08/09/2025

Part I : Equivalence for nonlinear random matrices

Data vectors $x_1\ x_2 \cdots x_n \overset{iid}{\sim} P(x)$ $\qquad x_i \in \mathbb{R}^d$

$E\, x_i = 0 \qquad E\, x_i x_i^T = I$

Inner product kernel matrix $A \in \mathbb{R}^{n \times n}$

$$A_{ij} = \begin{cases} \frac{1}{\sqrt{n}}\, \sigma\left(\frac{x_i^T x_j}{\sqrt{d}}\right) & i \neq j \\ 0 & i = j. \end{cases} \qquad \sigma : \mathbb{R} \to \mathbb{R}$$

Goal: understand the spectrum of A as $n, d \to \infty$.

Special case: $x_i \sim \mathrm{Unif}\{\pm 1\}^{\otimes d}$

and $\sigma(z) = \frac{z^2 - 1}{\sqrt{2}}$

$$\sqrt{2n}\, A_{ij} = \frac{(x_i^T x_j)^2}{d} - 1 = \frac{(\sum x_{ia} x_{ja})^2}{d}$$

$$= \frac{\sum_{ab} x_{ia} x_{ib}\ x_{ja} x_{jb}}{d} - 1$$

$$= 2 \sum_{a<b} x_{ia} x_{ib}\ x_{ja} x_{jb}$$

$$\Rightarrow A_{ij} = \sqrt{\tfrac{2}{n}} \langle f(x_i), f(x_j) \rangle$$

$f(x_i) \in \mathbb{R}^{\frac{d^2-d}{2}}$ $\qquad f(x) = (x_1 x_2, x_1 x_3 \cdots x_1 x_d, x_2 x_3 \cdots)$

$= (x_{ab})_{a<b}$ $\qquad f_u$ $\qquad M_2 = \{(i,j) : 1 \le i < j < d\}$

Idea: Linearization by lifting to the feature space.

$$A_{ij} = \begin{cases} \frac{\sqrt{2}}{\sqrt{n}\, d} \langle f(x_i), f(x_j) \rangle & i \neq j \\ 0 & i = j \end{cases}$$

$$A = \frac{1}{\sqrt{n}\sqrt{d^2/2}} \; \begin{array}{c} \frac{d^2-d}{2} \\ n \begin{bmatrix} f(x_1)^T \\ \vdots \\ f(x_n)^T \end{bmatrix} \end{array} \begin{bmatrix} f(x_1) & \cdots & f(x_n) \end{bmatrix} n \; - \operatorname{diag}(*)$$

Easy to show : $Ef = 0 \qquad E f f^T = I$

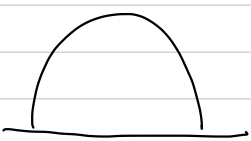
Universality: Replace f by $N(0, I_{\frac{d(d-1)}{2}})$

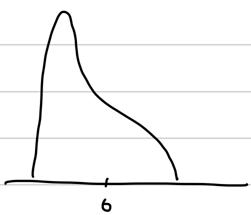
Then we have

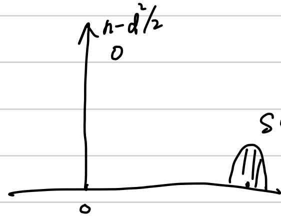
$$B = \frac{1}{\sqrt{n}\sqrt{d^2/2}} \; \overset{d^2/2}{\boxed{G^T}} \; \overset{n}{\boxed{G}} \; - \operatorname{diag}\{*\}$$

$n \propto d$ | $n \propto d^2$ | $n \propto d^3$

semicircle





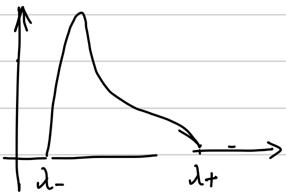


Idea: Linearization + Universality $\Rightarrow$ MP law

Scaling $n \propto d^\ell$ SC, MP or low-rank

Side note: Wishart ensemble and the MP law

$$A = \frac{1}{\sqrt{np}} \; \underset{n \times p}{\boxed{G^T}} \; \underset{p \times n}{\boxed{G}} \; - \; \mathrm{diag}\{*\}$$

