NATIONAL RESEARCH
UNIVERSITY

HSE University

Random Matrices: Theory and Applications

Alexey Naumov

e-mail: anaumov@hse.ru

Moscow 26.02.2021

"I think I understand it more or less, it uses a lot of newer mathematical techniques, things that were developed in the 80s and 90s. Non-communicative geometry, random matrices. I think I learned more mathematics this week than I did in three years of grad school!..." - Hal.

Film „The Proof".

Part 1 : Growing dimension

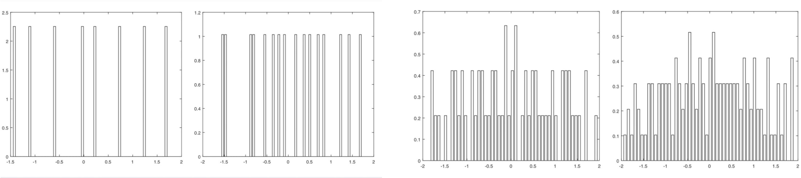
Symmetric matrices

- Let X be $n \times n$ random symmetric matrix ($X = X^T$)

Denote by $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ its eigenvalues.

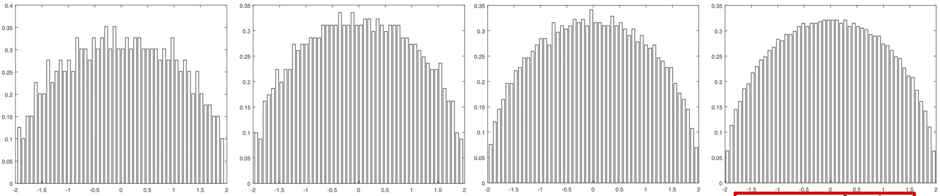
- Assume (for simplicity) $X_{\ell j} \overset{\text{i.i.d}}{\sim} N(0,1)$, $1 \leq \ell \leq j \leq n$

Calculate (e.g numerically) $\lambda_1, \ldots, \lambda_n$ and plot their histogram:



Increase n!

Semicircle law

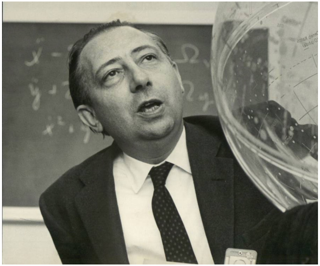


$\boxed{\text{semicircle law}}$

- **Empirical spectral distribution**

  Denote $\mu_n(A) := \frac{1}{n}\sum_{j=1}^{n} \delta_{\lambda_j/\sqrt{n}}(A)$,

  $A \subseteq \mathbb{R}$



E. Wigner

- Let $g_{sc}(\lambda) = \frac{1}{2\pi}\sqrt{(4-\lambda^2)_+}$, $(x)_+ = x \vee 0$

  be the density function of the **Wigner semicircle law**

Semicircle law (≈ 1955)

Theorem: For any $A \in \mathbb{R}$ with probability 1,

$$\lim_{n\to\infty} \mu_n(A) = \int_A g_{sc}(\lambda)\, d\lambda \qquad (1)$$

- What about λ_1 or λ_n? With probability 1,

$$\lim_{n\to\infty} \lambda_1(A)/\sqrt{n} = 2 \qquad (2)$$

(Bai, Yin)

$$A = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix} \quad \lambda_1(A) \sim 1$$

$$A = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \quad \lambda_1(A) \sim n$$

These laws are universal!

You need $E[X_{11}^2] = 1$ for (1)

$E[X_{11}^4] < C$ for (2)

Semicircular Law (Götze, Naumov, Tikhomirov 2015)

- Denote $B_i = \frac{1}{n} \sum_{j=1}^{n} \sigma_{ij}^2$, $\sigma_{ij}^2 = E[X_{ij}^2]$

- Assume:

(i) $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} |B_i^2 - 1| = 0$

(ii) $\max_{1 \le i \le n} B_i \le C$

(iii) $\lim_{n\to\infty} \frac{1}{n^2} \sum_{i,j=1}^{n} E[X_{ij}^2 \mathbb{1}(|X_{ij}| \ge \tau\sqrt{n})] = 0 \quad \forall \tau > 0$

Then the semicircle law holds!

