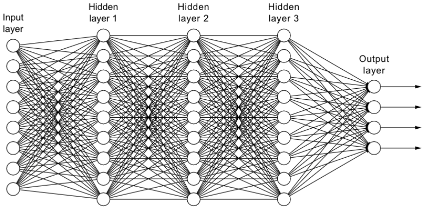
with Y_1, Y_3 arbitrary and Y_2 having i.i.d. $\mathcal{N}_{\mathbb{C}}(0, 1/2)$ entries.

- Rows and columns of X both have non-trivial correlation
- Entries of X are multivariate Gaussian with population covariance $Y_1 Y_1^* \otimes Y_3 Y_3^*$

Applications in several fields:

- Spatio-temporal statistics (rows = space, columns = time) [Storch-Zwiers ’99, Paul-Silverstein ’09]
- Matrix-variate statistics [Dawid ’81, Dutilleul ’99, Wang-West ’09, Allen-Tibshirani ’10, Hoff ’11, Leng-Tang ’12, Fosdick-Hoff ’12]
- Wireless communication [Tulino-Verdú ’04]
- Approximate factor models in economics [Onatski ’08]

Example: Deep neural network Jacobians (M large)



Feed-forward fully connected network with D layers of width N:

- **Weights** $W_1, \ldots, W_D \in \mathbb{R}^{N \times N}$ and **biases** $b_1, \ldots, b_D \in \mathbb{R}^N$.
- Given input $x = x^0 \in \mathbb{R}^N$, **activations** at layer k are:

  $$x^k = f(W_k \cdot x^{k-1} + b_k) \in \mathbb{R}^N$$

  for an **activation function** $f(x)$ applied element-wise.
- The **output** $F_\theta(x) \in \mathbb{R}^N$ is

  $$F_\theta(x) = x^D = f(b_D + W_D \cdot f(b_{D-1} + W_{D-1} \cdot f(\cdots)))$$

  for **parameters** $\theta = (W_1, \ldots, W_D, b_1, \ldots, b_D)$.

Example: Deep neural network Jacobians (M large)

At initialization: W_i has i.i.d. real Gaussian entries, $b_i = 0$.

$$F_\theta(x) = f(b_D + W_D \cdot f(b_{D-1} + W_{D-1} \cdot f(\cdots)))$$

Jacobian of output with respect to input is:

$$J(x) = Df(x^D) \cdot W_D \cdot Df(x^{D-1}) \cdots W_1,$$

where for $x \in \mathbb{R}^N$, $Df(x)$ is the diagonal matrix

$$Df(x) = \begin{bmatrix} f'(x_1) & & & \\ & f'(x_2) & & \\ & & \ddots & \\ & & & f'(x_N) \end{bmatrix}.$$

Example: Deep neural network Jacobians (M large)

Jacobian at initialization – with $U_1, \dots, U_D$ Haar unitary:

$$\begin{aligned} J(x) &= Df(x^D) \cdot W_D \cdot Df(x^{D-1}) \cdots W_1 \\ &\stackrel{d}{=} (Df(x^D)U_D) \cdot W_D \cdot (Df(x^{D-1})U_{D-1}) \cdot \cdots \cdot W_1 \end{aligned}$$

fits into our framework with $M = 2D$ and

$$Y_1 = Df(x^D)U_D, \qquad Y_2 = W_D, \qquad \dots$$

Typical values: depth $D = O(100)$ and width $N = O(10^5)$

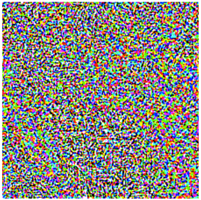
Conclusion: Asymptotic study requires $N, M \to \infty$ jointly

Example: Deep neural network Jacobians (M large)

$\boldsymbol{x}$

"panda"
57.7% confidence

$+ .007 \times$



$\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

"nematode"
8.2% confidence

$=$

$\boldsymbol{x} +$
$\epsilon \text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

"gibbon"
99.3 % confidence

[Goodfellow-Shlens-Szegedy '14]

- Adversarial examples on image classification models
- Slightly perturb input along gradient of $\mathbf{1}_{\text{gibbon}}$

Example: Deep neural network Jacobians (M large)

In training with loss $\ell(y, y')$ at data point (x_i, y_i), take step

$$\theta' = \theta - \alpha \cdot \nabla_\theta \ell\Big(y_i, F_\theta(x_i)\Big).$$

Expressed with $J_\theta F_\theta(x_i)$, which also has product structure.

For successful training, must make sure gradients are not:

- too large (**gradient explosion**), or
- too small (**gradient vanishing**).

