Tracy–Widom limit for sample covariance matrices

Kevin Schnelli

KTH Royal Institute of Technology

Joint work with Jong Yun Hwang (KAIST) and Ji Oon Lee (KAIST)

May 10, 2019

Sample covariance matrix

Consider a (mean-zero) multivariate Gaussian random variable:

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix} \in \mathbb{R}^M , \qquad \boldsymbol{y} \sim \mathcal{N}(0, \Sigma) .$$

$\Sigma = \mathbb{E}\boldsymbol{y}\boldsymbol{y}^*$, i.e., $\Sigma_{\alpha\beta} = \mathbb{E}y_\alpha y_\beta$: covariance matrix

Sample covariance matrix:

$$\widehat{\Sigma} := \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{y}_i \boldsymbol{y}_i^* , \qquad \boldsymbol{y}_i \ : \text{samples} .$$

Traditional setup: M fixed, $N \nearrow \infty$, then $\widehat{\Sigma} \to \Sigma = \mathbb{E}\boldsymbol{y}\boldsymbol{y}^*$.

High-dimensional setup:

$$N/M =: d_N \to d \in (0, \infty) , \qquad N \nearrow \infty .$$

Wishart matrix: $\Sigma = \mathrm{I}$

Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M$ denote the ordered eigenvalues of $\widehat{\Sigma}$. For any fixed interval J

$$\left| \mathcal{N}_J - \int_J \rho_{\mathrm{MP}}(x)\mathrm{d}x \right| \to 0\,, \qquad \text{with} \qquad \mathcal{N}_J := \frac{1}{M} \#\{j \,:\, \lambda_j \in J\}\,,$$

almost surely as $N \to \infty$, where ρ_{MP} is the Marchenko-Pastur law:

$$\rho_{\mathrm{MP}}(x)\mathrm{d}x = \begin{cases} \nu_d(x)\mathrm{d}x + (1-d)\delta_0(x)\mathrm{d}x\,, & \text{if } d < 1\,, \\ \nu_d(x)\mathrm{d}x\,, & \text{if } d \geq 1\,, \end{cases}$$

where

$$\nu_d(x) = \frac{d}{2\pi} \frac{\sqrt{(E_+ - x)(x - E_-)}}{x} 1_{[E_-, E_+]}(x) \qquad \text{and} \qquad E_\pm = (1 \pm d^{-1/2})^2\,.$$

Figure : Histogram of the eigenvalues of $\widehat{\Sigma}$: $\Sigma = \mathrm{I}$, $N = 2000$, $d = 2$.

Largest eigenvalue: $\Sigma = \mathrm{I}$

Theorem.

Let E_+ denote the upper endpoint of the support of ρ_{MP}. Then for any (small) $\varepsilon > 0$ and (large) $D > 0$

$$\mathbb{P}\Big(|\lambda_1 - E_+| \geq \frac{N^{\varepsilon}}{N^{2/3}}\Big) \leq \frac{1}{N^D}\,,$$

for $N \geq N_0(\varepsilon, D)$.

Hence, we expect that the fluctuations of the largest eigenvalue are of order $N^{-2/3}$.

Largest eigenvalue: $\Sigma = \mathrm{I}$

Theorem: (Johnstone 2001, Johansson 2000)

For real, respectively complex, Gaussians with $\Sigma = I$, there is a constant $\gamma = \gamma(d)$ such that

$$\lim_{N\to\infty} \mathbb{P}\left(\gamma N^{2/3}(\lambda_1 - E_+) \le s\right) = F_\beta(s)\,, \qquad \beta = 1, 2\,.$$

Tracy–Widom distributions:

$$F_2(s) := \exp\left(-\int_s^\infty (x-s)q(x)^2 dx\right)\,, \quad F_1(s) := \exp\left(-\frac{1}{2}\int_s^\infty q(x)dx\right)(F_2(s))^{1/2}\,,$$

where q satisfies

$$q'' = sq + 2q^3, \quad q(s) \sim Ai(s) \text{ as } s \to \infty.$$

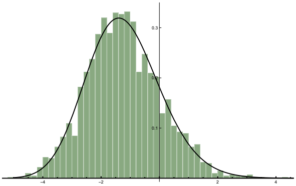

Figure : Histogram of 2000 samples of $\gamma N^{2/3}(\lambda_1 - E_+)$ of $N = 1000$ $d = 2$, real Gaussians.