Still open question: Are (i)–(iii) necessary conditions?

(Sample) covariance matrices

- Let $Y \in \mathbb{R}^d$, $E[Y] = 0$, $E[Y \otimes Y] = \Sigma$ (covariance matrix)

- PCA 1: find $\theta \in S^{d-1}$: $Var(\langle Y, \theta \rangle) \to \max_{\theta \in S^{d-1}}$

Solution:

- $Var(\langle Y, \theta\rangle) = E[\langle Y, \theta\rangle^2] = \sum_{j,k=1}^{n} \theta_j \theta_k E[Y_j Y_k] = \langle \Sigma\theta, \theta\rangle$
- Take θ - eigenvector of Σ corresponding to $\lambda_{max}(\Sigma)$

(Sample) covariance matrices

- PCA 2: find subspace $S \subseteq \mathbb{R}^d$ with $\dim S = k \ll d$ and Π_S projector onto S:

$$\mathbb{E}\left[\| Y - \Pi_S Y \|_2^2\right] \to \min_{S \subseteq \mathbb{R}^d,\, \dim S = k}$$

Solution:

- Take $S = \mathrm{span}(u_1, \dots, u_k)$

eigenvectors of Σ

Σ is usually unknown in practice :(

Sample covariance matrices

- Let $Y_1, \dots, Y_n$ be i.i.d. observations in $\mathbb{R}^d$: $E[Y_1] = 0$,
  $E[Y_1 \otimes Y_1] = E[Y_1 Y_1^T] = \Sigma$
- Σ is usually unknown in practice
- Replace Σ by $\hat{\Sigma} = \frac{1}{n} \sum_{j=1}^{n} X_j \otimes X_j$ $\quad (E[\hat{\Sigma}] = \Sigma)$
- There are many interesting questions
  - Distribution of eigenvalues of $\hat{\Sigma}$
  - Bounds for $\| \hat{\Sigma} - \Sigma \|_*$, $\| \cdot \|_*$ is some matrix norm
  - Stability of eigenvectors

Marchenko - Pastur Law

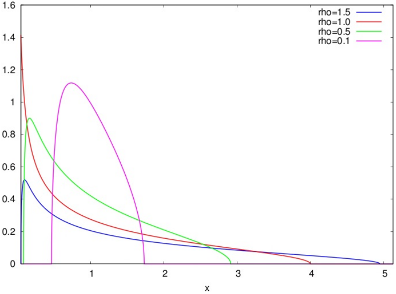
- Denote $p_{MP}(\lambda) = \frac{1}{2\pi\rho\lambda}\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}$

$\lambda_\pm = \sqrt{1 \pm \sqrt{\rho}}$, $\rho = \lim_{n\to\infty} \frac{d}{n} \in [0,1]$

- Let $s_1, \dots, s_d$ be the eigenvalues of $\hat{\Sigma}$

- Let $\mu_n(A) = \frac{1}{d}\sum_{j=1}^{d} \delta_{s_j}(A)$

V. Marchenko L. Pastur



Marchenko - Pastur Law ($\approx$ 1967)

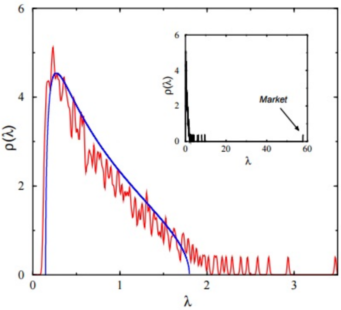
Theorem: Let $Y_1, \dots, Y_n$ be i.i.d, $EY_j = 0$, $EY_j \otimes Y_j = I$ for $\forall j \in [1,n]$. For any $A \subseteq R$ with probability 1

$$\lim_{n \to \infty} \mu_n(A) = \int_A p_{MP}(\lambda)\, d\lambda,$$

provided that $d/n \to p \in (0,1]$.