[Saxe-McClelland-Ganguli '14] [Pennington-Schoenholz-Ganguli '17]
[Chen-Pennington-Schoenholz '18] [Hanin '18] [Zhang-Dauphin-Ma '19]

I. Setting: Products of M random $N \times N$ matrices

II. Mathematical setup and results for fixed M

III. Main results: LLN and CLT with $N, M \to \infty$ jointly

IV. Method: Multivariate Bessel generating functions

V. Overview of statistical research areas

Random Matrix Theory: global regime

Recall $N \times N$ matrices $Y_1, \ldots, Y_M$:

$$X_{N,M} = Y_1 Y_2 \cdots Y_M.$$

Consider the empirical spectral measure of $X_{N,M}$

$$\nu_{N,M} := \frac{1}{N} \sum_{i=1}^{N} \delta_{\mu_i}$$

with $\mu_1 \geq \cdots \geq \mu_N$ singular values of $X_{N,M}$. As $N \to \infty$, want:

- **Law of Large Numbers:** Deterministic limit for $\nu_{N,M}$
- **Central Limit Theorem:** Gaussian fluctuations of $\nu_{N,M}$ about its expectation (after rescaling).

Global regime because they rely on singular values as a whole.

LLN for products with M fixed

Define the S-transform from free probability

$$S_\nu(z) = \frac{z+1}{z} M_\nu^{-1}(z) \qquad \text{with} \qquad M_\nu(z) = \int \frac{xz}{1-xz} d\nu(x).$$

Let A, B be right-invariant matrices whose singular values $\{a_i\}, \{b_i\}$ have empirical measures with deterministic limits

$$\nu_A := \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \delta_{a_i} \qquad \nu_B := \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \delta_{b_i}.$$

Theorem (Voiculescu, '80s)

The empirical singular value measure of $X = AB$ has deterministic limit ν_X satisfying

$$S_{\nu_X}(z) = S_{\nu_A}(z) S_{\nu_B}(z).$$

CLT for products with M fixed

Define the **height function** of X:

$$\mathcal{H}_N(t) = \#\{\mu_i \le t\} = N\,\nu_X\Big((-\infty, t]\Big).$$

Note that $N \to \infty$ limit of $\frac{1}{N}\mathbb{E}[\mathcal{H}_N(t)]$ is determined by LLN.

Theorem (Gorin-**S.** '18)

As $N \to \infty$, the limit of fluctuations of the height function

$$\xi(x) := \lim_{N\to\infty} \Big(\mathcal{H}_N(x) - \mathbb{E}[\mathcal{H}_N(x)]\Big)$$

is an **explicit** Gaussian log-correlated field $\xi(x)$, meaning

$$\mathbb{E}[\xi(x)\xi(y)] \approx -\frac{1}{2\pi^2}\log|x-y| \qquad \text{for } x \approx y.$$

Note: Fluctuations $\mathcal{H}_N(x) - \mathbb{E}[\mathcal{H}_N(x)]$ are random functions on $\mathbb{R}$ converging to the random **distribution** $\xi(x)$.

CLT for products with M fixed

Theorem (Gorin-**S.** '18)

As $N \to \infty$, the limit of fluctuations of the height function

$$\xi(x) := \lim_{N\to\infty} \Big(H_N(x) - \mathbb{E}[H_N(x)] \Big)$$

is an **explicit** Gaussian log-correlated field $\xi(x)$, meaning

$$\mathbb{E}[\xi(x)\xi(y)] \approx -\frac{1}{2\pi^2} \log|x-y| \qquad \text{for } x \approx y.$$

Additive analogue:

- 2nd order freeness: [Collins-Mingo-Śniady-Speicher '04]
- Stieltjes transform: [Pastur-Vasilchuk '07]

Multiplicative case:

- Sample covariance: [Jonsson '82, Bai-Silverstein '04]
- Separable covariance: [Bai-Li-Pan '16]
- Gaussianity: [Guionnet-Novak '15] [Arizmendi-Mingo '18]
- Explicit covariance + log-correlation: [Gorin-**S.** '18]

Fixed matrix size (N) and growing number (M)

Recall that

$$X_{N,M} = Y_1 Y_2 \cdots Y_M.$$

Consider N fixed as $M \to \infty$:

- Singular values grow exponentially in M
- **Lyapunov exponents** have deterministic limits

$$\lambda_i := \frac{1}{M} \log \mu_i$$

[Furstenberg-Kesten '60]

- Appears in dynamical systems from population ecology

Growing matrix size (N) and number (M) together

What if $N, M \to \infty$ together?

- Should consider Lyapunov exponents
- Taking $N \to \infty$ and then $M \to \infty$: multiplicative free central limit theorem
- Taking $M \to \infty$ and then $N \to \infty$: similarity to fixed N

Our results: LLN and CLT for all joint limits $N, M \to \infty$.

