Why do Random Matrices Explain Learning? An Argument of Universality Offered by GANs

Mohamed El Amine Seddik [1,2] Mohamed Tamaazousti [1] Romain Couillet [2,3]

[1]CEA List [2]CentraleSupélec [3]GIPSA-Lab University of GrenobleAlpes

Abstract

Context:

- Study of large **Gram** matrices for concentrated data.

Motivation:

- **Observation:** RMT predicts ML performances in high-dimension under Gaussian assumptions on data.
- BUT Real data are **unlikely close** to Gaussian vectors.
- Gaussian vectors fall within a **larger**, more **useful**, class of random vectors.

Results:

- GAN data [1] fall within the class of concentrated vectors.
- Only first and second order statistics of concentrated data matter to describe the behavior of **Gram** matrices.

Concentrated Vectors

Definition 1. Given a normed space $(E, \|\cdot\|_E)$ and $q \in \mathbb{R}$, a random vector $X \in E$ is q-*exponentially concentrated* if for any 1-**Lipschitz function** $\mathcal{F} : \mathbb{R}^p \to \mathbb{R}$, there exists $C, c > 0$ s.t.

$$\forall t > 0,\ \mathbb{P}\left\{|\mathcal{F}(X) - \mathbb{E}\mathcal{F}(X)| \geq t\right\} \leq C\, e^{-ct^q} \xrightarrow{\text{denoted}} \boxed{X \in \mathcal{O}(e^{-\cdot^q}) \text{ in } (E, \|\cdot\|_E)}$$

(P1) $X \sim \mathcal{N}(0, I_p)$ is 2-exponentially concentrated [2].

(P2) If $X \in \mathcal{O}(e^{-\cdot^q})$ and $\mathcal{G}$ is ℓ-**Lipschitz**, then $\mathcal{G}(X) \in \mathcal{O}(e^{-(\cdot/\ell)^q})$.

"Concentrated vectors are stable through Lipschitz maps."

Model & Assumptions

(A1) **Data matrix** (distributed in k classes $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_k$):

$$X = \Big[\underbrace{x_1, \ldots, x_{n_1}}_{\in \mathcal{O}(e^{-\cdot^{q_1}})}, \underbrace{x_{n_1+1}, \ldots, x_{n_2}}_{\in \mathcal{O}(e^{-\cdot^{q_2}})}, \ldots, \underbrace{x_{n-n_k+1}, \ldots, x_n}_{\in \mathcal{O}(e^{-\cdot^{q_k}})}\Big] \in \mathbb{R}^{p \times n}$$

Model statistics:

$$\mu_\ell = \mathbb{E}_{x_i \in \mathcal{C}_\ell}[x_i], \quad C_\ell = \mathbb{E}_{x_i \in \mathcal{C}_\ell}[x_i x_i^\intercal]$$