- Application (Test $H_0: \Sigma = I$)

- From MP to Deep neurall networks



Deep NN

- Let $x^0 = \{x^0_{j_0}\}_{j_0=1}^{n_0} \in \mathbb{R}^{n_0}$ – input
- $x^L = \{x^L_{j_L}\}_{j_L=1}^{n_L} \in \mathbb{R}^{n_L}$ – output
- $y^l = W^l x^{l-1} + b^l$, $x^l_{j_l} = \varphi(y^l_{j_l})$, $j_l = 1, .., n_l$, $l = 1, .., L$.

  W^l: $n_l \times n_{l-1}$ rect. matrices; b^l: Bias; φ: component-wise nonlinearity

- Input-output Jacobian

$$J^L_{\vec{n}_L} := \left\{ \frac{\partial x^L_{j_L}}{\partial x^0_{j_0}} \right\}_{j_0, j_L = 1}^{n_0, n_L} = \prod_{l=1}^{L} n_{l-1}^{-1/2} \cdot D^l X^l, \quad \vec{n}_L = (n_1, .., n_L)$$

$$D^l = \{ D^l_{j_l} \delta_{j_l k_l} \}_{j_l, k_l = 1}^{n_l}, \quad D^l_{j_l} = \varphi'\Big(n_{l-1}^{-1/2} \sum_{j_{l-1}=1}^{n_{l-1}} X^l_{j_l j_{l-1}} x^{l-1}_{j_{l-1}} + b^l_{j_l}\Big)$$

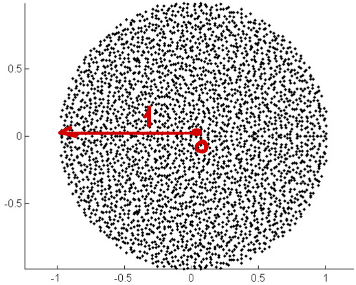
$$W^l = n_{l-1}^{-1/2} X^l$$

- One may calculate limit of ESD $\{J^L_{\vec{n}_L}\}^T J^L_{\vec{n}_L}$ when $\lim_{n_l \to \infty} \frac{n_{l-1}}{n_l} = c_l$

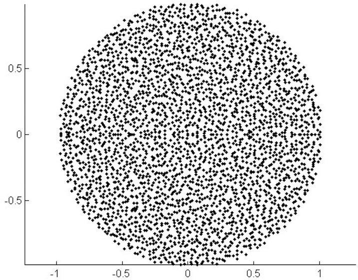
Non-Hermitian matrices (Circular law)

- Let X be $n \times n$ matrix, $X_{jk} \overset{i.i.d}{\sim} N(0,1)$ $1 \le j,k \le n$
- Let $\lambda_1, \dots, \lambda_n$ be eigenvalues of $X/\sqrt{n}$

Uniform distribution



Bernoulli



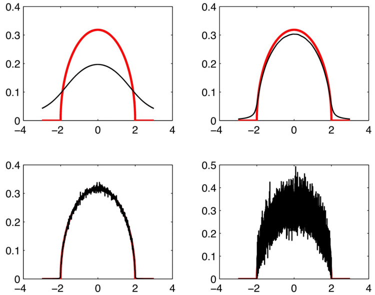
Gaussian

Girko (1985), Bai (2000), Gotze-Tikhomirov (2007), Tao-Vu (2010)

Ideas (Symmetric case)

- All information about eigenvalue is in the resolvent
  $$R(z) = (X/\sqrt{n} - zI), \quad z \in \mathbb{C}$$
- Stieltjes transform: Let μ be a (probability) measure
  $$s_\mu(z) = \int_{\mathbb{R}} \frac{\mu(d\lambda)}{\lambda - z} : \mathbb{C}^+ \to \mathbb{C}^+ \text{ analytic}$$
- $$m_n(z) = s_{\mu_n}(z) = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{\lambda_j/\sqrt{n} - z} = \boxed{\frac{1}{n} \operatorname{Tr} R(z)}$$
- $$\operatorname{Im} s_\mu(z) = v\int \frac{\mu(d\lambda)}{(\lambda-u)^2+v^2} \Rightarrow \frac{1}{\pi}\operatorname{Im} s_\mu(u+iv) = \mu \overset{\text{convolution}}{*} \underbrace{p_v(u)}_{\text{Cauchy dist}}$$
- $$\boxed{\frac{d\mu}{du}(u) = \lim_{v\to 0} \frac{1}{\pi}\operatorname{Im} s_\mu(u+iv)}$$

