Deep Random Matrix Theory

Dang-Zheng Liu

School of Mathematical Sciences
University of Science and Technology of China

17th Workshop on Markov Processes and Related Topics
November 25-27, 2022

Table of Contents

1 Introduction
- Origins
- Lyapunov exponents

2 Local statistics

3 Double scaling limits

Example 0.1 Feedforward neural network,I

Le Cun, Bengio & Hinton [Deep learning, Nature 2015]:

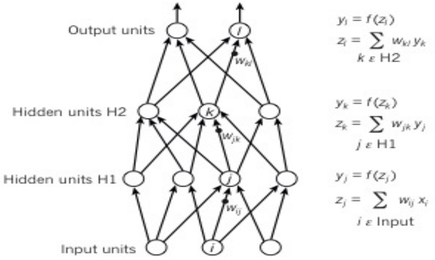


Figure: Multilayer architectures

The hierarchy of concepts allows the computer to learn complicated concepts by building them out of simpler ones. If we draw a graph showing how these concepts are built on top of each other, the graph is deep, with many layers. [Deep Learning by Goodfellow, Bengio &Courville]

Example 0.1 Feedforward neural network,II

Deep Neural Networks $\dashrightarrow$ Deep Random Matrix Theory (Deep RMT)

$$X_0 \in \mathbb{R}^N, \quad X_k = f(W_k X_{k-1}), k = 1, \ldots, L$$

where $f : \mathbb{R} \to \mathbb{R}$ is the activation function and W_k are (random) weight matrices of size N.
The input-output Jacobian reduces to a product of many (random) matrices

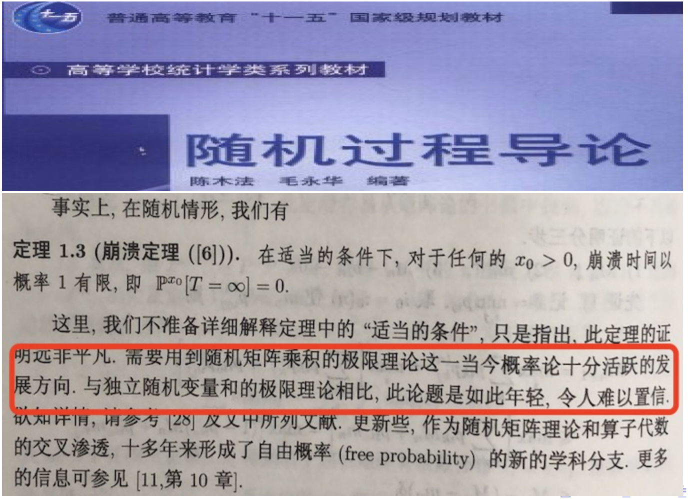
$$DW_L \cdots DW_2 DW_1$$

where D is a diagonal matrix. **Both width N and depth L are LARGE!**

Example 0.2 Stochastic Models of Economic Optimization

Input-Output Method (Chen, Mao, book page 8):

$$X_0 = X_L W_L \cdots W_2 W_1, \quad X_0, X_L \in \mathbb{R}^N$$



事实上, 在随机情形, 我们有

定理 1.3 (崩溃定理 ([6])). 在适当的条件下, 对于任何的 $x_0 > 0$, 崩溃时间以概率 1 有限, 即 $\mathbb{P}^{x_0}[T = \infty] = 0$.

这里, 我们不准备详细解释定理中的 "适当的条件", 只是指出, 此定理的证明远非平凡. 需要用到随机矩阵乘积的极限理论这一当今概率论十分活跃的发展方向. 与独立随机变量和的极限理论相比, 此论题是如此年轻, 令人难以置信. 欲知详情, 请参考 [28] 及文中所列文献. 更新些, 作为随机矩阵理论和算子代数的交叉渗透, 十多年来形成了自由概率 (free probability) 的新的学科分支. 更多的信息可参见 [11, 第 10 章].