Universality results

Replace Gaussian distributions by general distributions.

Definition.

Let $X = (x_{\alpha i})$ be an $M \times N$ matrix whose entries are i.i.d. real random variables such that

$$\mathbb{E}x_{\alpha i} = 0, \qquad \mathbb{E}|x_{\alpha i}|^2 = N^{-1}, \qquad \mathbb{E}|x_{\alpha i}|^k \leqslant \frac{C_k}{N^{k/2}}\,.$$

Assume that $M = M(N)$ with $N/M \to d \in (0, \infty)$ as $N \to \infty$.
Sample covariance matrix: $\widehat{\Sigma} = XX^*$.

Theorem (Pillai and Yin 2014).

Let E_+ denote the upper endpoint of the support of ρ_{MP}. Then for any (small) $\varepsilon > 0$ and (large) $D > 0$

$$\lim_{N \to \infty} \mathbb{P}\left(\gamma N^{2/3}(\lambda_1 - E_+) \le s\right) = F_1(s)\,,$$

for $N \geq N_0(\varepsilon, D)$, where $\gamma = d^{-1/6}(d^{1/4} + d^{-1/4})^{-1}$.

Sparse sample covariance matrices

Motivation: Biadjacency matrix of bipartite graph.

Two vertex sets: $V = \{v_\alpha\}_{\alpha=1}^M$ has size M; $W = \{w_i\}_{i=1}^N$ has size N. Edges only between V and W. Biadjacency matrix:

$$B = (b_{\alpha i})\,, \qquad 1 \le \alpha \le M\,, \quad 1 \le i \le N\,.$$

$$b_{\alpha i} = \begin{cases} 1\,, & \text{if there is an edge between } v_\alpha \text{ and } w_i, \\ 0\,, & \text{else } . \end{cases}$$

Sparse sample covariance matrices

Biadjacency matrix of bipartite Erdős-Rényi graph:

$$B = (b_{\alpha i})\,,$$

$(b_{\alpha i})$ are i.i.d. random variables satisfying

$$b_{\alpha i} = \begin{cases} 1\,, & \text{with probability } p \\ 0\,, & \text{with probability } 1-p\,, \end{cases}$$

Remarks:

- $\mathbb{E} b_{\alpha i}^k = p$, $k \geq 1$.
- We allow p to depend on N.

We say that B is sparse if $p \ll 1$.

Sparse sample covariance matrices

Center and normalize $B = (b_{\alpha i})$:

$$X_{\alpha i} := \frac{b_{\alpha i} - p}{\sqrt{Np(1-p)}} \,.$$

Then,

$$\mathbb{E} X_{\alpha i} = 0 \,, \qquad \mathbb{E} X_{\alpha i}^2 = \frac{1}{N} \,, \qquad \mathbb{E} X_{\alpha i}^k = \frac{1}{N (Np)^{(k-2)/2}} \big(1 + O(p)\big) \,, \qquad (k \geq 3) \,.$$

Introduce the **sparsity parameter** q by

$$q^2 := pN \,, \qquad 0 < q \leq c N^{1/2} \,.$$

Hence,

$$\mathbb{E} X_{\alpha i}^k = \frac{1}{N q^{k-2}} \big(1 + O(q^2/N)\big) \,, \qquad k \geq 3 \,.$$

Sparse sample covariance matrix

Definition.

