Traffic Distribution of Variance Profiled Non-Linear Matrices

Issa-Mbenard Dabo, joint work with Camille Mâle

Université de Bordeaux
Institut de Mathématiques

October 2, 2024

Our framework

We are interested by the singular-value distribution of this non-linear matrix:

$$H = \frac{1}{\sqrt{n}} h\left(\left\{\frac{WX}{\sqrt{p}}\right\}\right) \in \mathbb{R}^{m*n},$$

with h a function applied entry-wise on WX.

We suppose that $W \in \mathbb{R}^{m\times p}$ and $X \in \mathbb{R}^{p\times n}$ be random **variance profiled matrices**, i.e

$$W = \Gamma_w \circ W' \quad \text{and} \quad X = \Gamma_x \circ X',$$

such that Γ_w and Γ_x are deterministic matrices; W' and X' entries are i.i.d centered with finite moments of any order.

- H can be interpreted as a **single-layer neural network** applied on the data matrix X.

- The **singular value** distribution of H is related the train and test risk of the random feature regression.

- **The presence of variance profile** makes the data **heterogeneous**.

- The use of variance profile can lead to study data coming from **mixtures models**, relevant in the study of real data.

Péché-Benigni (2021) studied the case with **constant variance profiles**, retrieving the limiting eigenvalue distribution of HH^* and proving that

Theorem

H has the same limiting singular-value distribution as H_{lin}, with :

$$H_{lin} = \alpha \frac{WX}{\sqrt{np}} + \beta \frac{Z}{\sqrt{n}} + \gamma \frac{J}{\sqrt{p}},$$

with α, β, γ constant values depending on h, Z a random matrix with i.i.d $\mathcal{N}(0,1)$ entries and J a deterministic rank 2 matrix.

Benigni, L., Péché, S. (2021). Eigenvalue distribution of some nonlinear models of random matrices. Electronic Journal of Probability, 26, 1-37.

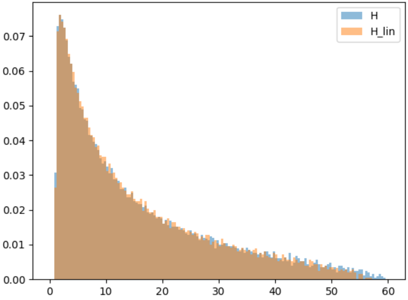

Figure: Comparison of the sigularvalue distributions of H and H_{lin} for constant profiles.

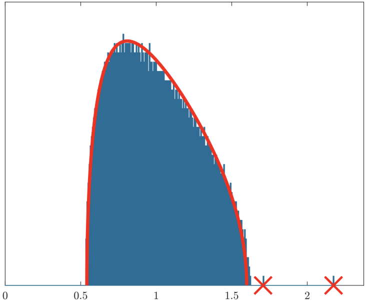

Figure: Appearance of outiliers in the spectrum of H.

Using a **variant of moment method**, we computed the **traffic distribution** of H, which encodes, among other things, the **eigenvalue distribution**.

Theorem

If h is odd then H has the same limiting singular eigenvalue distribution as H_{lin}, with

$$H_{lin} = \Theta_{WX} \circ \frac{WX}{\sqrt{np}} + \Theta_Z \circ \frac{Z}{\sqrt{n}} + \left(\Theta_J \circ \frac{J}{\sqrt{np}}\right),$$

with $\Theta_{WX}, \Theta_Z, \Theta_J$ deterministic matrices depending on h, Z a random matrix with i.i.d $\mathcal{N}(0,1)$ entries and J a deterministic rank 1 matrix.

D, I., Male, C. (2024). A traffic approach for profiled Pennington-Worah matrices. arXiv preprint arXiv:2409.13433.

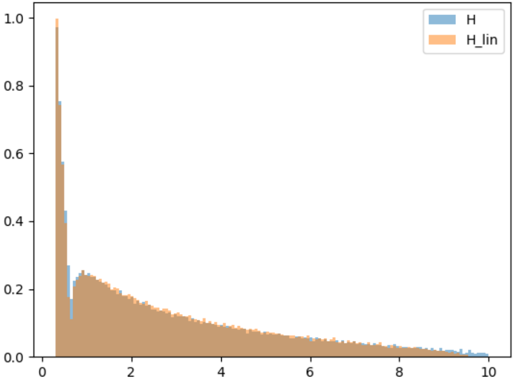

Figure: Comparison of the sigularvalue distributions of H and H_{lin}.

- This result led us to notice that H is **asymptotically bi-unitarily invariant (BUI)**:

$$H \sim U_n H V_n,$$

with U_n and V_n independent Haar unitary matrices.

- This result led us to notice that H is **asymptotically bi-unitarily invariant (BUI)**:

$$H \sim U_n H V_n,$$

with U_n and V_n independent Haar unitary matrices.

- This result allowed us to get a good **deterministic equivalent** of $Q(z) = (H^*H - zI_n)^{-1}$, denoted $Q^{\square}(z)$:

$$\left\| \Delta[Q(z)] - Q^{\square}(z) \right\| \to 0, \quad \text{a.s.}$$

Perspectives

- We are currently using these results in order to get deterministic equivalent for the train and test risk in the random feature regression context.
- Our results only hold for odd functions h, it would be interesting to consider more general functions.
- The rank one matrix, J, could imply the appearance of outliers in the spectrum of HH^*.

Thank you for your attention