Example 0.2 Stochastic Models of Economic Optimization

Chen, Mu-Fa, *Eigenvalues, Inequalities, and Ergodic Theory*, chapter 10

To conclude this chapter, let us make some remarks about the theory of random matrices. The theory is a traditional and important branch of mathematics and has a very wide range of applications including statistics,

physics, number theory, and even the Riemann hypothesis. Refer to M. Mehta (1991), V.L. Girko (1990), J.B. Conrey (2003), and references within.

Mainly, there are two topics in the study of eigenvalues. The first one is the estimation of the first few eigenvalues, as dealt with in this book. The second one, omitted in the book, is the asymptotic behavior of the eigenvalues. In the context of random matrices, concerning the second topic there is the famous beautiful Wigner's semicircle law (1955). For its modern generalization to operator algebras, called free probability, see D. Voiculescu, K. Dykema, and A. Nica (1992) and E. Haagerup (2002), for instance.

Four primary sources, Mathematics & Statistics

- A. Hurwitz, Über die Erzeugung der invarianten durch integration, Nachr. Ges. Wiss. Göttingen,1897. See Diaconis and Forrester, Hurwitz and the origins of random matrix theory in mathematics, Random Matrices: Theory and Applications, 2017
  Invariant measure on matrix groups $SO(N)$ and $U(N)$

Four primary sources, Mathematics & Statistics

- A. Hurwitz, Über die Erzeugung der invarianten durch integration, Nachr. Ges. Wiss. Göttingen,1897. See Diaconis and Forrester, Hurwitz and the origins of random matrix theory in mathematics, Random Matrices: Theory and Applications, 2017
  Invariant measure on matrix groups $SO(N)$ and $U(N)$
- J. Wishart, The generalized product moment distribution in samples from a normal multivariate population, Biometrika, 1928
  A sample covariance matrix with N samples

$$S_N = \frac{1}{N}\sum_{k=1}^{N} X_k X_k^t = \frac{1}{N} X X^t$$

where X is a $p \times N$ noisy array.

Four primary sources, Physics

E.P. Wigner, Characteristic vectors of bordered matrices with infinite dimensions, Ann. Math., 1955.

Truncated Hamiltonians of heavy-nuclei atoms into random matrices of finite size N

Figure: E. P. Wigner (1902–1995)

As $N \to \infty$, study asymptotic behaviors of eigenvalues and eigenvectors.

Four primary sources, Products of random matrices

♠ For a discrete-time evolution of a real or complex stochastic system

$$v(t+1) = X_{t+1} v(t), \quad t = 0, 1, 2, \ldots,$$

the total evolution is effectively driven by the **product of random matrices** at time $t = M$

$$\Pi_M = X_M \cdots X_2 X_1.$$

Four primary sources, Products of random matrices

♠ For a discrete-time evolution of a real or complex stochastic system

$$v(t+1) = X_{t+1} v(t), \quad t = 0, 1, 2, \ldots,$$

the total evolution is effectively driven by the **product of random matrices** at time $t = M$

$$\Pi_M = X_M \cdots X_2 X_1.$$

♡ Bellman [Limit theorems for non-commutative operations I. Duke Math., 1954], Furstenberg & Kesten [Products of random matrices., Ann. Math. Statist., 1960] initiated the study of products of random matrices and proved LLN & CLT.

Abel Prize Laureates 2020

The Norwegian Academy of Science and Letters has decided to award the Abel Prize for 2020 to

Hillel Furstenberg
Hebrew University of Jerusalem, Israel

Gregory Margulis
Yale University, New Haven, CT, USA

"for pioneering the use of methods from probability and dynamics in group theory, number theory and combinatorics."

A central branch of probability theory is the study of random walks, such as the route taken by a tourist exploring an unknown city by flipping a coin to decide between turning left or right at every cross. Hillel Furstenberg and Gregory Margulis invented similar random walk techniques to investigate the structure of linear groups, which are for instance sets of matrices closed under inverse and product. By taking products of randomly chosen matrices, one seeks to describe how the result grows and what this growth says about the structure of the group.

Furstenberg and Margulis introduced visionary and powerful concepts, solved formidable problems and discovered surprising and fruitful connections between group theory, probability theory, number theory, combinatorics and graph theory. Their work created a school of thought which has had a deep impact on many areas of mathematics and applications.