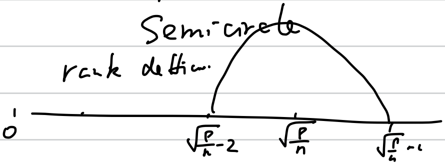


$\sqrt{\frac{n}{p}} + \sqrt{\frac{p}{n}} - 2 \qquad = \sqrt{\frac{n}{p}} + \sqrt{\frac{p}{n}} + 2$

Case I $\quad p \propto n \quad$ spectrum $O(1)$

Case II $\quad p \gg n \quad$ Semicircle

Case III $\quad p \ll n \quad$ rank deficien.



Question: General nonlinearity $\sigma(x)$

Approximation, $\sigma(x)$ is a finite-degree polynomial

$$\sigma(x) = \sum c_k x^k$$

$$\left(\sum X_{ia} X_{ja}\right)^3 = \sum_{abc} X_{ia} X_{ib} X_{ic} X_{ja} X_{jb} X_{jc} \cdots$$

$X_i \overset{iid}{\sim} \mathrm{Unif}(S^{d-1} \cdot \sqrt{d})$ [L, H.T. Yau, 2022]

$$\sigma(x) = \sum \mu_k g_k^d(x)$$

↑ Gegenbauer polynomials

$\left\{ 1,\ x,\ \frac{x^2-1}{\sqrt{2}} \sqrt{\frac{d+2}{d-1}},\ \ldots \right\}$

Very close to Hermite polynomial

$$g_k(x_i^T x_j) = \frac{1}{\sqrt{N_k}} f_k^T(x_i)\, f_k(x_j)$$

↑ spherical harmonics.

$N_k = O(d^k)$

$E f_k(x) = 0$ $\quad E f_k(x) f_k^T(x) = I_{N_k}.$ $\quad$ also $E f_k(x) f_\ell^T(x) = 0$

Idea: Replace $f_k(x)$ by $N(0, I_{d_k})$

$$A = \sum_{k=0}^{K} \mu_k \frac{1}{\sqrt{n}\sqrt{N_k}} \boxed{}\,\boxed{} - \mathrm{diag}\{*\}.$$

$$B = \sum_{k=0}^{K} \mu_k \frac{1}{\sqrt{n}\sqrt{N_k}} \boxed{G_k^T}\,\boxed{G_k} - \mathrm{diag}\{*\}$$

$$= \underbrace{\frac{\mu_0}{\sqrt{n}} 1}_{\text{spike}} + \underbrace{\mu_1 \frac{1}{\sqrt{nN_1}} \boxed{}\,\boxed{}}_{\text{spike } d \text{ e-vals}} + \underbrace{\mu_2 \frac{1}{\sqrt{nN_2}} \boxed{}\,\boxed{}}_{\text{MP}} + \underbrace{\frac{\mu_3}{\sqrt{nN_3}} \boxed{}\,\boxed{}}_{\text{SC}}$$

Gegenbaur polynomials and spherical harmonics are elegant but too special $\quad X \sim \mathrm{Unif}(\sqrt{d} \cdot S^{d-1})$

$X \overset{iid}{\sim} P(X) \qquad X_{i1}\ X_{i2} \cdots X_{id}$ independent

$E X_{ia} = 0 \quad E X_{ia}^2 = 1, \quad E X_{ia}^{2p} \le C_p < \infty \quad \forall p \in \mathbb{N}$

[Dubova, L. McKenna, Yau, 2023]

Write $\sigma(z) = \sum \mu_k h_k(x)$ ← normalized Hermite polynomials

$$\left\{ 1,\ x,\ \frac{x^2-1}{\sqrt{2}},\ \frac{x^3-3x}{\sqrt{6}},\ \frac{x^4-6x^2+3}{\sqrt{24}}\ \cdots \right\}$$

$$A = \sum \mu_k H_k$$

$$(H_k)_{ij} = \begin{cases} \frac{1}{\sqrt{n}} h_k\left(\frac{x_i^T x_j}{\sqrt{d}}\right) & i \neq j \\ 0 & i = j. \end{cases}$$