I. Setting: Products of M random $N \times N$ matrices

II. Mathematical setup and results for fixed M

III. Main results: LLN and CLT with $N, M \to \infty$ jointly

IV. Method: Multivariate Bessel generating functions

V. Overview of statistical research areas

$M \to \infty$, multiplicative case

Define i.i.d. $N \times N$ random matrices

$$Y_k := AU_k$$

with i.i.d Haar unitary matrices U_k and deterministic diagonal

$$A = \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_N \end{pmatrix}$$

with $a_j > 0$ so $\nu := \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \delta_{a_i}$ converges. Define

$$X_{N,M} := Y_1 Y_2 \cdots Y_M.$$

For $M \to \infty$, study Lyapunov exponents $\{\lambda_k\}$ defined by

$$\lambda_k := \frac{1}{M} \log \left(k^{\text{th}} \text{ singular value of } X_{N,M} \right)$$

$M \to \infty$ LLN, multiplicative case

Define:

$$S(z) := \frac{z+1}{z} M_\nu^{-1}(z) \qquad M_\nu(z) := \int \frac{xz}{1-xz} d\nu(x).$$

Theorem (Newman '86, Kargin '08, Tucci '10, Gorin-**S.** '18)

As $N, M \to \infty$ jointly, the empirical measure of Lyapunov exponents converges to the explicit measure

$$\lim_{N,M\to\infty} \frac{1}{N} \sum_{k=1}^{N} \delta_{\lambda_k} = \frac{-e^{-x}}{S'(S^{-1}(e^{-x}))} \mathbf{1}_{[-\log S(-1), -\log S(0)]} dx.$$

- CDF of limiting measure is $S^{-1}(e^{-x}) + 1$
- Limiting measure in LLN recovers original measure ν
- Holds for **any** relative rate of growth $N, M \to \infty$

$M \to \infty$ CLT, multiplicative case

Lyapunov exponents

$$\lambda_k = \frac{1}{M} \log \left(k^{\text{th}} \text{ singular value of } X_{N,M} \right)$$

and height function $H_{N,M}(t) = \#\{\lambda_k \leq t\}$.

Theorem (Gorin-**S.** '18)

As $N, M \to \infty$ jointly, rescaled fluctuations converge

$$M^{1/2} \Big(H_{N,M}(x) - \mathbb{E}[H_{N,M}(x)] \Big) \to \xi(x)$$

to explicit Gaussian field $\xi(x)$ with **white noise** component, i.e.

$$\mathbb{E}[\xi(x)\xi(y)] \approx \delta(x-y) \qquad \text{for } x \approx y.$$

Fluctuations go from **log-correlated** for M fixed to **white noise** for $M \to \infty$

$M \to \infty$, comparison to additive case

Define $X^{\text{add}}_{N,M} := \sum_{k=1}^{M} U_k A U_k^*$. As $N, M \to \infty$, have

$$\frac{1}{M} X^{\text{add}}_{N,M} \approx \frac{1}{N} \Big(\sum_{k=1}^{N} a_k \Big) \cdot \mathrm{Id}$$

$$\sqrt{\frac{N^2 - 1}{NM}} \Big(X^{\text{add}}_{N,M} - \mathbb{E}[X^{\text{add}}_{N,M}] \Big) \approx (\text{constant}) \cdot \mathrm{GUE}_{N,\mathrm{Tr}=0},$$

where

$$\mathrm{GUE}_{N,\mathrm{Tr}=0} = \begin{pmatrix} \text{traceless Hermitian matrix with i.i.d.} \\ \text{complex Gaussian entries} \end{pmatrix}$$

Theorem (Johansson '98)

Fluctuations of height function of GUE_N converge as $N \to \infty$ to explicit log-correlated Gaussian field.

Fluctuations **stay log-correlated** between M fixed and $M \to \infty$.

Why does white noise appear?

Consider additive decomposition

$$X^{\text{add}}_{N,M} = \mathbb{E}[X^{\text{add}}_{N,M}] + \left(X^{\text{add}}_{N,M} - \mathbb{E}[X^{\text{add}}_{N,M}] \right).$$

Expectation is multiple of identity: $\frac{1}{M}\mathbb{E}[X^{\text{add}}_{N,M}] \approx (\text{const}) \cdot \text{Id}$

$$(k^{\text{th}} \text{ eigenval. of } X^{add}_{N,M}) \approx \underbrace{\boxed{(\text{const}_1) \cdot M}}_{\mathbb{E}[X^{\text{add}}_{N,M}]} + \underbrace{\boxed{(\text{const}_2) \cdot \sqrt{M} \cdot \gamma_k}}_{X^{\text{add}}_{N,M} - \mathbb{E}[X^{\text{add}}_{N,M}]}$$

for $\gamma_k \stackrel{d}{=} (k^{\text{th}} \text{ eigenval. of } \text{GUE}_{N,\text{Tr}=0})$.

