Kernel Random Matrices of Large Concentrated Data: The Example of GAN-Generated Images

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 **kernel** matrices for **concentrated** data.

Motivation:

- **GAN** data are **close** to **real** data [1].
- **GAN** data are **concentrated** data by **design**.

Results:

- **Universality** of spectral clustering w.r.t. the data distribution.
- **Real** data behave similar to **concentrated** vectors.
- RMT allows for the **theoretical understanding** of ML methods for **real** data.

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 real function** $\mathcal{F}$, 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 $\mathcal{G}(X) \in \mathcal{O}(e^{-(\cdot/\ell)^q})$.

Model & Assumptions

Data matrix (distributed in k classes):

$$\mathbf{X} = \Big[\underbrace{\mathbf{x}_1, \ldots, \mathbf{x}_{n_1}}_{\in \mathcal{O}(e^{-\cdot^{q_1}})}, \underbrace{\mathbf{x}_{n_1+1}, \ldots, \mathbf{x}_{n_2}}_{\in \mathcal{O}(e^{-\cdot^{q_2}})}, \ldots, \underbrace{\mathbf{x}_{n-n_k+1}, \ldots, \mathbf{x}_n}_{\in \mathcal{O}(e^{-\cdot^{q_k}})}\Big]$$

Model statistics:

$$\text{(Means) } \mathbf{m} = \sum_{\ell=1}^{k} \frac{n_\ell}{n} \mathbf{m}_\ell, \quad \bar{\mathbf{m}}_\ell = \mathbf{m} - \mathbf{m}_\ell$$

$$\text{(Covariances) } \mathbf{C} = \sum_{\ell=1}^{k} \frac{n_\ell}{n} \mathbf{C}_\ell, \quad \bar{\mathbf{C}}_\ell = \mathbf{C} - \mathbf{C}_\ell$$

(A1) Growth rate assumptions:

As $p \to \infty$,

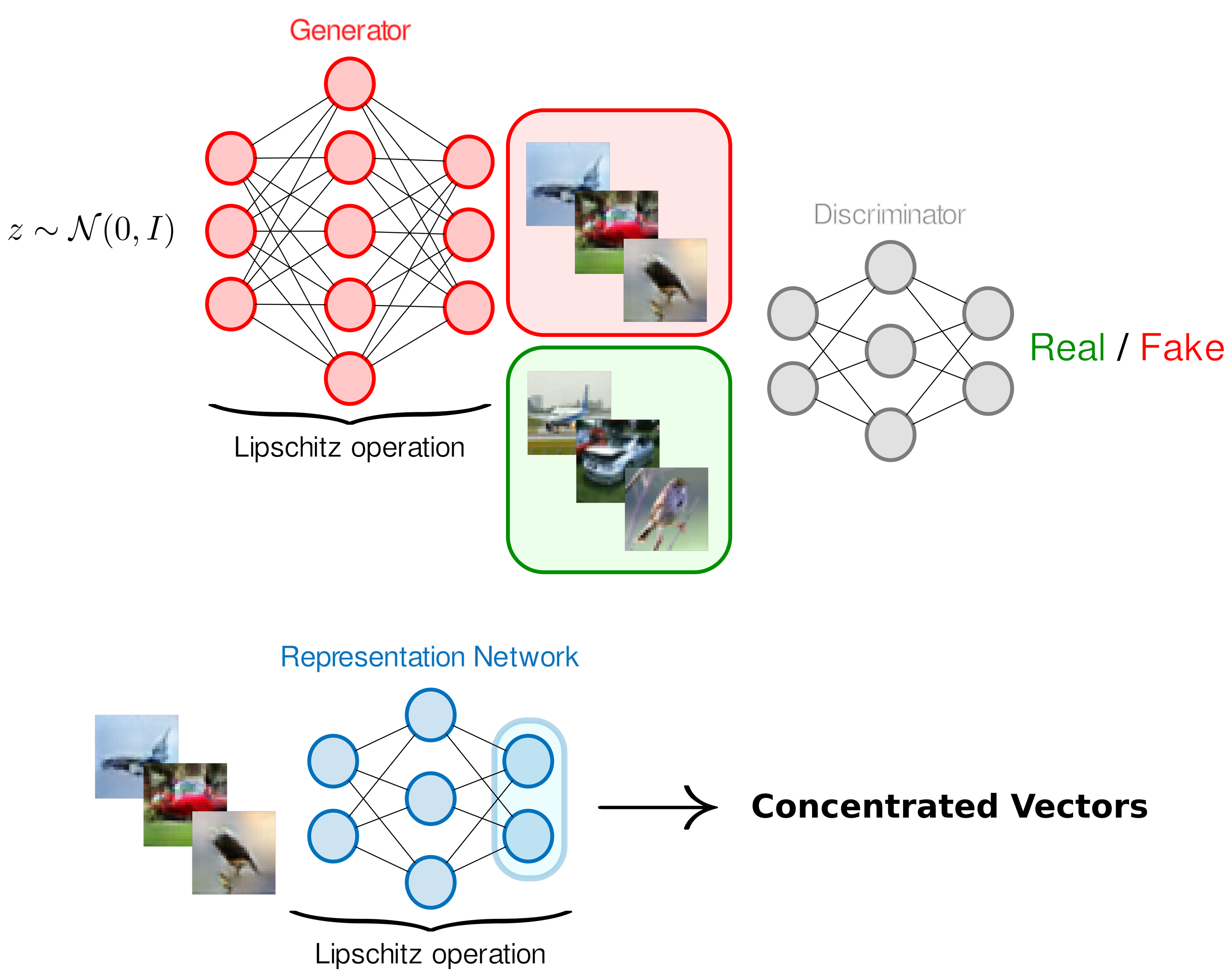
1. Data: $\frac{p}{n} \to c_0 \in (0, \infty)$, $\frac{n_\ell}{n} \to c_\ell \in (0, 1)$
2. Means: $\|\bar{\mathbf{m}}_\ell\| = \mathcal{O}(1)$, $\mathbb{E}\|\mathbf{x}_i\| = \mathcal{O}(\sqrt{p})$
3. Covariances: $\|\bar{\mathbf{C}}_\ell\| = \mathcal{O}(1)$, $\operatorname{tr} \bar{\mathbf{C}}_\ell = \mathcal{O}(\sqrt{p})$, $\operatorname{tr} \bar{\mathbf{C}}_a \bar{\mathbf{C}}_b = \mathcal{O}(\sqrt{p})$