Let $X = (X_{\alpha i})$ be an $M \times N$ matrix whose entries are real i.i.d. random variables such that

$$\mathbb{E}X_{\alpha i} = 0, \qquad \mathbb{E}|X_{\alpha i}|^2 = N^{-1}, \qquad \mathbb{E}|X_{\alpha i}|^k \le \frac{C_k}{Nq^{k-2}}$$

for some $q = N^{\phi}$, $0 < \phi \le \frac{1}{2}$. Assume that $M = M(N)$ with $N/M \to d \in (0, \infty)$ as $N \to \infty$.
Sample covariance matrix $\widehat{\Sigma} := XX^*$.

Remark: For any $\phi > 0$, the eigenvalues of $\widehat{\Sigma}$ follow the Marchenko-Pastur law on global scale.

Rescaled cumulants: Let $\kappa^{(k)}$ be the k-th cumulant of $X_{\alpha i}$:

$$\log \mathbb{E}\left[\mathrm{e}^{tX_{\alpha i}}\right] = \sum_{k=1}^{\infty} \kappa^{(k)} \frac{t^k}{k!}\,, \qquad \kappa^{(1)} = 0\,, \qquad \kappa^{(2)} = \frac{1}{N}\,.$$

Set, for $k \ge 3$,

$$s^{(1)} := 0\,, \qquad s^{(2)} := 1\,, \qquad s^{(k)} := Nq^{k-2}\kappa^{(k)}.$$

Estimates on largest eigenvalue

Theorem (Hwang-Lee-S. '18).

Consider the largest eigenvalue λ_1 of XX^*. Assume that $q \geq N^{\phi}$, $\phi > 0$. Then, there is $L_+ \in \mathbb{R}$ such that for any $\varepsilon > 0$ and $D > 0$,

$$\mathbb{P}\Big(\big|\lambda_1 - L_+\big| > N^{\varepsilon}\Big(\frac{1}{q^4} + \frac{1}{N^{2/3}}\Big)\Big) < N^{-D}\,,$$

for $N \geq N_0(\varepsilon, D)$, where

$$L_+ := \left(1 + \frac{1}{\sqrt{d}}\right)^2 + \frac{1}{\sqrt{d}}\left(1 + \frac{1}{\sqrt{d}}\right)^2 \frac{s^{(4)}}{q^2} + O(q^{-4})\,.$$

Tracy–Widom limit

Theorem (Hwang-Lee-S. '18).

Assume that $\phi > 1/6$ (so that $q \gg N^{1/6}$). Let λ_1 be the largest eigenvalue of XX^*. Then,

$$\lim_{N\to\infty} \mathbb{P}\big(\gamma N^{2/3}(\lambda_1 - L_+) \leqslant s\big) = F_1(s)$$

for all $s \in \mathbb{R}$, where $\gamma^{-1} = d^{1/6}(d^{1/4} + d^{-1/4})$.

Remark: For $N^{1/6} \ll q \leq N^{1/3}$, the spectral shift $L_+ - E_+ \simeq \frac{1}{q^2}$ is much larger than the TW-fluctuations.

Remark: If $\phi < 1/6$, it is believed that the limiting distribution will be Gaussian; cf. Huang-Landon-Yau '17.

(Some) previous results

- TW for Wishart matrix (X is Gaussian, $\Sigma = I$)
  - Complex case: Johansson (2000)
  - Real case: Johnstone (2001)
- Null case ($\Sigma = I$)
  - Edge universality: Pillai-Yin (2014), Ding-Yang (2018)
- (Phase transition for) Spiked Wishart matrix (X is Gaussian, Σ is a finite-rank perturbation of I)
  - Complex case: Baik-Ben Arous-Péché (2005)
  - Real case: Bloemendal-Virág (2011)
- Spiked sample covariance matrix (Σ is a finite-rank perturbation of I)
  - Edge universality: Bloemendal-Knowles-Yau-Yin (2014)
- Non-null case ($\Sigma \neq I$)
  - TW for Gaussian, complex case: El Karoui (2007), real case: Lee-S. (2016), Fan-Johnstone (2017)
  - Edge universality: Bao-Pan-Zhou (2015), Knowles-Yin (2017).