Height function fluctuations come **only** from matrix fluctuations.

Why does white noise appear?

Consider multiplicative decomposition

$$\log X_{N,M}^{\text{mult}}(X_{N,M}^{\text{mult}})^* = \mathbb{E}[\log X_{N,M}^{\text{mult}}(X_{N,M}^{\text{mult}})^*] + \Big(\log X_{N,M}^{\text{mult}}(X_{N,M}^{\text{mult}})^* - \mathbb{E}[\log X_{N,M}^{\text{mult}}(X_{N,M}^{\text{mult}})^*] \Big).$$

Expectation $\mathbb{E}[\log X_{N,M}^{\text{mult}}(X_{N,M}^{\text{mult}})^*]$ has non-trivial spectrum

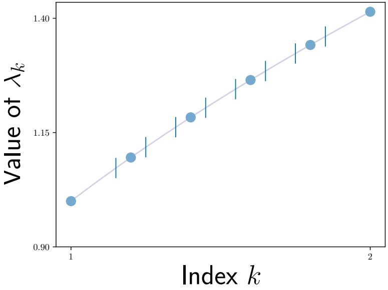
- k^{th} eigenvalue of $\log X_{N,M}^{\text{mult}}(X_{N,M}^{\text{mult}})^*$ not determined from spectra of its expectation and fluctuations
- fluctuations $\Big(\log X_{N,M}^{\text{mult}}(X_{N,M}^{\text{mult}})^* - \mathbb{E}[\log X_{N,M}^{\text{mult}}(X_{N,M}^{\text{mult}})^*] \Big)$ are distributed along the spectrum of $\mathbb{E}[\log X_{N,M}^{\text{mult}}(X_{N,M}^{\text{mult}})^*]$

Why does white noise appear?

For **fixed** N, $M \to \infty$: limit of Lyapunov exponents is

$$(\lambda_1, \dots, \lambda_N) \approx \mathbb{E}\Big[(\lambda_1, \dots, \lambda_N)\Big] + \frac{1}{\sqrt{M}} \mathcal{N}(0, \Sigma)$$

[Akemann-Burda-Kieburg '14], [Forrester '15], [Reddy '16], [Kieburg-Kosters '17]



Empirical measure $\frac{1}{N}\sum_{k=1}^{N} \delta_{\lambda_k}$ has non-trivial limit, so...

- $\frac{1}{\sqrt{M}}$ fluctuation in height function (giving $\mathbf{M}^{1/2}$ factor)
- white noise due to **independence** between coordinates

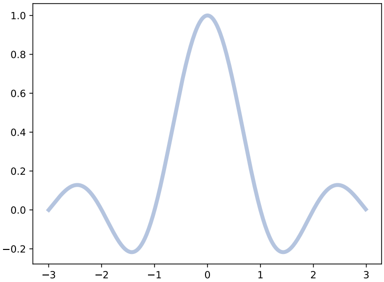
Does the relative rate of growth of N, M matter?

Global LLN and CLT: Our result holds for **any** relative rate

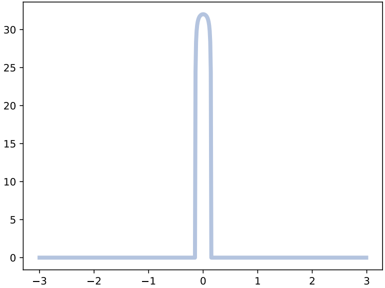
For **local statistics** (correlations in a $O(N^{-1})$ neighborhood), as $N, M \to \infty$ jointly for complex Ginibre:

- for $N \gg M$: sine kernel correlations
- for $N \ll M$; transition to delta function statistics

[Akemann-Burda-Kieburg '18] [Liu-Wang-Wang '18]



Sine kernel



Approximate Delta function kernel

I. Setting: Products of M random $N \times N$ matrices

II. Mathematical setup and results for fixed M

III. Main results: LLN and CLT with $N, M \to \infty$ jointly

IV. Method: Multivariate Bessel generating functions

V. Overview of statistical research areas

Overview of method

Combine **integrable probability** and **moment method**:

1. Define Bessel generating function $\phi_X^{\text{mult}}(s)$ of measure
2. Find differential operators D_k in $s_1, \ldots, s_N$ giving moments

$$\mathbb{E}\left[p_{k_1}(x) \cdots p_{k_r}(x)\right] = D_{k_1} \cdots D_{k_r} \phi_X^{\text{mult}}(\rho),$$

   where $p_k(x) = x_1^k + \cdots + x_N^k$.