(A2) **Growth rate assumptions:** As $p \to \infty$,

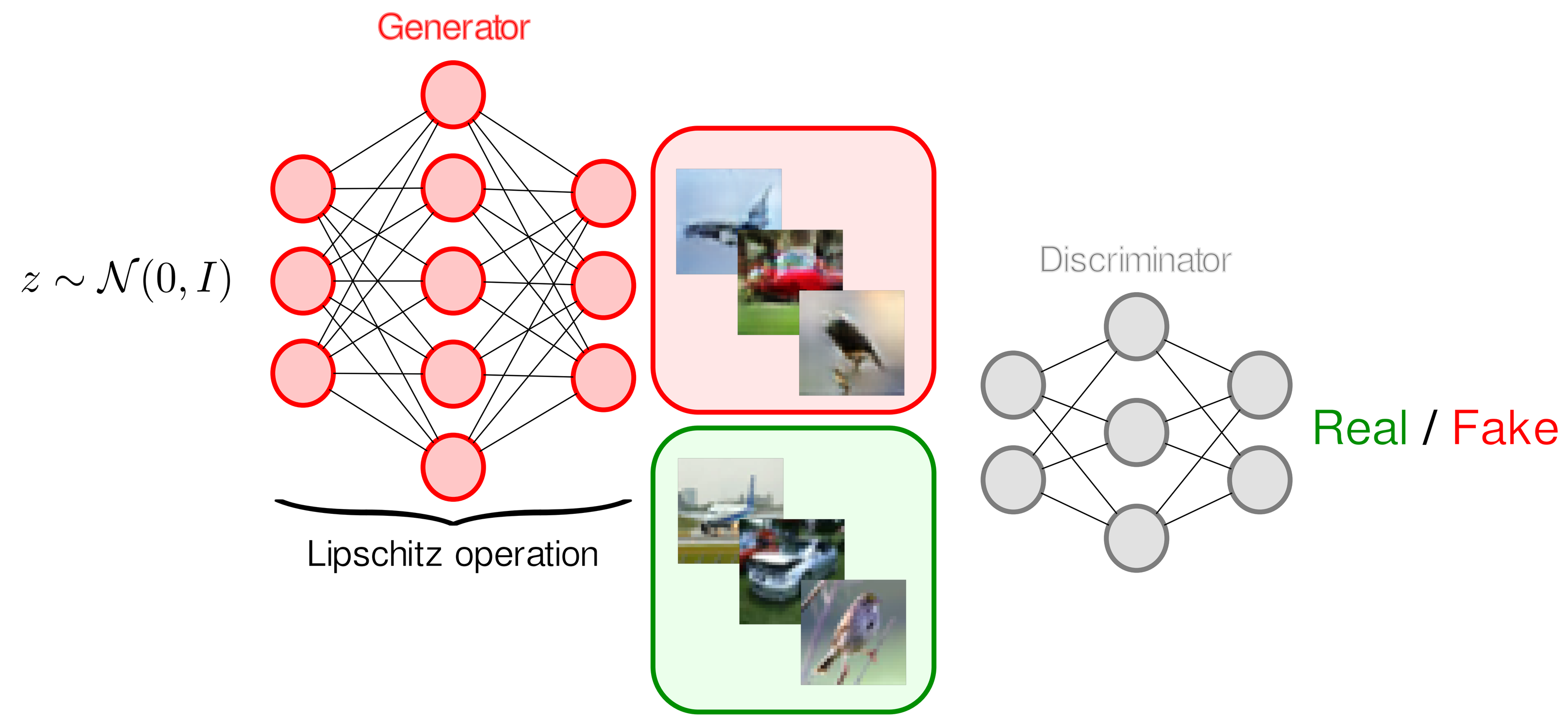
1. $p/n \to c \in (0, \infty)$.
2. The number of classers k is bounded.
3. For any $\ell \in [k]$, $\|\mu_\ell\| = \mathcal{O}(\sqrt{p})$.

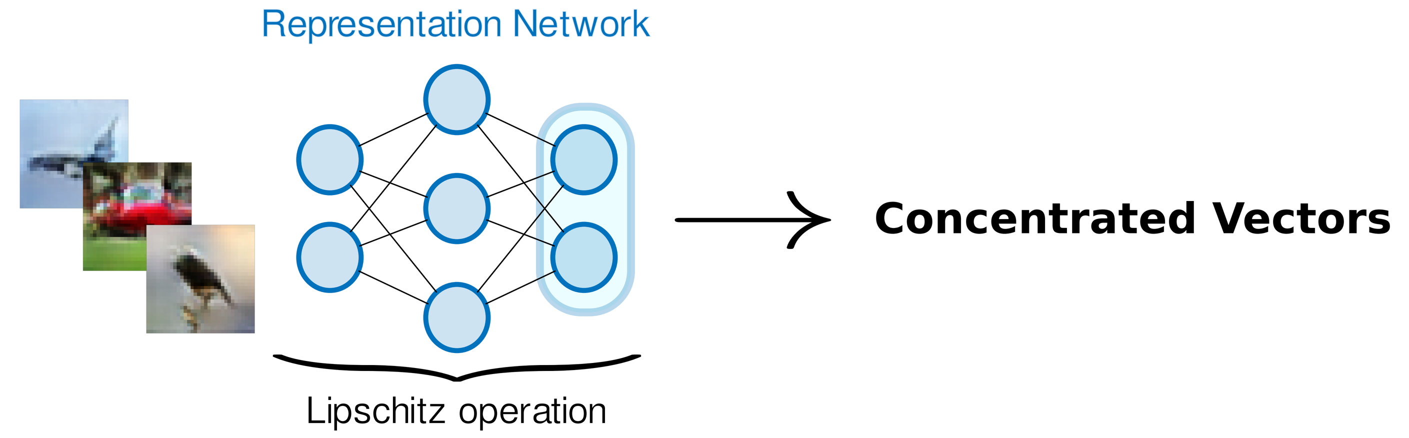
Gram matrix and its resolvent:

$$\boxed{G = \frac{1}{p} X^\intercal X, \ \ Q(z) = (G + zI_n)^{-1}}$$

$$m_L(z) = \frac{1}{n}\operatorname{tr}(Q(-z)),\ UU^\intercal \frac{1}{2\pi i}\oint_\gamma Q(-z)dz$$

Why Concentrated Vectors?