(A2) **Kernel function:** Let $f : \mathbb{R}_+ \to \mathbb{R}_+$ three-times continuously **differentiable** function in $\tau \equiv \frac{2}{p} \operatorname{tr} \mathbf{C}$.

Kernel matrix:

$$\boxed{\mathbf{K} \equiv \left\{ f\left(\frac{1}{p}\|\mathbf{x}_i - \mathbf{x}_j\|^2\right) \right\}_{i,j=1}^{n}}$$

Why Concentrated Vectors?



Between and Within Class Vectors are "equidistant" in High-dimension

Proposition 1. Denote $\tau \equiv \frac{2}{p} \operatorname{tr} \mathbf{C}$. Under (A1), with probability $1 - \delta$

$$\max_{1 \leq i \neq j \leq n} \left\{ \left| \frac{1}{p}\|\mathbf{x}_i - \mathbf{x}_j\|^2 - \tau \right| \right\} = \mathcal{O}\left(\frac{\log\left(\frac{n}{\sqrt{\delta}}\right)^{1/q}}{\sqrt{n}} \right)$$

Random Matrix Equivalent for $\mathbf{K}$

Proposition 2. Under (A1) and (A2), Taylor expanding $\mathbf{K}$ entry-wise leads to

$$\mathbf{K} \propto \underbrace{\mathbf{J}\mathbf{A}\mathbf{J}^\intercal}_{\text{Information}} + \underbrace{f'(\tau)\mathbf{Z}^\intercal\mathbf{Z} + *}_{\text{Noise}}, \quad \mathbf{Z} = (\mathbf{X} - \mathbf{M}\mathbf{J}^\intercal)/\sqrt{p}$$

R1 $\mathbf{K}$ behaves as a **spiked** RMT model.