$f_k(x) \in \mathbb{R}^{N_k = \binom{d}{k}}$

$$\left.\begin{aligned} f_0(x) &= 1 \\ f_1(x) &= x \\ f_2(x) &= (X_i X_j)_{i<j} \\ f_3(x) &= (X_{i_1} X_{i_2} X_{i_3})_{i_1<i_2<i_3} \\ &\ \vdots \end{aligned}\right\}$$

multilinear polynomial

Polynomial chaos

Walsh

$$h_k\left(\frac{x_i^T x_j}{\sqrt{d}}\right) = \frac{\sqrt{k!}}{d^{k/2}} \langle f_k(x_i), f_k(x_j)\rangle + O_<\left(\frac{1}{\sqrt{d}}\right)$$

Note: Ignore $O_<(\frac{1}{\sqrt{d}})$ will not change the global spectrum

* How to prove this?

Key ingredient:

$b_k \in \mathbb{R}^{N_k}$ deterministic

What's the typical size of $\langle b_k, f_k(x) \rangle$

Note: Cauchy - Schwartz

$$|\langle b_k, f_k(x) \rangle| \le \|b_k\| \cdot \|f_k(x)\| \le \|b_k\| \, \mathcal{O}(\sqrt{N_k})$$

But if it were truly gaussian

$\langle b_k, f_k \rangle \sim \mathcal{N}(0, \|b_k\|^2)$

Rotational-invariance

So $\langle b_k, f_k \rangle = O_p(\|b_k\|)$

Key ingredient

$$\langle b_k, f_k \rangle = O_p(\|b_k\|)$$

In fact: $\left\| \frac{\langle b_k, f_k \rangle}{\|b_k\|} \right\|_{L^p} \le C_p < \infty$

Note: If $x \sim \mathcal{N}(0, I_d)$, easy consequence of hypercontractivity

Consequence: concentration of quadratic forms

$$\frac{1}{N_k} f_k^T(x) \, F \, f_k(x) = \frac{1}{N_k} \operatorname{tr} F + O_<\left(\frac{1}{\sqrt{d}}\right)$$

Related model

(1) Assaly & Benigni '25

$X \in \mathbb{R}^{d\times n}$ $\quad Y \in \mathbb{R}^{d\times n}$ $\quad n = \alpha d^2$

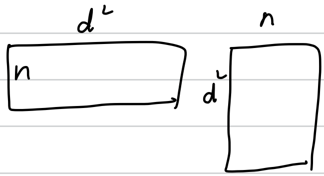
$$A = \frac{1}{d}(X^T X) \odot \frac{1}{d}(Y^T Y)$$

$$A_{ij} = \frac{1}{d^2}\left(\sum_a X_{ia} X_{ja}\right)\left(\sum_b Y_{ib} Y_{jb}\right)$$

$$= \frac{1}{d^2} \sum_{ab} X_{ia} X_{ib} \; Y_{ia} Y_{ib}$$

$$A_{ij} = \frac{1}{d^2} \langle f_i, f_j \rangle$$

$f_i = (X_{ia} Y_{ib})_{a,b}$ a subset of second-order polynomial chaos of $\begin{pmatrix} X_i \\ Y_i \end{pmatrix}$

$= \frac{1}{d^2}$ 

(2) NTK kernel [Benigni & Paquette '24+]

$$K = K_1 + K_2$$

$$K_1 = (X^T X) \odot \left(\sigma'(X^T W^T) \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_p \end{pmatrix} \sigma'(WX)\right)$$

$$K_2 = \sigma(X^T W)\, \sigma(WX)$$

$p/d \to \alpha$

$n/d^2 \to \beta$

Simplify: K_1 $\quad \sigma(x) = x^2/2$

$$K_1 = (X^T \cdot X) \odot (X^T \underbrace{W^T \overset{a}{\diagdown} W}_{= U \Lambda U^T} X)$$

$$\overset{(d)}{=} (X^T X) \odot (X^T \Lambda X)$$

$$(K_1)_{ij} = \Big(\sum_a X_{ia} X_{ja}\Big)\Big(\sum_b X_{ib} X_{jb} \lambda_b\Big)$$

$$= \sum_{ab} X_{ia} X_{ib} \lambda_b X_{ja} X_{jb}$$

$$= \sum_a X_{ia}^2 \lambda_a X_{ja}^2 + 2 \sum_{a<b} X_{ia} X_{ib} \lambda_b X_{ja} X_{jb}.$$

$$2 \quad f_i^T \begin{pmatrix} \diagdown \end{pmatrix} f_j$$ (conjecture)

Thought: approximate rotational invariance

$$\langle b_k, f_k(x) \rangle = O_<(\|b_k\|)$$

Do we have CLT?