Starting from the study of random products of matrices, in 1963, Hillel Furstenberg introduced and classified a notion of fundamental importance, now called Furstenberg boundary. Using this, he gave a Poisson type formula expressing harmonic functions on a general group in terms of their boundary values. In his works on random walks at the beginning of the '60s, some in collaboration with Harry Kesten, he also obtained an important criterion for the positivity of the largest Lyapunov exponent.

Motivated by Diophantine approximation, in 1967, Furstenberg introduced the notion of disjointness of ergodic systems, a notion akin to that of being coprime for integers. This natural notion turned out to be extremely deep and have applications to a wide range of areas including signal processing and filtering questions in electrical engineering, the geometry of fractal sets, homogeneous flows and number theory. His "×2 ×3 conjecture" is a beautifully simple example which has led to many further developments. He considered the two maps taking squares and cubes on the complex unit circle, and proved that the only closed sets invariant under both these maps are either finite or the whole circle. His conjecture states that the only invariant measures are either finite or rotationally invariant. In spite of efforts

Lyapunov exponents

♡ In 1892, Lyapunov studied the stability of solutions of equations

$$\dot{v}(t) = X_t v, \quad v(0) = v_0 \in \mathbb{R}^N,$$

where $X_{(\cdot)}$ is a continuous and bounded function from $\mathbb{R}_+$ to the space of $N \times N$ real matrices, and proved that the Lyapunov exponent of a solution

$$\lambda(v_0) := \limsup_{t\to\infty} \frac{1}{t} \log \|v(t)\|.$$

was finite for every solution with $v_0 \neq 0$.

Lyapunov exponents

♡ In 1892, Lyapunov studied the stability of solutions of equations

$$\dot{v}(t) = X_t v, \quad v(0) = v_0 \in \mathbb{R}^N,$$

where $X_{(\cdot)}$ is a continuous and bounded function from $\mathbb{R}_+$ to the space of $N \times N$ real matrices, and proved that the Lyapunov exponent of a solution

$$\lambda(v_0) := \limsup_{t \to \infty} \frac{1}{t} \log \|v(t)\|.$$

was finite for every solution with $v_0 \neq 0$.

♠ Through the works of Furstenberg, Kesten, Oseledets, Kingman, Ruelle, Margulis, Avila and other mathematicians, Lyapunov exponents have recently emerged as an important concept in (stochastic) dynamical systems, products of random matrices and maps, spectral theory of (random) Schrödinger operators; see e.g. Wilkinson, What are Lyapunov exponents, and why are they interesting? BAMS 2017

Connection with

- ♠ Random Schrödinger operators
  Bougerol&Lacroix. Products of random matrices with applications to Schrodinger operators, 1985
- ♠ Dynamical systems
- ♠ Wireless communication, MIMO networks
- ♠ Free probability theory
- ♠ Deep Neural Networks ⇢ Deep Random Matrix Theory (Deep RMT)
- ♡ Random walks on matrix groups [Benoist&Quint, Random Walks on Reductive Groups] ⇝ Random walks on BIG matrix groups

Benoist&Quint, Random Walks on Reductive Groups

1 Introduction 1
- 1.1 What Is This Book About? 1
- 1.2 When Did This Topic Emerge? 2
- 1.3 Is This Topic Related to Sums of Random Numbers? 2
- 1.4 What Classical Results Should I Know? 3
- 1.5 Can You Show Me Some Nice Sample Results from This Topic? 5
- 1.6 How Does One Prove These Nice Results? 8
- 1.7 Can You Answer Your Own Questions Now? 9
- 1.8 Where Can I Find These Answers in This Book? 10
- 1.9 Why Is This Book Less Simple than These Samples? 10
- 1.10 Can You State These More General Limit Theorems? 11
- 1.11 Are the Proofs as Simple as for the Simple Samples? 12
- 1.12 Why Is the Iwasawa Cocycle so Important to You? 13
- 1.13 I Am Allergic to Local Fields. Is It Safe to Open This Book? . . 14
- 1.14 Why Are There so Many Chapters in This Book? 15
- 1.15 Whom Do You Thank? 16