Deterministic Equivalent for $Q(z)$

Theorem Under the assumptions (A1) and (A2).
We have $Q(z) \in \mathcal{O}(e^{-(\sqrt{p}\,\cdot)^q})$ in $(\mathbb{R}^{n\times n}, \|\cdot\|)$. Furthermore,

$$\left\|\mathbb{E}[Q(z)] - \tilde{Q}(z)\right\| = \mathcal{O}\left(\sqrt{\frac{\log p}{p}}\right) \quad \text{where } \tilde{Q}(z) = \frac{1}{z}\Lambda(z) + \frac{1}{p\,z} J\Omega(z)J^\intercal$$

with $\Lambda(z) = \left\{\frac{1_{n_\ell}}{1+\delta_\ell(z)}\right\}_{\ell=1}^k$ and $\Omega(z) = \{\mu_\ell^\intercal \tilde{R}(z)\mu_\ell\}_{\ell=1}^k$

$$\tilde{R}(z) = \left(\frac{1}{k}\sum_{\ell=1}^k \frac{C_\ell}{1+\delta_\ell(z)} + zI_p\right)^{-1}$$

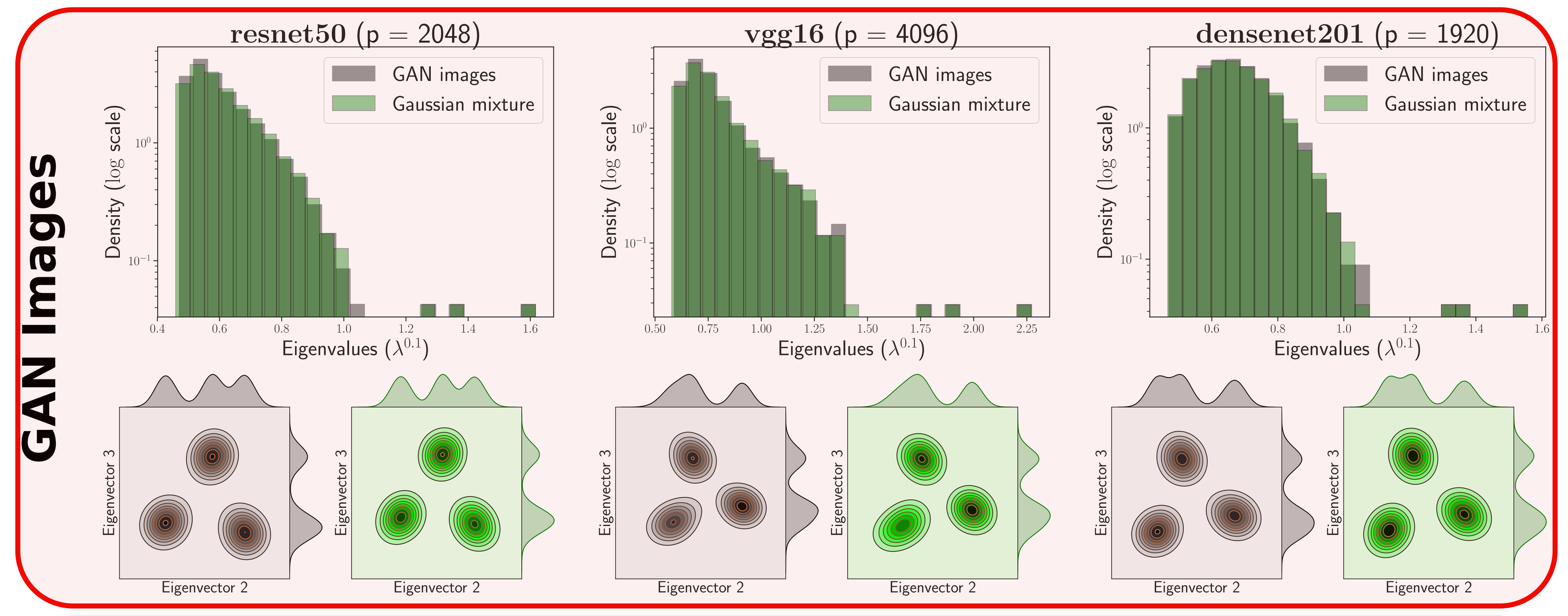
with $\delta(z) = [\delta_1(z), \ldots, \delta_k(z)]$ is the unique fixed point of the system of equations