Tools: Stieltjes transform and Green function

- Given a probability measure μ on $\mathbb{R}$, its Stieltjes transform is defined as

$$m_\mu(z) := \int_{\mathbb{R}} \frac{\mathrm{d}\mu(x)}{x - z}\,, \qquad z \in \mathbb{C}^+\,.$$

- For μ the Marchenko-Pastur law, we have

$$m_{\mathrm{MP}}(z) = \frac{\left(z + 1 - \frac{1}{d}\right) + \sqrt{\left(z + 1 - \frac{1}{d}\right)^2 - 4z}}{2z}\,.$$

- Hence, $m_{\mathrm{MP}}(z)$ is the solution to

$$1 + z m_{\mathrm{MP}} + \left(z m_{\mathrm{MP}} + 1 - \frac{1}{d}\right) m_{\mathrm{MP}}(z) = 0\,,$$

  with $m_{\mathrm{MP}}(z) \in \mathbb{C}^+$, for $z \in \mathbb{C}^+$.

- In the sparse setup we make the ansatz

$$1 + z\widetilde{m}(z) + \left(z\widetilde{m}(z) + 1 - \frac{1}{d}\right)\widetilde{m}(z) + \frac{s^{(4)}}{q^2}\left(z\widetilde{m}(z) + 1 - \frac{1}{d}\right)^2 \widetilde{m}(z)^2 = 0\,,$$

  and pick the solution with $\widetilde{m}(z) \in \mathbb{C}^+$, $z \in \mathbb{C}^+$.

Tools: Stieltjes transform and Green function

- In the sparse setup we make the ansatz

$$1 + z\widetilde{m}(z) + \big(z\widetilde{m}(z) + 1 - \frac{1}{d}\big)\widetilde{m}(z) + \frac{s^{(4)}}{q^2}\left(z\widetilde{m}(z) + 1 - \frac{1}{d}\right)^2 \widetilde{m}(z)^2 = 0\,,$$

  and pick the solution with $\widetilde{m}(z) \in \mathbb{C}^+$, $z \in \mathbb{C}^+$.

- Then there is a probability measure $\widetilde{\rho}$ such that $\widetilde{m}(z) = \int_{\mathbb{R}} \frac{\mathrm{d}\widetilde{\rho}(x)}{x-z}$ and with support on $[L_-, L_+]$ where

$$L_\pm = \left(1 \pm \frac{1}{\sqrt{d}}\right)^2 \pm \frac{1}{\sqrt{d}}\left(1 \pm \frac{1}{\sqrt{d}}\right)^2 \frac{s^{(4)}}{q^2} + O(q^{-4})\,.$$

  Moreover, we have $\widetilde{\rho}(x) \sim \sqrt{(L_+ - x)_+}$, for x near the upper edge.

- Green function/resolvent of $Q := X^*X$ (N by N matrix)

$$G^{X^*X}(z) := \frac{1}{X^*X - z}\,.$$

- By spectral calculus

$$\frac{1}{N}\mathrm{Tr}G^{X^*X}(z) = \sum_{i=1}^{N} \frac{1}{\lambda_i(Q) - z}\,.$$

Tools: Stieltjes transform and Green function

- Green function/resolvent of $Q := X^*X$ (N by N matrix)

$$G^{X^*X}(z) := \frac{1}{X^*X - z}\,.$$

- By spectral calculus

$$\frac{1}{N}\mathrm{Tr}G^{X^*X}(z) = \sum_{i=1}^{N} \frac{1}{\lambda_i(Q) - z}\,.$$

Local law at the edge:

$$\left|\frac{1}{N}\mathrm{Tr}G^{X^*X}(z) - \widetilde{m}(z)\right| \le N^{\varepsilon}\left(\frac{1}{q^2} + \frac{1}{N\,\mathrm{Im}\,z}\right)\,,$$

with high probability, for all $z = E + \mathrm{i}\eta$ with $E \in [L_+ - c, L_+ + 1]$, $N^{-1+\varepsilon'} \le \eta \le 1$.

