2020 暑期学校 RMT 笔记

马宇恒

Summer 2020

1 Random Matrix Theory

1.1 Intuition

$c = \frac{p}{n}$ 往往默认比较小，但当 c 不再是很小，而是以一个缓慢的速度收敛到 0，收敛到远离 0 的常数，大于 0，甚至 p 为 n 的多项式的时候，矩阵的特征值将不再呈现原始分布。课程要解决的问题包括：

- 如何变换才能还原特征值的分布？
- 给定观察 Y=X+S，X 为噪声，三者为 n*p 矩阵，如何去噪声？

1.2 两种模型

先给两种不常见的关系。

Definition 1. *Roughly equal. For some constant c, order of X, a RV, is roughly c iff anyϵ small and some large fixed D, when n is large enough,* $P(||X|| \geq n^{\epsilon}\sqrt{p}) \leq n^{-D}$

Definition 2. *Q norm.*

$$||X||_q = (E|X|^q)^{\frac{1}{q}}$$

给了常用的两种矩阵模型。

Definition 3. *Wigner Matrix.* $H = (h_{ij})_{n\times n}$ 为对称矩阵，且满足：

- h_{ij} 独立，$1 \le i \le j \le n$
- $Eh_{ij} = 0$
- $Eh_{ij}^2 = \frac{1}{n}(1 + \delta_{ij})$=*Var*($h_{ij}$)
- $||\sqrt{n}h_{ij}||_q$ *bdd by some constant* M_q，但是后面会讲到这一点是可以弱化的

若这里的每个小 h 都满足正态分布,则 H 为一个 *GOE*(*Gaussian Orthogonal Ensemble*)。

Definition 4. *Wishart Matrix.* 和多元里一样。*Let* $W = XX^T$ *where* X *is some* $p \times n$ *matrix and has*

- $X = (x_1, x_2, \cdots, x_n)$, x_i 是一个观察
- $Ex_{ij} = 0$
- $||\sqrt{n}x_{ij}||_q$ *bdd by some* M_q

1.3 四种方法

随机矩阵特征值分布主要有四种方法来分析。

- Moment method. 思路是如果有足够多的矩，那么我们想要的东西都可以通过这些信息求出来。缺点是无法明确写出分布，并且需要分布的空间满足一定条件（指数族 for instance）。不是这节课的重点。想看可以找白志东老师的书。
- Dynamic approach. 重点。
- Free probability theory. 结合了统计物理（还是理论物理来着。。）
- orthogonal polynomial. （听起来有点像数值那一套）

1.4 以 Wigner 为例

先给几个定义。

Definition 5. *Empirical spectrum distribution.* $\mu_n(x) = \frac{1}{n}\sum_{i=1}^{\infty} 1(\lambda_i \le x)$

Definition 6. *Resolvent or Green function of* H *is* $G(z) = (H - zI)^{-1}$. *Where* z 是一个复数，$z = E + i\eta$ *and* $\eta > 0$

关于这个 Resolvent 我们会有一系列的恒等式来求或者估计它的元素，他基本上也代表了 H 的大部分信息，我们下面的大部分工作也都是为了计算 resolvent。理由就是下面这个东西。

Definition 7. *Stieltjes transform*。傅立叶变换的推广。$m_n(z) = \int \frac{1}{x-z} du_n(x)$. 这里的 u_n 看作一个通常意义的分布函数。它可以完全被 *resolvent* 决定，经过计算有 $m_n(z) = \frac{1}{n}\sum G_{ii}(z) = \frac{1}{n}tr(G(z))$。

我们下一个定义，并且陈述一个有关事实，这是接下来我们要证的东西和未来几天的工具。

Definition 8. *Converge weakly in probability. For any bdd continuous f on R, any* ϵ*>0,* u_n *converges weakly in probability to u iff* $P(|\int f d(u_n - u)| > \epsilon) \to 0$.

Here's a fact: for a sequence of u_n (ESD in practice), if their corresponding $m_n(z)$ converges to some m(z) for all z, then m(z) corresponds to a u and u_n converges to u weakly in probability.