3. Obtain **LLN** and **CLT** with integral covariance kernel in terms of asymptotics for derivatives of $\log \phi_X^{\text{mult}}(s)$.
4. Obtain **exact asymptotics** for derivatives of $\log \phi_X^{\text{mult}}(s)$.

Analogous to moment generating functions for 1-D measures.

1-D LLN and CLT from moment generating function

Let X_N be a sequence of real-valued random variables.

1. Moment generating function is

$$\phi_{X_N}(s) = \mathbb{E}\left[e^{sX_N}\right].$$

2. Moments of X_N are obtained by derivatives

$$\mathbb{E}[X_N^k] = \phi_{X_N}^{(k)}(0).$$

3. Cumulants of X_N are obtained by log-derivatives so that

$$\text{LLN} \iff \kappa_2(X_N) = \frac{d^2}{ds^2}[\log \phi_{X_N}(s)]\Big|_{s=0} = o(1)$$

$$\text{CLT} \iff \kappa_k(X_N) = \frac{d^k}{ds^k}[\log \phi_{X_N}(s)]\Big|_{s=0} = o(1) \text{ for } k \geq 3.$$

4. Get LLN and CLT from log-derivatives for specific $\phi_{X_N}(s)$.

Step 1: Multivariate Bessel generating functions

Multivariate Bessel function is defined by

$$\mathcal{B}(s,x) := \frac{\det(e^{s_i x_j})_{i,j=1}^{N}}{\prod_{i<j}(s_i - s_j)\prod_{i<j}(x_i - x_j)}(N-1)!\cdots 1!.$$

For measure $\nu(x)$ on N-tuples $(x_1 \geq \cdots \geq x_N)$ and $\rho = (N-1,\ldots,0)$, the **Bessel generating function** is

$$\phi_\nu(s) := \mathbb{E}_\nu\left[\frac{\mathcal{B}(s;x)}{\mathcal{B}(\rho,x)}\right].$$

- normalized by $\phi_\nu(\rho) = 1$
- $\phi_\nu(s)$ is analogue of Schur generating function for discrete measures (ρ replaced by 0^N in [Bufetov-Gorin '13-'17])

Step 1: Bessel generating functions and products

Recall for $\rho = (N-1, \dots, 0)$ and a measure ν on $(x_1, \dots, x_N)$:

$$\phi_\nu(s) := \mathbb{E}_\nu \left[\frac{\mathcal{B}(s; x)}{\mathcal{B}(\rho, x)} \right]$$

Let ν be the measure on **log-singular values** (scaled by 2) of a random matrix X. Define

$$\phi_X^{\text{mult}}(s) := \phi_\nu(s).$$

Proposition

For independent right-unitarily invariant matrices X, Y:

$$\phi_{XY}^{\text{mult}}(s) = \phi_X^{\text{mult}}(s) \cdot \phi_Y^{\text{mult}}(s).$$

Step 2: Moments from Bessel generating functions

Consider differential operators

$$D_k := \prod_{i<j}(s_i - s_j)^{-1} \circ \sum_{i=1}^{N} \partial_i^k \circ \prod_{i<j}(s_i - s_j).$$

Proposition (Gorin-S. '18)

If $\phi_\mu(s)$ is Bessel generating function for measure ν on $(x_1 \geq \cdots \geq x_N)$, moments of ν are

$$\mathbb{E}[p_{k_1}(x) \cdots p_{k_r}(x)] = D_{k_1} \cdots D_{k_r} \phi_\nu(\rho)$$

for $p_k(x) = x_1^k + \cdots + x_N^k$.

Proof: Analytic continuation from $D_k \phi_\nu(s) = p_k(x) \phi_\nu(s)$ via

$$\frac{\mathcal{B}(s,x)}{\mathcal{B}(\rho,x)} = \frac{\det(e^{s_i x_j})_{i,j=1}^N}{\prod_{i<j}(s_i - s_j)} \frac{\prod_{i<j}(\rho_i - \rho_j)}{\det(e^{\rho_i x_j})_{i,j=1}^N}.$$

Step 3: LLN from Bessel generating functions

Theorem (Gorin-S. '18)

If $\phi_X^{\text{mult}}(s)$ for prob. measure $d\mu(x)$ on $(x_1 \geq \cdots \geq x_N)$ satisfies

$$\frac{1}{N}\partial_{r_i}[\log \phi_X^{\text{mult}}(rN)]\Big|_{r_k=\rho_k/N, k\neq j} \to \Psi'(r_i),$$

have convergence in probability for fixed M:

$$\lim_{N\to\infty} \frac{1}{N} p_k(x) = \frac{1}{k+1} \oint \Big(\log(u/(u-1)) + \Psi'(u)\Big)^{k+1} \frac{du}{2\pi \mathbf{i}}$$

and for $\psi_X(s) = \phi_X^{\text{mult}}(s)^M$ with $M \to \infty$:

$$\lim_{N\to\infty} \frac{1}{N} p_k(x) = \oint \log(u/(u-1))\Psi'(u)^k \frac{du}{2\pi \mathbf{i}}.$$