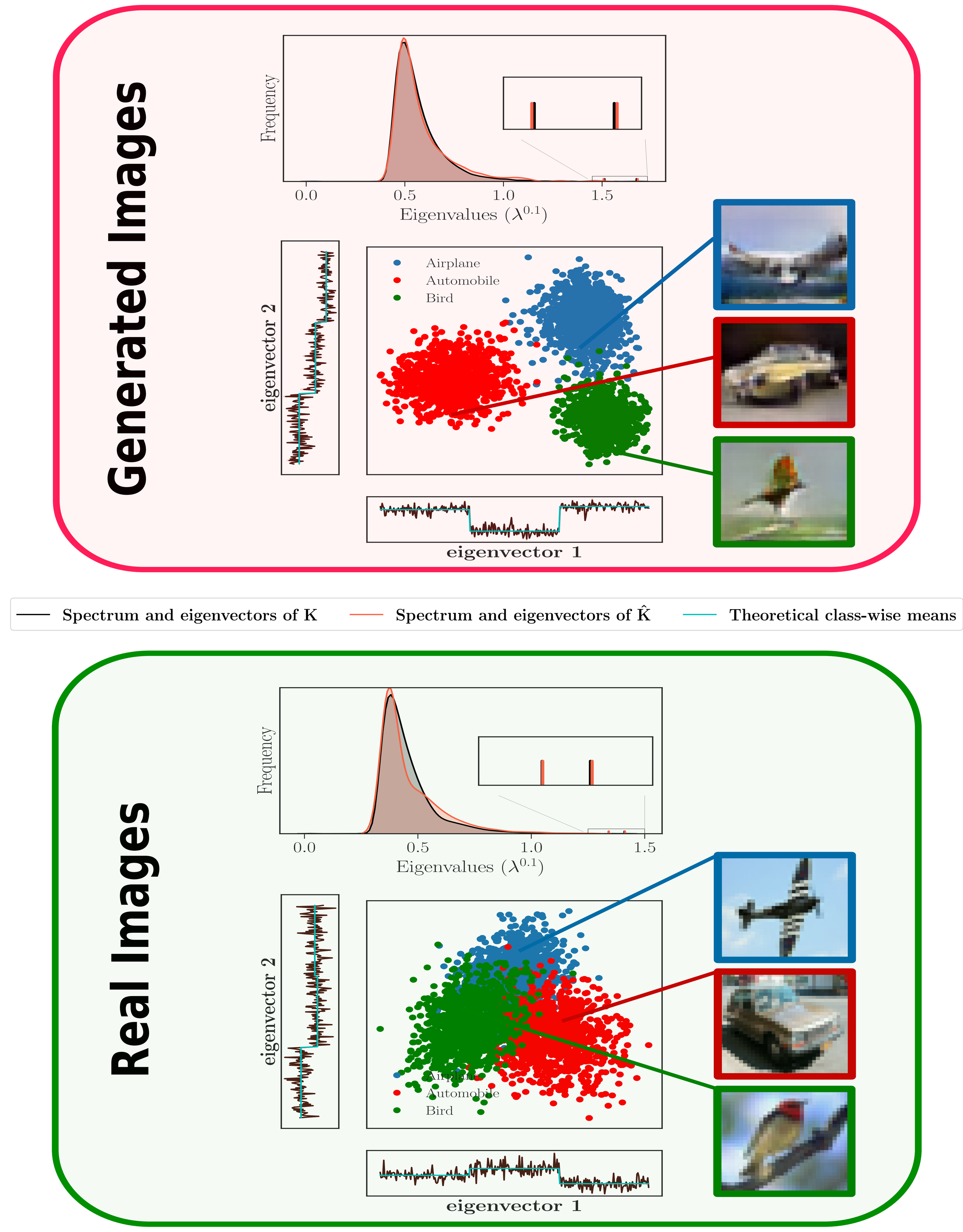
R2 The classification **performance** depends on $f'(\tau)$, $f''(\tau)$, $\mathbf{M}$, $\mathbf{t}$ and $\mathbf{T}$.

R3 No other **informative** statistics $\Rightarrow$ **universality** of spectral clustering.

$\mathbf{A}$ is a **low-rank** matrix depending only on $f'(\tau)$, $f''(\tau)$, $\mathbf{M}$, $\mathbf{t}$ and $\mathbf{T}$, where

$$\mathbf{J} = [\mathbf{j}_1, \ldots, \mathbf{j}_k], \quad \mathbf{M} = [\bar{\mathbf{m}}_1, \ldots, \bar{\mathbf{m}}_k], \quad \mathbf{t} = \left\{ \frac{\operatorname{tr} \bar{\mathbf{C}}_\ell}{\sqrt{p}} \right\}_{\ell=1}^{k}, \quad \mathbf{T} = \left\{ \frac{\operatorname{tr} \bar{\mathbf{C}}_a \bar{\mathbf{C}}_b}{p} \right\}_{a,b=1}^{k}$$

Application to GAN-Generated Images



Perspectives

- Prove a CLT under the concentration assumption.
- Generalize to other ML tasks (Classification, Regression)
- Apply to the dynamics of neural networks and GANs.

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 Advances in neural information processing systems, 2014.

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

ICASSP'19, Brighton, UK - **Contact:** mohamedelamine.seddik@cea.fr