那么我们如何找到这个 u？有 Inversion formula 如下。

Thm

Definition 9. *Inversion formula.* *u(x) is distribution and* ρ *is density, which is* $du(x)$*=*ρdx. $\rho(x) = \lim_{\eta} \frac{Im(m(x+i\eta))}{\pi}$.

这就是说分布的 stieltjes transform 在 z 逼近实数轴时可以 recover 出对应的 u。综上，有 resolvent 就有 stieltjes transform，就有 density，所以我们的任务就是找 resolvent。下面列出一些恒等式或者事实以及对应的注释，详细的证明在手写文档。

1.5 计算 Resolvent

Notation 1. *For a wigner matrix H and a set of integers ranging from 1 to n,* $H^{(T)}$ *is to remove column and row where their index is in T.* $G^{(T)} = (H^{(T)} - Iz)^{-1}$

Notation 2. *For a wigner matrix H and a set of integers ranging from 1 to n,* $H^{[T]}$ *is to set* $\{h_{ij}|i,j \in T\}$ *to 0. and* $G^{[T]} = (H^{(T)} - Iz)^{-1}$

课上老师做了一些修改，需要加这么一个 notation，notation 2 似乎只在 Wishart 用到?

Notation 3. $\sum_{ij}^{(T)} H_{ij}$ *is summation over* $\{(i,j)|i,j \notin T\}$

Theorem 1. *Schurs's Complement. A is n*n, B is n*m, C is m*n, D is m*m, A,D invertible.*

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} \left(A - BD^{-1}C\right)^{-1} & -\left(A - BD^{-1}C\right)^{-1} BD^{-1} \\ -D^{-1}C\left(A - BD^{-1}C\right)^{-1} & D^{-1} + D^{-1}C\left(A - BD^{-1}C\right)^{-1} BD^{-1} \end{bmatrix}$$

这是线性代数的基本结论，但是是接下来计算中最重要的技巧。

Theorem 2. *Diagonal entries.*

$$G_{ii}(z) = \frac{1}{H_{ii} - z - \sum_{kl}^{(i)} H_{ik} G_{kl}^{(i)} H_{li}}$$

证明见手写文档证明 1. 有一个没有证明的事实就是我们会期待每个 G_{ii} 差别不会太大，所以我们事实上可以期待有这样一个关系 $m_n(z) \approx G_{ii}(z)$. 注意到定义中 $Eh_{ii}^2 = \frac{2}{n}$，所以该项不太大。先抛出“G 的非对角元素都很小”和“G 不会受到局部变化影响”这两个结论，于是我们有 $\frac{1}{m_n(z)} \approx 0 - z - \sum H_{ik} G_{kl}^{(i)} H_{li} \approx \frac{1}{m_n(z)} \approx 0 - z - \sum H_{ik} G_{kk}^{(i)} H_{ki} \approx 0 - z - \frac{1}{n} \sum G_{kk}^{(i)} \approx -z - m$，拿到了一个关于自守函数的 stieltjes transform 的等式，如此我们便可以解出特征值的分布，称为 semicircle law。那么我们就来验证 G 的非对角元素很小。

Theorem 3. *A useful equation.*

$$G_{ij}^{(T)} = G_{ij}^{T\cup k} + \frac{G_{ij}^{(T)} G_{kj}^{(T)}}{G_{kk}^{(T)}}, i, j, k \notin T$$

In particular when T *is empty.*

$$G_{ij} = G_{ij}^{(k)} + \frac{G_{ik} G_{kj}}{G_{kk}}$$

证明有两种思路，一个是使用 Woodbury matrix identity，参考白志东老师书。另一个就是使用技巧，对矩阵元素求导 (element wise)，参考手写文档。证明过程中使用了下一个定理。这个定理说明了我们 “G 不会受到局部变化的影响”，即我们 remove 一个 index 并没有什么太大区别。不过下这个结论还需要剩下的余项很小，即下一个结论阐述的事情。