(One to one correspondence)



Ideas (Symmetric case)

- $s_{sc}(z) = -\frac{z}{2} + \sqrt{\frac{z^2}{4} - 1}$ or $1 + z\, s_{sc}(z) + s_{sc}^2(z) = 0$

- $1 + z\, m_n(z) + m_n^2(z) = T_n$ with $|T_n| \sim \frac{1}{nv^2}$

- Fixed $v > 0$: Take $n \to \infty$. Then $T_n \to 0$ and $E[m_n(z)] \to s_{sc}(z)$
(macroscopic scale, global regime)

- $v \sim \frac{\log n}{n}$. Information about individual eigenvalues, rates of convergence
(microscopic scale, local regime)

See e.g. L. Erdös – H.-T. Yau, ..., T. Tao – V. Vu, Götze - N. - Tikhomirov, ... (2007 --)

- $\tilde m_n(z) = E[m_n(z)]$, $m_n(z) \approx \tilde m_n(z)$ (concentration)

- $\tilde m_n(z) = \frac{1}{n} E[\mathrm{Tr}\, R(z)] = \frac{1}{n} E[\mathrm{Tr}\{(X/\sqrt{n} - zI)^{-1}\}]$

- $(X/\sqrt{n} - zI)\, R(z) = I \implies 1 + z\tilde m_n(z) - \frac{1}{n^{3/2}} E[\mathrm{Tr}\{X R(z)\}] = 0$

- $E[\mathrm{Tr}\{X R(z)\}] = \sum_{j,k=1}^{n} E[X_{jk} R_{kj}(z)] = -\frac{1}{\sqrt{n}} \sum_{j,k=1}^{n} E[R_{jj}(z) R_{kk}(z)]$

$- \frac{1}{\sqrt{n}} \sum_{j,k} R_{jk}(z) R_{kj}(z) =$

$= -\frac{1}{\sqrt{n}} E[\{\mathrm{Tr}\, R(z)\}^2] - \frac{1}{\sqrt{n}} \sum_{j,k} E[R_{jk}^2(z)]$

Stein:
$E[\xi f(\xi)] = E[f'(\xi)]$
$\iff \xi \sim N(0,1)$
$f \in C_b^{(1)}$

- $E[\{\mathrm{Tr}\, R(z)\}^2] \approx \{E[\mathrm{Tr}\, R(z)]\}^2$, $\|R\|_F^2 \le \frac{n}{v^2}$

- $1 + z\tilde m_n(z) + \tilde m_n^2(z) = T_n$, $T_n \approx \frac{1}{nv^2}$

Non-Hermitian case

- Eigen-spectrum is very unstable

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 \cdots & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \cdots & & 1 \\ 0 & 0 & \cdots & & 0 \end{pmatrix} \qquad B = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 \cdots & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \cdots & & 1 \\ \varepsilon_n & 0 & \cdots & & 0 \end{pmatrix}$$

$$\lambda_1(A) = \ldots = \lambda_n(A) = 0$$

δ_0 (Dirac at 0)

$$B^n = \varepsilon_n I \Rightarrow \lambda_k(B) = \varepsilon_n^{1/n} e^{2k\pi i/n}$$

uniform dist

$\varepsilon_n^{1/n} \to 1 \quad n \to \infty$

- Singular-spectrum is stable!

$$AA^T = \mathrm{diag}(1, \ldots, 1, 0) \quad , \quad BB^T = (1, \ldots, 1, \varepsilon_n^2)$$

Non-Hermitian case. Relation between spectras.