Part I The Law of Large Numbers

2 Stationary Measures 19
- 2.1 Markov Operators 19
  - 2.1.1 Markov Chains on Standard Borel Spaces 19
  - 2.1.2 Measure-Preserving Markov Operators 20
  - 2.1.3 Ergodicity of Markov Operators 23
- 2.2 Ergodicity and the Forward Dynamical System 24
- 2.3 Markov–Feller Operators 26
- 2.4 Stationary Measures and the Forward Dynamical System 27
- 2.5 The Limit Measures and the Backward Dynamical System . . . 30
- 2.6 The Two-Sided Fibered Dynamical System 34
- 2.7 Proximal Stationary Measures 35

Table of Contents

1. Introduction
   - Origins
   - Lyapunov exponents
2. Local statistics
3. Double scaling limits

Law of Large Numbers

Furstenberg & Kesten [Ann. Math. Statist., 1960] initiated the study of products of random matrices and proved LLN & CLT.

♡ Classical LLN & CLT for $N = 1$:

$$\frac{1}{M} \log |\Pi_M| = \frac{1}{M} \sum_{k=1}^{M} \log |X_k|.$$

Law of Large Numbers

Furstenberg & Kesten [Ann. Math. Statist., 1960] initiated the study of products of random matrices and proved LLN & CLT.

♡ Classical LLN & CLT for $N = 1$:

$$\frac{1}{M} \log |\Pi_M| = \frac{1}{M} \sum_{k=1}^{M} \log |X_k|.$$

♠ [Furstenberg & Kesten '60] For any fixed N, the largest Lyapunov exponent with probability 1

$$\lambda_{\max} := \lim_{M\to\infty} \frac{1}{M} \log \|\Pi_M\|$$

exists. Further, all Lyapunov exponents exist by Multiplicative Ergodic Theorem [Oseledets, Trans. Moscow Math. Soc., '68],

$$\lambda_k := \lim_{M\to\infty} \frac{1}{2M} \log \left(k^{\text{th}} \,\text{largest eigenvalue of}\, \Pi_M^* \Pi_M\right), \quad k = 1, 2, \ldots, N.$$

Calculating Lyapunov exponents

Very hard to find explicit formulae of Lyapunov exponents, posed as an outstanding problem by [Kingman, Ann. Probab., '73].

- ♡ Largest exponent for a finite set of matrices with positive entries [Pollicott, Invent. Math., '10]
- ♠ All the Lyapunov spectrum for real Ginibre (zero-mean normal) ensemble [Newman, CMP'86] and complex Ginibre [Forrester, J. Stat. Phys., ' 13],

$$\lambda_k = \frac{1}{2}\left(\log \frac{2}{\beta} + \psi\big(\frac{\beta}{2}(N-k+1)\big)\right), \quad k = 1, \ldots, N,$$

where $\psi(x) := \Gamma'(x)/\Gamma(x)$ denotes the digamma function.

DPP structure

♢ Eigenvalue PDF of $\log(\Pi_M^* \Pi_M)$ (Determinantal point process):

$$P(x_1, \ldots, x_N) = C_N \det \left[g_{j-1}(x_i) \right]_{i,j=1}^N \prod_{1 \le j < k \le N} (e^{x_j} - e^{x_k}),$$

see [Akemann, Kieburg, Wei, J. Phys.A, '13]

♡ DPP with correlation kernel

$$P(x_1, \ldots, x_N) = \frac{1}{N!} \det \left[K_N(x_j, x_k) \right]_{i,j=1}^N,$$

$$K_N(x, y) = \int_{c-i\infty}^{c+i\infty} \frac{ds}{2\pi i} \oint_\Sigma \frac{dt}{2\pi i} \frac{e^{xt-ys}}{s-t} \frac{\Gamma(t)}{\Gamma(s)} \left(\frac{\Gamma(s+N)}{\Gamma(t+N)} \right)^{M+1},$$

where $c > 0$ and Σ encircles $0, -1, \ldots, 1-N$; [Kuijlaars-Zhang, CMP '14].