Theorem 4. *Off-diagonal entries.*

$$\begin{aligned} G_{ij}^{(T)} &= -\, G_{ii}^{(T)} \sum_{k}^{T\cup\{i\}} H_{ik} G_{kj}^{T\cup\{i\}} \\ &= -\, G_{jj}^{(T)} \sum_{k}^{T\cup\{j\}} G_{ik}^{T\cup\{j\}} H_{kj} \end{aligned}$$

In particular, when T *is empty,*

$$\begin{aligned} G_{ij} &= -\, G_{ii} \sum_{k}^{(i)} H_{ik} G_{kj}^{(i)} \\ &= -\, G_{jj} \sum_{k}^{(j)} G_{ik}^{(j)} H_{kj} \end{aligned}$$

证明还是使用 Schur's compliment，参考手写文档。由此我们可以说明 “非对角元素很小” 的结论。考虑 $G_{12} = -G_{11} \sum_k^{(1)} H_{1k} G_{k2}^{(1)}$，根据下文的 Concentrating Inequality，我们有 G_{ij} 是能被 G_{ii} 以 $\frac{1}{\sqrt{n}}$ 限制住，从而达到目的。到此为止，我们的所有 Resolvent 的元素都是可以计算的。

1.6 Wishart Matrix 和 Resolvent

Definition 10. *Resolvent of Wishart matrix. Recall that $W = XX^T$ is p by p. Then two resolvent functions are defined as:*

$$G(z) = (XX^T - zI)^{-1}, z \in C^+$$

$$g(z) = (X^T X - zI)^{-1}, z \in C^+$$

Theorem 5.

$$Tr(g(z)) - Tr(G(z)) = \frac{p-n}{2}$$

证明只需要考虑特征值，利用 AB 与 BA 非零特征值集合相等。还是首先计算对角元素，依旧是使用 Schur 来证明。

Theorem 6.

$$G_{ii}(z) = \frac{1}{-z - z(X_i^T G^{(i)} X_i)}$$

证明中用到了恒等式 .$x\left(x^T x - zI\right)^{-1} x^T = I + z\left(xx^T - zI\right)$.
至此，如果我们延续上面的分析，会有

$$\begin{aligned} m_n(z) &= G_{ii}(z) \\ &= \frac{1}{-z - z(\frac{Tr(g)}{p} + \frac{p-n}{2})} \\ &= \frac{1}{1 - z - c - zcm} \end{aligned}$$

于是有自守函数

$$\rho(x) = \frac{1}{2\pi c}\sqrt{\frac{((1+\sqrt{c})^2 - x)(x - (1-\sqrt{c})^2)}{x^2}}$$

称为 Marchenko-Pauthr Law。那么这些事情成立依赖于以下几点。

Theorem 7.

$$g_{ij}(z) = g_{ij}^{(k)}(z) + \frac{g_{ik}(z) g_{kj}(z)}{g_{kk}(z)}, i, k \neq k$$

$$g_{ij}(z) = z g_{ii}(z) g_{jjj}(z) + x_i^T G^{(ij)}(z) x_j, i \neq j$$

1.7 Concentrating Inequality

很重要的结论，用来证明非对角元素很小。总共四个。

Theorem 8. *For $x_i \in R$, independent, $Ex_i = 0$, $||x_i||_q < C_q$, b_i's are fixed real number. Then*

$$||\sum_{i=1}^{n} b_i x_i||_q \leq \sqrt{C_q' q} C_q \sqrt{\sum |b_i|^2}$$

Remark 1. *If x's and b's are independent RVs, thm still holds except it is bounded by another RV.*

Theorem 9. *x_i and y_i are all independent, q norm bdd, $a_{ij} \in R$ fixed (or independent from x and y), then*

$$||\sum_{ij} x_i a_{ij} y_j||_q \leq C_q' C_q^2 (\sum |a_{ij}|^2)^{\frac{1}{2}}$$

Theorem 10. *Same setting of last thm,*

$$||\sum_{i \neq j} x_i a_{ij} y_j||_q \leq C_q' C_q^2 (\sum_{i \neq j} |a_{ij}|^2)^{\frac{1}{2}}$$

Theorem 11.