- $$U_{\mu_A}(z) = -\int_{\mathbb{C}} \log|\lambda - z|\, d\mu_A(z) = -\frac{1}{n}\log|\det(A-zI)| =$$

ESD of A

$$= -\frac{1}{n}\log\det\left(\sqrt{(A-zI)(A-zI)^*}\right) = -\int_0^\infty \log t\, d\nu_{A-zI}(t)$$

ESD of Hermitian matrix

Victory transform (V) by V.Girko!

Smallest singular value

- Let A be $n \times n$ matrix
- Singular values : $s_1(A) = \sup_{\|x\|_2=1} \|Ax\|_2$
  $s_n(A) = \inf_{\|x\|_2=1} \|Ax\|_2$
- $s_1(A) \lesssim \sqrt{n}$ with high probality (4 moments !)
- von Neumann : $s_n(A) \sim n^{-1/2}$.
- Edelman : $P(s_n(A) \leq \varepsilon n^{-1/2}) \lesssim \varepsilon$ in the Gaussian case
- Kahn, Komlós, Szemerédi : $P(\text{random sign matrix } A \text{ is singular}) < c^n$ for some $c \in (0,1)$.

K. Tikhomirov proved that $c = \frac{1}{2} + \bar{o}(1)$, Annals of math, 2020

Smallest singular value

Theorem. Let $\xi_1, \dots, \xi_n$ be independent, mean zero, $E[\xi_i^2] \geq 1$, sub-gaussian random variables. Let A be $n \times n$ matrix whose rows are independent copies of $(\xi_1, \dots, \xi_n)$. Then for every $\varepsilon \geq 0$

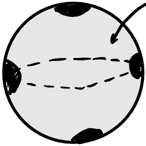
$$P\left(s_n(A) \leq \varepsilon n^{-1/2}\right) \leq C\varepsilon + c^n,$$

where $C > 0$ and $c \in (0,1)$.

- Conjectured by Spielman and Teng. Proved by Rudelson and Vershynin. (≈ 2007)
- $A \leftarrow A + M$, M is some non-random matrix. With high probability $\frac{s_1(A+M)}{s_n(A+M)} \sim n^c$.

Smallest singular value (Idea)

- $S^{n-1} = \text{Comp} \cup \text{Incomp}$



Incomp $\approx$ delocalized

Comp $\approx$ Sparce (most coordinates are zero)

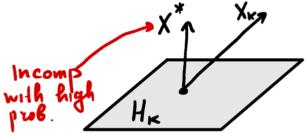
- Comp could be controlled by union Bound (covered by ε-nets)

- Incomp

$$P\Big(\inf_{x \in \text{Incomp}} \|Ax\|_2 \lesssim \varepsilon n^{-1/2}\Big) \lesssim \frac{1}{n} \sum_{k=1}^{n} P(\operatorname{dist}(X_k, H_k) < \varepsilon) \sim \varepsilon + n^{-1/2}$$

under CLT

$$A = [X_1, \dots, X_n], \quad H_k = \operatorname{span}(X_\ell, \ell \neq k)$$



$$\operatorname{dist}(X_k, H_k) \geq |\langle X_k, X^* \rangle|, \quad X^* \perp H, \quad \|X^*\|_2 = 1$$

independent

Use anti-concentration !

Johnson - Lindenstrauss Lemma (Dimension reduction)

- Data : set of N points in $\mathbb{R}^n$
- Compress the data : represent it as a set of N points in $\mathbb{R}^m$, $m \ll n$

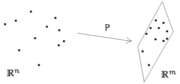
Theorem Let $\mathcal{X}$ be a set of N points in $\mathbb{R}^n$ and $\varepsilon \in (0,1)$. Consider an $m \times n$ matrix A whose rows are independent, mean zero, isotropic and sub-gaussian vectors in $\mathbb{R}^n$. Define

$$P = \frac{A}{\sqrt{m}}. \quad \text{and assume that} \quad m \geq C\varepsilon^{-2} \log N.$$

Then with high probality

$$(1-\varepsilon)\|x-y\|_2 \leq \|Px - Py\|_2 \leq (1+\varepsilon)\|x-y\|_2 \quad \forall\, x, y \in \mathcal{X}$$

Almost isometry !