Linearization

- Define an $(N+M)\times(N+M)$ matrix (the linearization of Q)

$$H(z) := \left(\begin{array}{c|c} -zI_{N\times N} & X^* \\ \hline X & -I_{M\times M} \end{array}\right), \qquad z \in \mathbb{C}^+ .$$

- Introduce the "Green function" $G(z) := H(z)^{-1}$. By the Schur complement formula,

$$G_{ij}(z) = \left(\frac{1}{X^*X - zI}\right)_{ij}, \qquad 1 \le i,j \le N ,$$

$$G_{\alpha\beta}(z) = z\left(\frac{1}{XX^* - zI}\right)_{\alpha\beta}, \qquad N+1 \le \alpha,\beta \le N+M .$$

Weak local law (Ding-Yang 2018)

$$|G_{ab}(z) - \Pi_{ab}(z)| \le N^{\varepsilon}\left(\frac{1}{q} + \sqrt{\frac{\operatorname{Im} m_{MP}(z)}{N\operatorname{Im} z}} + \frac{1}{N\operatorname{Im} z}\right), \qquad 1 \le a,b \le N+M ,$$

with high probability, for all $z = E + \mathrm{i}\eta$ with $E \in [L_+ - c, L_+ + 1]$, $N^{-1+\varepsilon'} \le \eta \le 1$, where

$$\Pi(z) := \left(\begin{array}{c|c} m_{\mathrm{MP}}(z) I_{N\times N} & 0 \\ \hline 0 & (1+m_{\mathrm{MP}}(z))^{-1} I_{M\times M} \end{array}\right) .$$

Moment estimate

Define $P : \mathbb{C}^+ \to \mathbb{C}$ by

$$P(m) := 1 + zm + \big(zm + 1 - \frac{1}{d}\big)m + \frac{s^{(4)}}{q^2} m^2 \left(zm + 1 - \frac{1}{d}\right)^2 .$$

Then we had $P(\widetilde{m}(z)) = 0$. Hence we expect that

$$\left| P\left(\frac{1}{N} \mathrm{Tr} \frac{1}{X^* X - z}\right)\right| \leq N^{\varepsilon} \Psi(z) ,$$

with high probability for some "small" control parameter $\Psi(z)$. In other words:

$$\mathbb{E}\left[\overline{P(m(z))}^{D} P(m(z))^{D}\right] \leq N^{2D\varepsilon} \Psi(z)^{2D} , \qquad m(z) := \frac{1}{N} \mathrm{Tr} \frac{1}{X^* X - z} .$$

Recursive moment estimate I

We would hope for a recursive estimate of the form

$$\mathbb{E}\left[|P(m(z))|^{2D}\right] \lesssim N^{\varepsilon}\mathbb{E}\left[\left(\frac{1}{q^4} + \frac{\operatorname{Im}\widetilde{m}(z)}{N\eta}\right)|P(m(z))|^{2D-1}\right], \qquad \eta = \operatorname{Im} z\,.$$

Then Hölder or Young would give

$$\mathbb{E}\left[|P(m(z))|^{2D}\right] \leq CN^{\varepsilon}\left(\frac{1}{q^4} + \frac{\operatorname{Im}\widetilde{m}(z)}{N\eta}\right)^{2D},$$

and we would be in good shape to apply Markov.