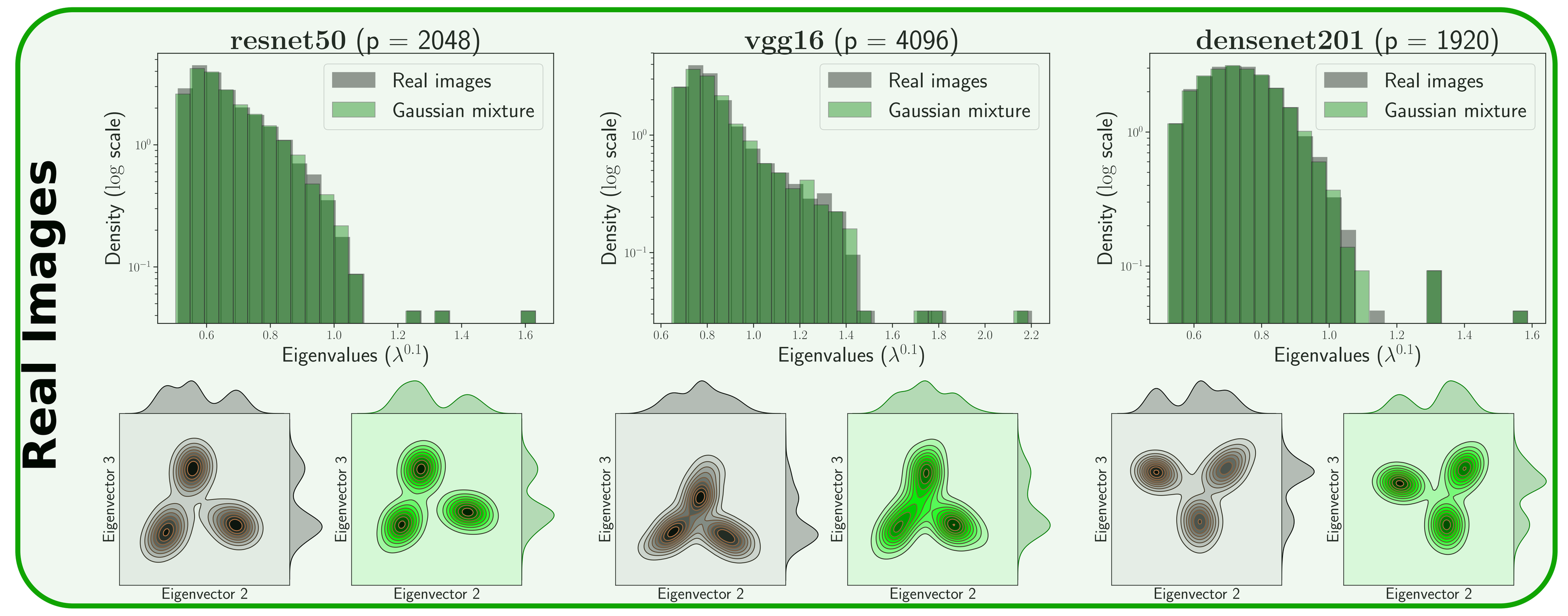
$$\delta_\ell(z) = \frac{1}{p}\operatorname{tr}\left(C_\ell\left(\frac{1}{k}\sum_{j=1}^k \frac{C_j}{1+\delta_j(z)} + zI_p\right)^{-1}\right) \quad \text{for each } \ell \in [k].$$

Key Observation: Only first and second order statistics matter!

Application to GAN-Generated Images

Figure 1. Images generated by the BigGAN model [3].





Perspectives

- **Generalize** to other ML tasks (Classification, Regression and TL).
- **Understand** and **improve** GANs by adding statistic constraints.

References

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets", in NIPS 2014.

[2] Terence Tao, "Topics in random matrix theory, volume 132". American Mathematical Society Providence, RI, 2012.

[3] Andrew Brock, Jeff Donahue, and Karen Simonyan, "Large scale GAN training for high fidelity image synthesis", in ICLR 2019.

DATAIA Days on Safety and AI 2019 - **Contact:** mohamedelamine.seddik@cea.fr