Theorem 12. *Ward Identity.*

$$\sum_j |G_{ij}|^2 = \frac{1}{\eta} Im G_{ii}$$

Remark 2. *Recall that we need to constrain $\sum_{k \neq l}^{(i)} H_{ik} G_{kl}^{(i)} H_{li}$, note that G can be treated as constant since independence. Then*

$$|\sum_{k \neq l}^{(i)} H_{ik} G_{kl}^{(i)} H_{li}| \leq C \frac{1}{n} (\sum G_{kl}^{(i)})^{\frac{1}{2}}$$

And by ward identity this is

$$\sqrt{\sum_{k}^{(i)} \frac{1}{n^2 \eta} Im G_{kk}^{(i)}}$$

1.8 Some Remark

我们来考察自守函数和他的 st transform。对于 Wigner Matrix，经过整理有 Im(m(z)) is of same order of $\frac{\eta}{\sqrt{K+\eta}}$ when |E|>2 and Im(m(z)) is of$\sqrt{K+\eta}$ when |E|<2, where K=|E-2|. 后面我们会考虑 spiked matrix，这是与统计结合的较为紧密的内容，常用来做信号存在的假设检验。如果 spike 进来的信号特征值很大，那么他会跑出-2 到 2 的区域，并且也会脱离 $\frac{\eta}{\sqrt{K+\eta}}$ 的控制。

1.9 Local Law of RMT

之前讲的 MP law 和 semicircle law 均属于 global law，阐述了密度的（弱）收敛性。而 local law 量化了收敛的速度。

Definition 11. *Spectrum domine.*

$$S(\tau) = \{z = E + i\eta||E| < \frac{1}{\tau}, n^{-1+\tau} \leq \eta \leq \frac{1}{\tau}\}$$

Theorem 13. *Local law of Wigner matrix.*

$$|G_{ij} - \delta_{ij}m(z)| \sim \phi(z)$$

where $\phi(z) = \sqrt{\frac{Im(m(z))}{n\eta}} + \frac{1}{n\eta}$.

$$|m_n(z) - m(z)| \sim \frac{1}{n\eta}$$

这个定理证明非常繁复。基本思路是先证明一个 weak local law，用一个很容易满足的初始条件推进到 strong local law 需要满足的某些条件，还结合了 bootstrapping。详情参照 literature 前两篇。Weak local law 的简略证明也可以参照手写文档。值得注意的一点是，我们前面用到的估计是 $\frac{1}{G_{ii}} - \frac{1}{m}$，$x^{-1}$ 是病态函数，所以我们的结果是不够用的。

Theorem 14. *Stability condition. For self-consistent equation* $u^2+zu+1 = 0$, $z \in S(\tau)$, *if there are two solutions* m_1 *and* m_2, *which satisfy*

- $|m_1 - m_2| \sim \sqrt{K+\eta}$

- *u_0 solves $u^2 + zu + 1 + r = 0$ where $|r|<1$*

then $min(|u - m_1|, |u - m_2|) \leq \frac{c|r|}{\sqrt{K+\eta+|r|}}$

证明见手写文档。如此一来，在 Wigner 中有 $|m_1 - m_2| = |\sqrt{z^2-4}| \approx \sqrt{K+\eta}$, 满足条件，所以

$$
\begin{aligned}
&1 + m_n G_{ii} + z G_{ii} \\
=&(1 + m_n^2 + z m_n) + (1 + m_n G_{ii} + z G_{ii}) - (1 + m_n^2 + z m_n) \\
=&1 + m_n^2 + z m_n + O(n^{-\frac{1}{2}})
\end{aligned}
$$