Recursive moment estimate II

We would hope that

$$\mathbb{E}\left[|P(m(z))|^{2D}\right] \lesssim N^{\varepsilon}\mathbb{E}\left[\left(\frac{1}{q^4}+\frac{\operatorname{Im}\widetilde{m}(z)}{N\operatorname{Im}z}\right)|P(m(z))|^{2D-1}\right].$$

Then Hölder or Young would give $\mathbb{E}\left[|P(m(z))|^{2D}\right] \le CN^{\varepsilon}\left(\frac{1}{q^4}+\frac{\operatorname{Im}\widetilde{m}(z)}{N\operatorname{Im}z}\right)^{2D}$, and we would be in good shape.

Lemma.

$$\begin{aligned}
\mathbb{E}|P(m)|^{2D} \leqslant{}& N^{\varepsilon}\mathbb{E}\Big[\Big(\frac{1}{q^4}+\frac{\operatorname{Im}m}{N\eta}+\frac{N-M}{N^2}\Big)|P(m)|^{2D-1}\Big] \\
&+N^{-\varepsilon/4}q^{-1}\mathbb{E}\Big[|m-\widetilde{m}|^2|P(m)|^{2D-1}\Big]+N^{\varepsilon}q^{-8D} \\
&+N^{\varepsilon}q^{-1}\sum_{s=2}^{2D}\sum_{u'=0}^{s-2}\mathbb{E}\Big[\Big(\frac{\operatorname{Im}m}{N\eta}+\frac{N-M}{N^2}\Big)^{2s-u'-2}|P'(m)|^{u'}|P(m)|^{2D-s}\Big] \\
&+N^{\varepsilon}\sum_{s=2}^{2D}\mathbb{E}\Big[\Big(\frac{1}{N\eta}+\frac{1}{q}\Big(\frac{\operatorname{Im}m}{N\eta}+\frac{N-M}{N^2}\Big)^{1/2}+\frac{1}{q^2}\Big) \\
&\qquad\times\Big(\frac{\operatorname{Im}m}{N\eta}+\frac{N-M}{N^2}\Big)^{s-1}|P'(m)|^{s-1}|P(m)|^{2D-s}\Big] =: \Phi(z)\,.
\end{aligned}$$

Recursive moment estimate III

Recall that

$$P(m) = 1 + zm + \big(zm + 1 - \frac{1}{d}\big)m + \frac{s^{(4)}}{q^2}\left(zm + 1 - \frac{1}{d}\right)^2 m^2 \,.$$

Recall the definition of the "Green function" $H(z)G(z) = I$, in components,

$$1 + zG_{ii}(z) = \sum_{\alpha=N+1}^{M+N} (X^*)_{i\alpha} G_{\alpha i}(z) = 1\,, \qquad 1 \le i \le N\,,$$

and

$$m(z) = \frac{1}{N}\sum_{i=1}^{N} G_{ii}(z)\,.$$

Hence

$$\mathbb{E}\big[(1+zm)P(m)^{D-1}\overline{P(m)}^{D}\big] = \frac{1}{N}\sum_{i,\alpha}\mathbb{E}\big[X_{\alpha i}G_{\alpha i}P(m)^{D-1}\overline{P(m)}^{D}\big]\,.$$

Cumulant expansion

Stein Lemma

Fix $\ell \in \mathbb{N}$ and let $F \in C^{\ell+1}(\mathbb{R};\mathbb{C}^+)$. Let Y be a centered random variable with finite moments to order $\ell+2$. Then,

$$\mathbb{E}[YF(Y)] = \sum_{r=1}^{\ell} \frac{\kappa^{(r+1)}(Y)}{r!}\mathbb{E}[F^{(r)}(Y)] + \mathbb{E}[\Omega_\ell(YF(Y))]\,,$$

where $\kappa^{(r+1)}(Y)$ denotes the $(r+1)$-st cumulant of Y and $F^{(r)}$ denotes the r-th derivative of the function F. The error is controlled in terms of $F^{(\ell+1)}$ and $\kappa^{(\ell+2)}$.