Step 3: CLT from Bessel generating functions

Theorem (Gorin-**S.** '18)

If $\phi_X^{\text{mult}}(s)$ for prob. measure $d\mu(x)$ on $(x_1 \geq \cdots \geq x_N)$ satisfies

$$\frac{1}{N}\partial_{r_i}[\log \phi_X^{\text{mult}}(rN)]\Big|_{r_k=\rho_k/N, k\neq i} \to \Psi'(r_i)$$

$$\partial_{r_i}\partial_{r_j}[\log \phi_X^{\text{mult}}(rN)]\Big|_{r_k=\rho_k/N, k\neq i,j} \to F^{(1,1)}(r_i, r_j)$$

have Gaussian limit for $\{p_k(x) - \mathbb{E}[p_k(x)]\}_{k\in\mathbb{N}}$ with $\mathrm{Cov}(p_k, p_l)$:

$$\oint\oint \Big(\log(u/(u-1)) + \Psi'(u)\Big)^k \Big(\log(w/(w-1)) + \Psi'(w)\Big)^l \Big(\frac{1}{(u-w)^2} + F^{(1,1)}(u,w)\Big)\frac{du}{2\pi\mathbf{i}}\frac{dw}{2\pi\mathbf{i}}.$$

For $M \to \infty$: Similar theorem with $\psi_X^{\text{mult}}(s) = \phi_X^{\text{mult}}(s)^M$

Step 4: Asymptotics of Bessel generating functions

For LLN and CLT, need to find Ψ and F so that

$$\frac{1}{N}\partial_{r_i}[\log \phi_X^{\mathrm{mult}}(rN)]\Big|_{r_k=\rho_k/N, k\neq i} \to \Psi'(r_i)$$

$$\partial_{r_i}\partial_{r_j}[\log \phi_X^{\mathrm{mult}}(rN)]\Big|_{r_k=\rho_k/N, k\neq i,j} \to F^{(1,1)}(r_i, r_j).$$

For $X = AU$ with A diagonal and U Haar unitary

$$\phi_X^{\mathrm{mult}}(s) = \frac{\mathcal{B}(s, a)}{\mathcal{B}(\rho, a)}.$$

LLN $\iff$ asymptotics for s differing from ρ in 1 coordinate:

$$s = \Big(r_1 N, N-1, \ldots, \widehat{b_1 N}, \ldots, 0\Big).$$

Step 4: LLN asymptotics

Theorem (Gorin-**S.** '18)

If the empirical measure of diagonal entries of A limits to ν

$$\lim_{N\to\infty} \frac{1}{N}\partial_{r_i}[\log \phi_X^{\mathrm{mult}}(rN)]\Big|_{r_k=\rho_k/N, k\neq i} = -\log S_\nu(r_i - 1).$$

Lemma (Kieburg-Kosters '15, Gorin-**S.** '18)

If X is right unitarily invariant with QR-decomposition $X = QR$

$$\phi_X^{\mathrm{mult}}(s) = \mathbb{E}\Big[\prod_{k=1}^{N} R_{kk}^{2(s_k-\rho_k)}\Big].$$

Corollary (Gorin-**S.** '18)

Let X be right unitarily invariant with singular value measure converging to ν. For $t \in [0,1]$, we have

$$-\log S_\nu(t-1) = \lim_{N\to\infty} \mathbb{E}[2\log R_{\lfloor tN\rfloor,\lfloor tN\rfloor}].$$

Summary

1. Global fluctuations of sums and products of M independent $N \times N$ unitarily-invariant random matrices converge to explicit Gaussian fields as $N \to \infty$.
   - sums: log-correlated fields for M fixed and $M \to \infty$
   - products: **log-correlated** for M fixed to **white noise** for $M \to \infty$
2. Uses differential operators acting on **multivariate Bessel generating functions** of empirical measures of Lyapunov exponents.