Johnson - Lindenstrauss Lemma (Dimension reduction)

- Let $z = \frac{x-y}{\|x-y\|_2} \in S^{n-1}$

- $(1-\varepsilon) \leq \|Pz\|_2 \leq 1+\varepsilon$

- $(Pz)_i = \frac{1}{\sqrt{m}} \langle \xi_i , z \rangle$, $P = \begin{bmatrix} \xi_1 \\ \xi_m \end{bmatrix} \in \mathbb{R}^{m \times n}$

- $\left| \frac{1}{m} \sum_{i=1}^{m} \langle \xi_i , z \rangle^2 - 1 \right| \leq \varepsilon$

- Use concentration of measure

$$P\left(\left| \frac{1}{m} \sum_{i=1}^{m} \langle \xi_i , z \rangle^2 - 1 \right| \geq \varepsilon \right) \leq 2 e^{-mc\varepsilon^2}$$

- $P\left(\max_z \left| \frac{1}{m} \sum_{i=1}^{m} \langle \xi_i , z \rangle^2 - 1 \right| \geq \varepsilon \right) \leq N^2 e^{-mc\varepsilon^2}$

- Take $m \geq c\, \varepsilon^{-2} \log N$

Part 2: Fixed dimension

Sums of random matrices

- Let $X_1, \dots, X_n$ be i.i.d random matrices in $\mathbb{R}^{d \times d}$, $E[X_1] = A$
- Form $S_n = \frac{1}{n} \sum_{j=1}^{n} X_j$. Does it concentrate around A?
- Example 1. Recall sample covariance matrices

$$\underbrace{\hat{\Sigma}}_{S_n} = \frac{1}{n} \sum_{j=1}^{n} \underbrace{Y_j \otimes Y_j}_{X_j}, \quad A = \Sigma$$

Sums of random matrices

- Example 2. Randomized matrix multiplications

Let $B \in \mathbb{R}^{d_1 \times N}$, $C \in \mathbb{R}^{N \times d_2}$. Calculate $\underbrace{BC}_{\substack{O(N d_1 d_2) \\ \text{operations}}}$.

$$BC = \sum_{j=1}^{N} \underbrace{b_{:j}}_{\substack{j\text{-th} \\ \text{column} \\ \text{of } B}} \underbrace{c_{j:}}_{\substack{j\text{-th} \\ \text{row} \\ \text{of } C}} .$$

Define random matrix $R = \frac{1}{p_j} b_{j:} c_{j:}$ with probability p_j

$$p_j = \frac{\|b_{:j}\|^2 + \|c_{j:}\|^2}{\|B\|_F^2 + \|C\|_F^2}, \quad j = 1, \dots, N$$

$$E[R] = BC \ !!!$$

Form $S_n = \frac{1}{n} \sum_{j=1}^{n} R_j$, $R_j \overset{i.i.d}{\sim} R$ $\quad O(n d_1 d_2)$ operations

Sums of random matrices

- **Theorem** (Matrix Bernstein inequality)
  Let $X_1, \dots, X_n$ be independent, $E[X_j] = 0$, $X_j = X_j^T$, $\|X_j\| \le K$. Then for any $t \ge 0$

  $$P\left(\left\|\sum_{j=1}^n X_j\right\| \ge t\right) \le 2d \exp\left(-\frac{t^2/2}{\sigma^2 + Kt/3}\right)$$

  Here $\sigma^2 = \left\|\sum_{j=1}^n E[X_j^2]\right\|$ "matrix variance"

- In the non-Hermitian case

  $$X \to \begin{pmatrix} X & 0 \\ 0 & X^T \end{pmatrix}$$

- $e^{X+Y} \ne e^X e^Y$:(But $E_Y[\operatorname{tr}\exp(X+Y)] \le \operatorname{tr}\exp(X + \log E[e^Y])$
  X, Y symmetric