Applied to our recursive moment estimate

$$\mathbb{E}[(1+zm)P^{D-1}\overline{P^D}] = \frac{1}{N}\sum_{r=1}^{\ell}\frac{1}{r!}\frac{s^{(r+1)}}{Nq^{r-1}}\mathbb{E}\Big[\sum_{i,\alpha}\partial_{\alpha i}^r\big(G_{\alpha i}P^{D-1}\overline{P^D}\big)\Big] + \mathbb{E}\big(\Omega_\ell\big)\,,$$

where $\partial_{\alpha i}^r = \frac{\partial^r}{(\partial X_{\alpha i})^r}$. We stop expanding at order $\ell = 8D$.

First order terms $r = 1$

Need to compute

$$\frac{1}{N}\frac{s^{(2)}}{N}\mathbb{E}\Big[\sum_{i,\alpha}\partial_{\alpha i}\big(G_{\alpha i}P^{D-1}\overline{P^D}\big)\Big],$$

where $s^{(2)} = 1$. Using $\partial_{\alpha i}G_{\alpha i} = -G_{ii}G_{\alpha\alpha} - G_{\alpha i}G_{\alpha i}$, we get

$$\begin{aligned}\partial_{\alpha i}\big(G_{\alpha i}P^{D-1}\overline{P^D}\big) = &\,(-G_{ii}G_{\alpha\alpha} - G_{\alpha i}G_{\alpha i})P^{D-1}\overline{P^D}\\ &+ (D-1)P'(m)G_{\alpha i}\frac{1}{N}\sum_j \partial_{\alpha i}G_{jj}P^{D-2}\overline{P^D}\\ &+ D\overline{P'(m)}G_{\alpha i}\frac{1}{N}\sum_j \partial_{\alpha i}\overline{G_{jj}}P^{D-1}\overline{P^{D-1}}\,.\end{aligned}$$

Leading order one term:

$$-\mathbb{E}\Big[\frac{1}{N^2}\sum_{i,\alpha}G_{ii}G_{\alpha\alpha}P^{D-1}\overline{P^D}\Big] = -\mathbb{E}\Big[m\frac{1}{N}\sum_{\alpha}G_{\alpha\alpha}P^{D-1}\overline{P^D}\Big]\,.$$

Since X^*X and XX^* share the same non-zero eigenvalues, we have

$$\frac{1}{N}\sum_{\alpha}G_{\alpha\alpha} = \frac{1}{N}\sum_{\alpha}\left(\frac{z}{XX^*-z}\right)_{\alpha\alpha} = \frac{z}{N}\sum_i\left(\frac{1}{X^*X-z}\right)_{ii} + \frac{N-M}{N} \quad = zm + 1 - \frac{1}{d}\,.$$

First order terms $r = 1$

Hence

$$\mathbb{E}\Big[(1+zm)P(m)^{D-1}\overline{P(m)}^{D}\Big] = \frac{1}{N^2}\mathbb{E}\Big[\sum_{i,\alpha}\partial_{\alpha i}\big(G_{\alpha i}P^{D-1}\overline{P^D}\big)\Big] + \mathrm{h.o.t}(r \geq 2) + \Phi(z)$$
$$= -\mathbb{E}\Big[m\left(zm + 1 - \frac{1}{d}\right)P^{D-1}\overline{P^D}\Big]\Big] + \mathrm{h.o.t}(r \geq 2) + \Phi(z)\,.$$

and we get cancellation to leading order in the cumulant expansion.

Higher order terms in the cumulant expansion: Terms involving the third cumulant $r = 2$ are negligible: Third moment does not matter.