V. Gorin, Y. S., Gaussian fluctuations for products of random matrices, `arXiv:1812.06532`.

I. Setting: Products of M random $N \times N$ matrices

II. Mathematical setup and results for fixed M

III. Main results: LLN and CLT with $N, M \to \infty$ jointly

IV. Method: Multivariate Bessel generating functions

V. Overview of statistical research areas

Covariance estimators in mixed-effects models

Mixed effects linear model

$$Y = U_1\alpha_1 + \cdots + U_k\alpha_k + \varepsilon \in \mathbb{R}^{n\times p}$$

- $\alpha_r \in \mathbb{R}^{n_r\times p}$ with Gaussian rows realizing random effects
- $U_r \in \mathbb{R}^{n\times n_r}$ known deterministic incidence matrices
- error term $\varepsilon \in \mathbb{R}^{n\times p}$

Commonly used in genome-wide association studies for polygenic traits, where effects of SNPs are modeled as random.

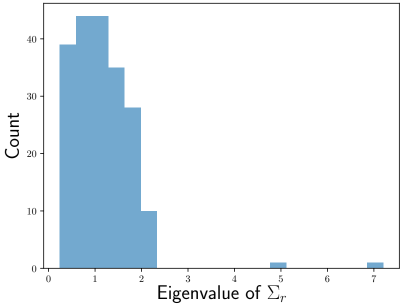
Classical setting ($n, n_r \to \infty$, $p < \infty$): MANOVA estimator

$$\widehat{\Sigma}_r = Y^\mathsf{T} B_r Y$$

for estimation matrix B_r recovers α_r's population covariance Σ_r.

Covariance estimators in mixed-effects models

Model population covariance Σ_r as **bulk** plus outlier **spikes.**



Goal: In high-dimensional setting ($n, n_r, p \to \infty$ jointly):

- Find **RMT correction** to spectrum of MANOVA estimator

$$\widehat{\Sigma}_r = Y^\mathsf{T} B_r Y.$$

- Use correction to **infer** Σ_r from data.

Covariance estimators in mixed-effects models

High-dimensional setting ($n, n_r, p \to \infty$ jointly):

Isotropic bulk $\Sigma_{r,\text{bulk}} = \sigma_r^2 \cdot \text{Id}$: [Fan-Johnstone-**S.** '18]

- LLN and CLT for outlier eigenvalues $\widehat{\theta}_i$ of $\widehat{\Sigma}_r$
- Asymptotic angle between sample and population eigenvectors $\widehat{v}_i$ and v_i.
- Asymptotically consistent inference for population outlier eigenvalues θ_i, debiasing of $\widehat{v}_i$
- Stieltjes transform methods

General bulk $\Sigma_{r,\text{bulk}}$: [Fan-**S.**-Wang '19]

- LLN for outlier eigenvalues $\widehat{\theta}_i$ of $\widehat{\Sigma}_r$
- Asymptotic angle between eigenvectors $\widehat{v}_i$ and v_i.
- Free probability methods

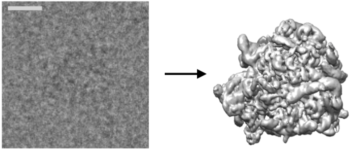
Learning group orbits

Let $G \subset O(d)$ be discrete subgroup of orthogonal group. For unknown $\theta_* \in \mathbb{R}^d$, consider data drawn according to

$$x_i \sim g_i \cdot \theta_* + \varepsilon_i$$

with g_i Haar uniform on G and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2 \cdot \mathrm{Id})$.

- Multi-reference alignment: $G = \mathbb{Z}/d\mathbb{Z}$ with cyclic rotation
- Cryo-EM: $G = SO(3)$ with rotation (ignoring projection)



$10^{4\text{-}7}$ projection images

3D electron density

[Zhong-Bepler-Davis-Berger '19]

Question: How can we estimate θ_* from observations of x_i?

Learning group orbits

Recall: Observations from

$$x_i \sim g_i \cdot \theta_* + \varepsilon_i$$

with g_i Haar uniform on G and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2 \cdot \mathrm{Id})$.

Moment method: [Bandeira-Blum-Smith-Kileel-Perry-Weed-Wein '17]

- Estimate G-invariant moments for large noise $\sigma \to \infty$
- Sample complexity in terms of σ and **ring of invariants**
- Information-theoretic result (large polynomial system)

Question: What about maximum likelihood (MLE) and expectation-maximization (EM)?