Local limits

♡ For any fixed N and as $M \to \infty$, N finite Lyapunov exponents for Π_M are asymptotically independent Gaussian random variables [Akemann, Burda, Kieburg, J. Phys.A, '14].

Local limits

♡ For any fixed N and as $M \to \infty$, N finite Lyapunov exponents for Π_M are asymptotically independent Gaussian random variables [Akemann, Burda, Kieburg, J. Phys.A, '14].

♠ For any fixed M and as $N \to \infty$, with the largest eigenvalue $x_{\max} = \max\{x_1, \ldots, x_N\}$, [L.-Wang-Zhang, AIHP Probab. Stat., '16]
$\mathrm{P}\left(\frac{2^{1/3}N^{2/3}}{(M+1)^{2/3}}\left(x_{max} - \log N^M - \log\frac{(M+1)^{M+1}}{M^M}\right) \leq t\right) \to F_2(t)$
where the GUE Tracy-Widom distribution

$$F_2(t) = 1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \int_{(t,\infty)^k} \det[K_{\mathrm{Airy}}(x_i, x_j)]_{i,j=1}^{k} d^k x,$$

$$K_{\mathrm{Airy}}(x, y) = \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}(y)\mathrm{Ai}'(x)}{x - y}$$

GUE Tracy-Widom distribution

'At the Far Ends of a New Universal Law' from Quanta Magazine.

A potent theory has emerged explaining a mysterious statistical law that arises throughout physics and mathematics

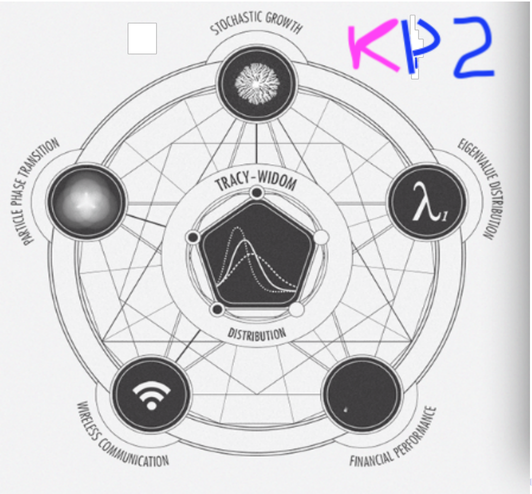


Table of Contents

1 Introduction
- Origins
- Lyapunov exponents

2 Local statistics

3 Double scaling limits

Many products of large matrices

Known results about products of Ginibre matrices $\Pi_M = X_M \cdots X_2 X_1$:
(1) Finite N & $M \to \infty$, Gaussian
(2) Finite M & $N \to \infty$, RMT statistics
A very natural question arises:
What happens when both M and N go to infinity?
Motivation 1:
Connect limit theorems in classical Probability Theory to RMT statistics in Random Matrix Theory
Motivation 2:
In Some Open Problems in Random Matrix Theory and the Theory of Integrable Systems. II [SIGMA, '17], P. Deift (ICM 2006, plenary; Poincare Prize) ended with "*There are many other areas, closely related to the problems in the above list, where much progress has been made in recent years, and where much remains to be done. These include: ...,* **singular values of** n **products of** $m \times m$ **random matrices as** $n, m \to \infty$, *and many others*".

Phase transitions

Finite Lyapunov exponents for Π_M, equivalently, eigenvalues of $\log(\Pi_M^* \Pi_M)$, two independent papers:

[1] G. Akemann, Z. Burda, and M. Kieburg. From integrable to chaotic systems: Universal local statistics of Lyapunov exponents, arXiv:1809.05905. Europhysics Letter, 126 (4), 40001: p1-p7, 2019.