$r = 3$ terms

$$\frac{1}{3!}\frac{s^{(4)}}{Nq^2}\frac{1}{N}\sum_{i,\alpha}\mathbb{E}\left[\partial_{\alpha i}^3\left(G_{\alpha i}P^{D-1}\overline{P^D}\right)\right] = \frac{1}{3!}\frac{s^{(4)}}{N^2q^2}\mathbb{E}\Big[\sum_{i,\alpha}(\partial_{\alpha i}G_{\alpha i})P^{D-1}\overline{P^D}\Big] + O(\Phi(z))\,.$$

Need to compute

$$\frac{1}{3!}\partial_{\alpha i}^3 G_{\alpha i} = -G_{ii}^2 G_{\alpha\alpha}^2 + \text{terms involing off-diagonal Green function entries}\,.$$

Leading term for $r = 3$:

$$-\frac{s^{(4)}}{q^2}\mathbb{E}\Big[\Big(\frac{1}{N}\sum_i G_{ii}^2\Big)\Big(\frac{1}{N}\sum_\alpha G_{\alpha\alpha}^2\Big)P^{D-1}\overline{P^D}\Big]$$

Lemma.

$$-\frac{s^{(4)}}{q^2}\mathbb{E}\Big[\frac{1}{N^2}\sum_{i,\alpha}G_{ii}^2G_{\alpha\alpha}^2P^{D-1}\overline{P^D}\Big] = -\frac{s^{(4)}}{q^2}\mathbb{E}\left[m^2\left(\frac{1}{N}\sum_\alpha G_{\alpha\alpha}\right)^2P^{D-1}\overline{P^D}\right] + O(\Phi)$$

$$= -\frac{s^{(4)}}{q^2}\mathbb{E}[m^2(zm + 1 - \frac{1}{d})^2P^{D-1}\overline{P^D}] + O(\Phi)\,.$$

Wrapping things up:

- We argued that

$$\mathbb{E}|P(m)|^{2D} \leq N\Phi(z)\,.$$

  Hence we get a high probability estimate on $|P(m)|$.

- Study the local stability of the equation $P(\widetilde{m}(z)) = 0$, yields the local law.
- Recursive moment estimate combined with local law give an estimate on $|\lambda_1(X^*X) - L_+|$.

Ideas to prove the Tracy–Widom fluctuations

- The fluctuations of the largest eigenvalues can be extracted from the expectation of the Green function when the spectral parameter is close to the edge.
- Introduce a continuous interpolation between the given sparse sample covariance matrix and the Wishart covariance matrix: Dyson matrix flow/Ornstein-Uhlenbeck process.

$$\mathrm{d}X_{\alpha i} = \frac{1}{\sqrt{N}}\mathrm{d}B_{\alpha i} - \frac{1}{2}X_{\alpha i}\mathrm{d}t \qquad \Longrightarrow \qquad q = q_t\,, \qquad L_+ = L_+(t)\,.$$

- Follow the associated flow of the Green function and estimate its change over time.

$$\mathrm{d}\mathbb{E}[N\,\mathrm{Im}\, m(L_+ + \kappa + \mathrm{i}\eta_0)] = \sum_{\alpha,i,j}\mathrm{Im}\,\mathbb{E}\Big[-\frac{1}{2}\frac{\partial G_{ii}}{\partial X_{\alpha j}}X_{\alpha i} + \frac{1}{2N}\frac{\partial^2 G_{ii}}{(\partial X_{\alpha j})^2}\Big]\mathrm{d}t\,.$$

- This change is offset by changing the spectral parameter, time-dependent spectral parameter.

$$\mathbb{E}\big[\sum_{i,j}\dot{L}_+ G_{ij}G_{ji}\big]\mathrm{d}t\,, \qquad \text{with} \qquad \dot{L}_+ \simeq 2\frac{1}{\sqrt{d}}\big(1+\frac{1}{\sqrt{d}}\big)^2\mathrm{e}^{-t}s^{(4)}q_t^{-2}\,.$$

- Non-trivial technical estimates are required to assure that the expectations of certain linear combinations of the random variables are smaller than their naive sizes.

Thank you for your attention, and the hospitality!