- Example 1. Recall sample covariance matrices

$$\underbrace{\hat{\Sigma}}_{S_n} = \frac{1}{n}\sum_{j=1}^{n} \underbrace{Y_j \otimes Y_j}_{X_j}, \quad A = \Sigma$$

Assume that $\|Y_j\| \lesssim E\|Y_j\|_2^2 = \operatorname{tr}\Sigma$. Then

$$E\|\hat{\Sigma} - \Sigma\| \lesssim \|\Sigma\|\left(\sqrt{\frac{r(\Sigma)\log n}{n}} + \frac{r(\Sigma)\log n}{n}\right),$$

$$r(\Sigma) = \frac{\operatorname{Tr}\Sigma}{\|\Sigma\|} \quad \text{effective dimension}$$

Products of random matrices

- We would like to solve $F(\theta) = 0$, $F : \mathbb{R}^m \to \mathbb{R}^n$
- For some reason we cannot compute $F(x)$ at any desirable point and have only some unbiased estimate $F(\theta, \xi)$: $E[F(\theta, \xi)] = F(\theta)$.
- Robbins - Monro (1951):
  1. Set some $\theta_0 \in \mathbb{R}^m$
  2. $\theta_{k+1} = \theta_k + \alpha_{k+1} F(\theta_k, \xi_{k+1})$

  This procedure converges !

Linear stochastic approximation

- $F(\theta) = -A\theta + b$. We have $\bar{A}(\cdot), \bar{b}(\cdot)$: $E[\bar{A}] = A$, $E[\bar{b}] = b$
- Linear SA: $\theta_{k+1} = \theta_k + \alpha_{k+1}\left(-\bar{A}(Z_{k+1})\theta_k + \bar{b}(Z_{k+1})\right)$
- $\theta_n - \theta^* = \sum_{j=1}^{n} \alpha_j \prod_{\ell=j+1}^{n} (I - \alpha_\ell \bar{A}(Z_\ell))\, \bar{\varepsilon}(Z_j) + \prod_{j=1}^{n} (I - \alpha_j \bar{A}(Z_j)) \{\theta_0 - \theta^*\}$

$\bar{\varepsilon}(Z_j) = \bar{b}(Z_j) - b - (\bar{A}(Z_j) - A)\,\theta^*$.

- Examples:

- stochastic gradient methods (with momentum)
- TD (λ) learning in Reinforcement learning

Stability (submitted to COLT 2021)

- If we wish to estimate

$$E\left[\|\theta_n - \theta^*\|_2^p\right]$$

we need to study exponential stability

$$\forall m \le n, \; E\left[\left\|\prod_{j=m+1}^{n} (I - \alpha_j \bar{A}(Z_j))\right\|^p\right] \le C_p \exp\left(-a_p \sum_{j=m+1}^{n} \alpha_j\right)$$

$\{Z_j\}_{j \ge 1}$ could be i.i.d or Markov chain

- Compare with

$$\left\|\prod_{j=m+1}^{n} (I - \alpha_j A)\right\| \le K_Q \exp\left(-a \sum_{j=m+1}^{n} \alpha_j\right),$$

provided that $-A$ is Hurwitz ($\mathrm{Re}\,\lambda(-A) < 0$).

$\exists\; Q \succ 0 : \quad A^T Q + Q A = I, \quad K_Q = \sqrt{\dfrac{\lambda_{max}(Q)}{\lambda_{min}(Q)}}$

If $A = A^T \Rightarrow a = \lambda_{min}(A)$.

References

1. F. Götze, A. Naumov, A. Tikhomirov. Limit theorems for two classes of random matrices with dependent entries, 2015

2. L. Pastur. On random matrices arising in deep neural networks. Gaussian case, 2020

3. R. Vershynin. High-dimensional probality. 2019

4. J. Tropp. An Introduction to matrix concentration inequalities

5. T. Tao Topics in RMT, 2012

6. L. Erdös, H.T. Yau. A dynamical approach to RMT, 2017

Thank you!