Assume $x \sim N(0, I_d)$

$f_0(x) = 1$

$f_1 = x \in \mathbb{R}^d$

$f_2 = (x_i x_j)_{i<j} + (h_2(x_i))_i$

$f_3 = (x_i x_j x_k)_{i<j<k} + (h_3(x_i))_i + (h_2(x_i) x_j)_{i<j}$

$f_k \in \mathbb{R}^{N_k}$ $\qquad N_k = \binom{d+k-1}{k}$

$H_0 = 1$

$H_1 = x \in \mathbb{R}^d$

$H_2 = \frac{xx^T - I}{\sqrt{2}} \in (\mathbb{R}^d)^{\otimes 2}$

$H_k \in (\mathbb{R}^d)^{\otimes k}$

$b_k^T f_k = tr(B_k H_k)$

Ex: $k=2$ $\qquad tr\left(B \frac{xx^T - I}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}}(x^T B x - tr\, B)$

WLOG $\sqrt{d} B = diag\{\lambda_1 \cdots \lambda_d\}$

$$Z = \frac{\frac{1}{\sqrt{2}} \sum \lambda_i (x_i^2 - 1)}{\sqrt{d}} \qquad E Z^2 = \frac{\sum \lambda_i^2}{d} = 1$$

If λ_i bounded $\to$ CLT

If λ_1 λ_2 spikes

$$\underbrace{c \frac{x_1^2 - 1}{\sqrt{2}}}_{\chi^2} + \underbrace{\frac{1}{\sqrt{d}} \sum_{i \ge 2} \lambda_i h_2(x_i)}_{CLT}$$

A key component : CLT

$\theta^T \sigma(W x)$

$x \sim N(0, I_d)$ $\quad W \in \mathbb{R}^{p\times d}$ $\quad (WW^T)_{ii} = 1 \quad \forall i.$

$\theta \in \mathbb{R}^p$

$\sigma(z) = \sum_{k=1}^{K} \mu_k h_k(z)$

$$\theta^T \sigma(Wx) = \sum_{k=1}^{K} \mathrm{tr}(T_k H_k)$$

$$T_k = \sum_{i=1}^{p} \theta_i (w_i)^{\otimes k} \in (\mathbb{R}^d)^{\otimes k}$$

Malliavin Calculus
Normal Approximation with

General CLT for Wiener Chaos: [Nourdin & Peccati]

$Z_k = \mathrm{tr}(T_k H_k)$ $\quad \mathbb{E} Z_k^2 = \|T_k\|_F^2 = 1$

$Z_k \xrightarrow{law} N(0,1)$ iff $\mathbb{E} Z_k^4 \to 3$

$T_{k,r} \in \mathbb{R}^{d^r \times d^{k-r}}$

$\mathbb{E} Z_k^4 \to 3$ iff $\|T_{k,r} T_{k,r}^T\|_F^2 \to 0 \quad \forall 1 \le r \le k-1$

Joint work with Fan, Hu, Misiakiewicz, Wen. '25

$\theta^{*T} \sigma(Wk) \to N(0, \sigma^2)$ if

$$\left\| (WW^T)^{\odot r} \begin{pmatrix} \theta_1 & & \\ & \ddots & \\ & & \theta_p \end{pmatrix} (WW^T)^{\odot (k-r)} \begin{pmatrix} \theta_1 & & \\ & \ddots & \\ & & \theta_p \end{pmatrix} (WW^T)^{\odot r} \right\|_{op}$$

$\to 0$ for all $\underset{1 \le r < k}{k \le K}$

In linear scaling

$\| \cdot \|_{op} \leq \|W^T W\|_{op}^{k+r} \|\theta\|_\infty^2$

So CLT as long as $\|W^T W\|_{op}^{2k} \|\theta\|_\infty^2 \to 0$