- MLE is efficient (moment method is only $\sqrt{n}$-consistent)
- EM is most widely used algorithm in practice

Learning group orbits

MLE: approximate negative log-likelihood with population risk:

$$R(\theta) = -\mathbb{E}_\varepsilon \left[\log \mathbb{E}_{g,g'} \left[\exp\left(-\frac{||g\cdot\theta_* + \varepsilon - g'\cdot\theta||^2}{2\sigma^2} \right) \right] \right]$$

Landscape is non-convex: could have **spurious local minima.**

Small noise $\sigma \to 0$: [Fan-**S.**-Wang-Wu '20+]

- Loss landscape is good: all local minima are global minima

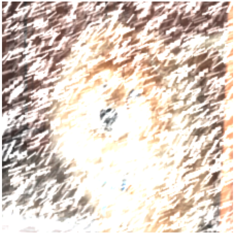
Large noise $\sigma \to \infty$: [Fan-**S.**-Wang-Wu '20+]

- Good loss landscape for $\mathbb{Z}/d\mathbb{Z}$ on $\mathbb{R}^2$
- Examples of spurious local minima for MRA on $\mathbb{R}^6$ and $\mathbb{R}^7$
- Eigenvalues of Fisher information have very different magnitudes according to structure of ring of invariants
- Uses detailed analysis of series expansion of $R(\theta)$ in σ^{-1}

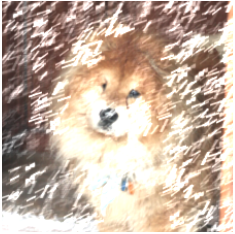
Robustness against unforeseen adversaries

Neural networks classify clean images well..

Original



Initialization

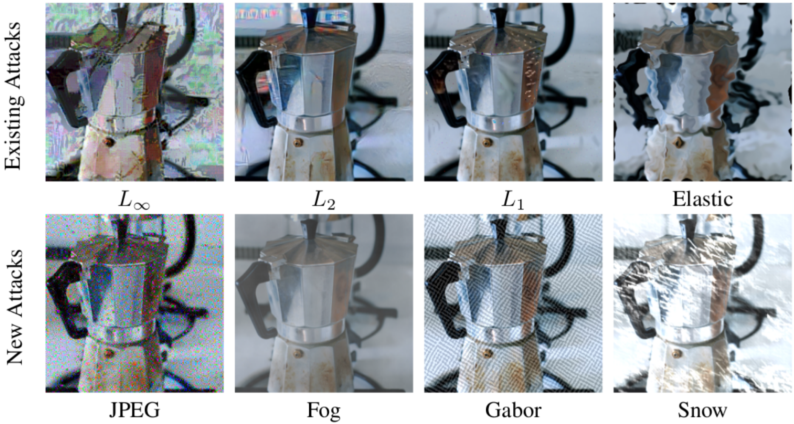


Optimized

[Kang*-**S.***-Hendrycks-Brown-Steinhardt ’19]

.. but are vulnerable to adversarially crafted attacks.

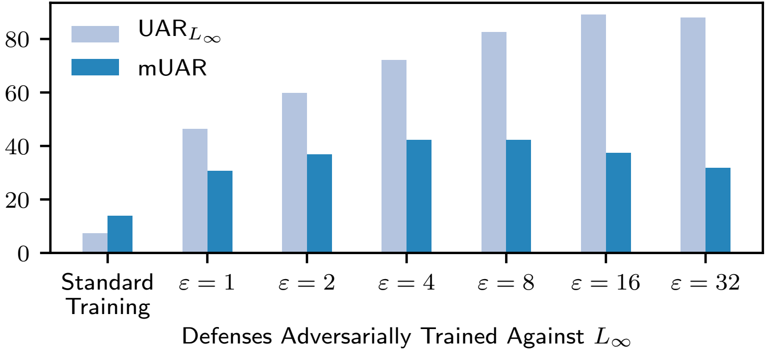
Robustness against unforeseen adversaries



[Kang*-**S.***-Hendrycks-Brown-Steinhardt '19]

- Existing attacks optimize specific unrealistic constraints
- We introduce **four** new attacks for evaluation

Robustness against unforeseen adversaries



[Kang*-**S.***-Hendrycks-Brown-Steinhardt '19]

- New metric UAR for performance of defense vs. attack
- UAR shows existing defenses generalize poorly to **unforeseen** attacks
- Used 10000+ GPU-hour experiments done with OpenAI

References

Fluctuations for products of random matrices

- V. Gorin and **Y. S.**, Gaussian fluctuations for products of random matrices, `arXiv:1812.06532`.

Covariance estimators in mixed-effects models

- Z. Fan, I. Johnstone, and **Y. S.**, Spiked covariances and principal components analysis in high-dimensional random effects models, `arXiv:1806.09529`.
- Z. Fan, **Y. S.**, and Z. Wang. Principal components in linear mixed models with general bulk, `arXiv:1903.09592`.

Learning group orbits

- Z. Fan, **Y. S.**, T. Wang, and Y. Wu. Likelihood landscape for discrete group mixture models, in preparation.

Robustness against unforeseen adversaries

- D. Kang*, **Y. S.*** , D. Hendrycks, T. Brown, and J. Steinhardt. Testing robustness against unforeseen adversaries, `arXiv:1908.08016`.

Funding: NSF DMS-1701654, Simons Foundation, OpenPhil