[2] L., D. Wang(UCAS) , Y. Wang(HENU), Lyapunov exponent, universality and phase transition for products of random matrices, arXiv:1810.00433. To appear in CMP

A phase transition for the largest finite Lyapunov exponents

(I) Weakly correlated regime as $M/N \to \infty$, Gaussian

(II) Intermediate regime as $M/N \to \gamma \in (0, \infty)$, New distribution $F_{\text{crit}}(\gamma; t)$

(III) Strongly correlated regime of $M/N \to 0$, Tracy-Widom F_2

Criticality

$$F_{\text{crit}}(\gamma; t) = 1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \int_{(t,\infty)^k} \det[K_{\text{crit}}(\gamma; x_i, x_j)]_{i,j=1}^k d^k x,$$

where

$$K_{\text{crit}}(\gamma; \xi, \eta) = \int_{1-i\infty}^{1+i\infty} \frac{ds}{2\pi i} \oint_{\Sigma_{-\infty}} \frac{dt}{2\pi i} \frac{1}{s-t} \frac{\Gamma(t)}{\Gamma(s)} \frac{e^{\frac{\gamma s^2}{2} - \eta s}}{e^{\frac{\gamma t^2}{2} - \xi t}},$$

with $\Sigma_{-\infty}$ being a contour starting from $-\infty - i\epsilon$, looping around $\{0, -1, -2, \ldots\}$ positively, and then going to $-\infty + i\epsilon$.
Transition: From criticality, as $\gamma \to \infty$, Gaussian, while $\gamma \to 0$, GUE Tracy-Widom F_2 after proper scalings.

Relevant works, I

- ♠ Hanin, Nica, Products of Many Large Random Matrices and Gradients in Deep Neural Networks, arXiv:1812.05994, CMP 2020,
- ♠ Gorin, Sun, Gaussian fluctuations for products of random matrices, arXiv: 1812.06532, Am J Math 2022
  log-correlated Gaussian field (finite M) and Gaussian field
- ♡ Ahn, Fluctuations of β-Jacobi Product Processes, arXiv:1910.00743, PTRF2022
  truncated orthogonal, unitary, symplectic matrices real, complex, quaternion
- ♡ Hanin, Paouris, Non-asymptotic Results for Singular Values of Gaussian Matrix Products, arXiv: 2005.08899, GAFA 2021
- ♢ ...

Relevant works, II

Complex eigenvalues of $\Pi_M = X_M \cdots X_1$

♡ **A very similar transition for complex eigenvalues occurs!!!** [L., Yanhui Wang, Phase transitions for infinite products of large non-Hermitian random matrices, arXiv:1912.11910]

(I) Weakly correlated regime as $M/N \to \infty$, Gaussian
(II) Intermediate regime as $M/N \to \gamma \in (0, \infty)$, New distribution
(III) Strongly correlated regime of $M/N \to 0$, Ginibre statistics

♠ Jiang, Qi, Spectral Radii of Large Non-Hermitian Random Matrices, Journal of Theoretical Probability, 2017. Gumbel, a new distribution, logarithmic normal

Empirical Distributions of Eigenvalues of Product Ensembles Journal of Theoretical Probability, 2019 Uniform on the disk

♠ Eigenvalue statistics for Products of real Gaussian matrices, in progress, with Yanhui Wang

Future Problems

Product of independent random matrices $\Pi_M = X_M \cdots X_1$ where all X_K are of size N, consider relevant problems in the large N limit

- ♠ Question 1: Product of large (doubly) stochastic matrices
  Cf. Stochastic Models of Economic Optimization, Chen, Mu-Fa, *Eigenvalues, Inequalities, and Ergodic Theory*, chapter 10
- ♠ Question 2: Universality for finite-size Lyapunov exponents
  where all entries of X_k may be assumed to be i.i.d. with certain higher moments
- ♡ Far-seeing Plan:
  I) Random Walks on Big Matrix Groups
  II) Large-Dimensional Dynamical Systems

Weinan E, ICM 2022, *A Mathematical Perspective on Machine Learning*

Concluding remarks: This is all about math in high dimension

- **Compared with polynomials, neural networks provide a much more effective tool for approximating functions in high dimension.**
- Opens up a new subject in mathematics: **high dimensional analysis**.
  - supervised learning: high dimensional functions
  - unsupervised learning: high dimensional probability distributions
  - reinforcement learning: high dimensional Bellman equations
  - time series: high dimensional dynamical systems

Thank you!