之前的陈述是正确的。

1.10 Rigidity of Eigenvalues

Local law 的一个应用，描述了单个特征值收敛到响应 quantile 的速度。

Theorem 15. *Rigidity of Eigenvalues.*

$$|\lambda_i(H) - \gamma_i| \sim n^{-\frac{2}{3}} (min(i, n+1-i))^{-\frac{1}{3}}$$

where $\int_{\gamma_i}^{\infty} du = \frac{i}{n}$

i 很小或者很大的时候，收敛速度慢些，分布是有更明确的 distribution 的，参考 literature 部分。在中间的时候形成了某种不可解释的 process。这个结果也解释了得到的图像中点分布的密度。密度高的地方打点密些，低了就稀疏一些。

对于 Wishart 来说有一样的结果。同样，对于最大的特征值，有如下结果。

Theorem 16.

$$n^{\frac{2}{3}} (\lambda_1(H) - 2) \to TW$$
$$n^{\frac{2}{3}} \frac{\lambda_1(H) - (1+\sqrt{c})^2}{\sigma} \to TW$$

TW 表示 Tracy–Widom 分布，其中 σ 的值可以在文献中找到。注意到既然有了明确的分布，那么假设检验的相关理论就可以用了。

1.11 Complete Eigenvector Delocolization

Theorem 17. *For a Wigner matrix, element of any of its eigenvector is roughly* $\frac{1}{n}$.

证明参考手写文档。

1.12 Spiked Covariance Matrix Model

Definition 12.

$$\tilde{H} = H + dvv^T$$

is a spiked covariance matrix.

这部分内容原本与主成分分析有关，但是没讲。这个矩阵有很多结论，比如特征值包含 λ s.t. $1 + dm(\lambda) = 0$, 也就是 $d + \frac{1}{d}$. 参照手写笔记中的步骤我们发现，当 v 的结构相对简单（稀疏），我们很容易能够限制得到上面的结论。但是当 v 的结构逐渐复杂，原来可以消去的余项变大了。于是我们有下一个 section。

1.13 Isotropic Local Law

这个 local law 是不会受到特征向量结构所影响的。

Theorem 18. *Isotropica Local Law.*

$$||u^T G(z)v - m(z)u^T v|| \sim \sqrt{\frac{Im(m(z))}{n\eta} + \frac{1}{n\eta}}$$

1.14 General Covariance Matrix Structure

Definition 13. *For some* $x_1, \cdot, x_p$, $E(x_i) = 0$, $Cov(x_i) = \Sigma$, *then*

$$W_1 = \Sigma^{\frac{1}{2}} YY^T \Sigma^{-\frac{1}{2}}$$

is a generalized cov matrix. Note that Wishart matrix is when Σ*=I.*

这类矩阵也被称为 Deformed Wishart 或者 Sample Cov Matrix。这类矩阵 Spiked 以后特征值收敛到 1+d+c+$\frac{c}{d}$.

Theorem 19. *Deformed MP law. For deformed Wishart matrix, their selfconsistent equation is*

$$\frac{1}{m} = -z + c\int \frac{x}{1+mx}\pi dx$$

1.15 Literature review

Local Law of Wigner Matrix: Rigidity of Eigenvalues of Generalized Wigner Matrices, Jun Yin，一般的情形足够使用，不需要读 Generalized 情形。着重看矩存在的必要性。

Local Law of Wishat Matrix: Universality of covariance matrices, Jun Yin. 最小特征值很困难，重点看。

Isotropic Local Laws: Isotropic Local Laws for Sample Covariance and Generalized Wigner Matrices, Jun Yin.

Deformed MP Law: Anisotropic local laws for random matrices, Jun Yin.

Spiked Cov Matrix: On the principal components of sample covariance matrices, Jun Yin.
ASYMPTOTICS OF SAMPLE EIGENSTRUCTURE FOR A LARGE DIMENSIONAL SPIKED COVARIANCE MODEL, Debashis Paul.

Matrix denoising: high dimensional deformed rectangular matrices with applications in matrix denoising, Xiucai Ding.
Singular vector and singular subspace distribution for the matrix denoising model, Bao Zhigang.

General Introduction of RMT to Statistics: high dimensional statistical inference and random matrices, Iain Johnstone.
random matrix theory in statistics a review, Paul A.