Spectral clustering and random matrices

Florent BENAYCH-GEORGES (a)
joint work with Romain COUILLET (b)

(a) Université Paris Descartes
(b) CentraleSupélec

April 3, 2017
Lyon

Clustering

Clustering

Goal : cluster observations $x_1, \ldots, x_n$ with maximum similarity intra classes and minimum similarity inter classes

$$x_1 \quad , \ldots , \quad x_n \quad \Longrightarrow \mathcal{C}_1 = \{x_3, x_{18}, \ldots\}, \ldots, \mathcal{C}_k = \{x_1, x_{20}, \ldots\}$$

Clustering

Goal : cluster observations $x_1, \ldots, x_n$ with maximum similarity intra classes and minimum similarity inter classes

$$
\begin{array}{ccccc}
x_1 & , \ldots , & x_n & \Longrightarrow & \mathcal{C}_1 = \{x_3, x_{18}, \ldots\}, \ldots, \mathcal{C}_k = \{x_1, x_{20}, \ldots\} \\
\downarrow & & \downarrow & & \\
\begin{bmatrix} x_{1,1} \\ \vdots \\ x_{1,p} \end{bmatrix} & & \begin{bmatrix} x_{n,1} \\ \vdots \\ x_{n,p} \end{bmatrix} & \left.\vphantom{\begin{bmatrix} x \\ \vdots \\ x \end{bmatrix}}\right\} & p \text{ coordinates for each observation : } x_i \in \mathbb{R}^p
\end{array}
$$

Clustering

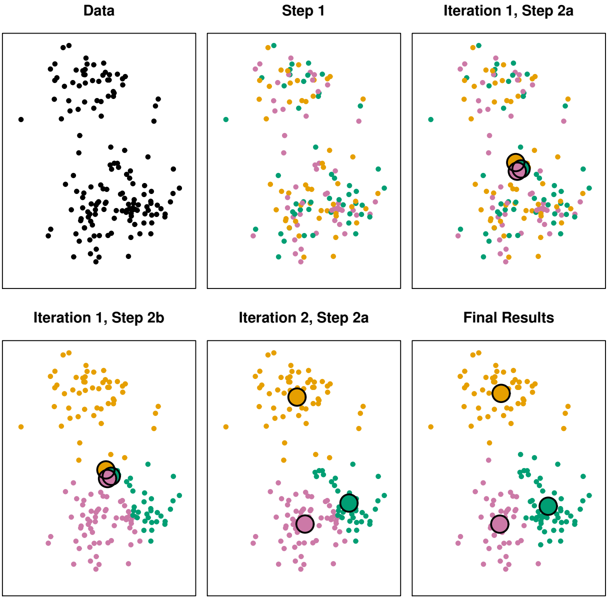
Goal : cluster observations $x_1, \dots, x_n$ with maximum similarity intra classes and minimum similarity inter classes

$$\begin{array}{ccccc} x_1 & , \dots , & x_n & \Longrightarrow & \mathcal{C}_1 = \{x_3, x_{18}, \dots\}, \dots, \mathcal{C}_k = \{x_1, x_{20}, \dots\} \\ \downarrow & & \downarrow & & \\ \begin{bmatrix} x_{1,1} \\ \vdots \\ x_{1,p} \end{bmatrix} & & \begin{bmatrix} x_{n,1} \\ \vdots \\ x_{n,p} \end{bmatrix} & \left.\vphantom{\begin{bmatrix} x \\ \vdots \\ x \end{bmatrix}}\right\} & p \text{ coordinates for each observation : } x_i \in \mathbb{R}^p \end{array}$$

Examples : k-means, EM, hierarchical clustering

Example : the k-means algorithm

Example : the k-means algorithm



Example : the k-means algorithm

Disadvantages :

Example : the k-means algorithm

Disadvantages :

- Not efficient in large dimension (when $p \gg 1$)

Example : the k-means algorithm

Disadvantages :

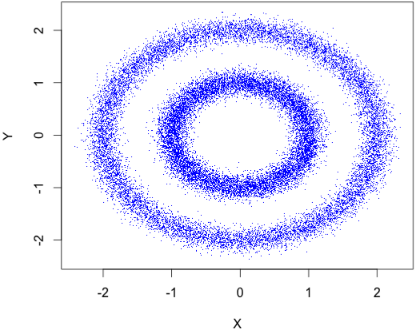
- Not efficient in large dimension (when $p \gg 1$)
- Even in low dimension : means are not always relevant :

Example : the k-means algorithm

Disadvantages :

- Not efficient in large dimension (when $p \gg 1$)

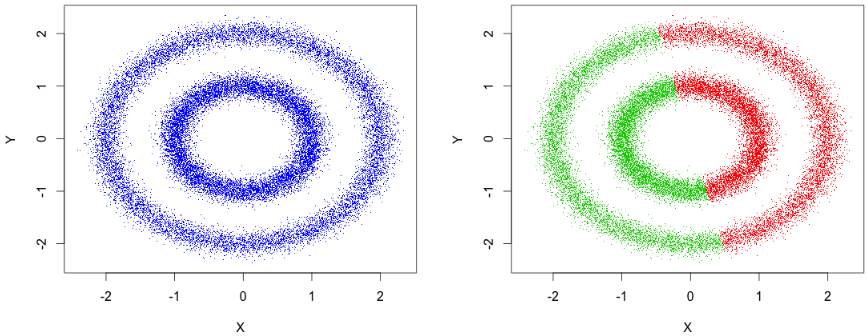
- Even in low dimension : means are not always relevant :



Example : the k-means algorithm

Disadvantages :

- Not efficient in large dimension (when $p \gg 1$)
- Even in low dimension : means are not always relevant :

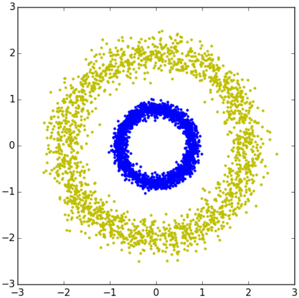


Spectral clustering : example and principle (1)

First **transform** the observations :

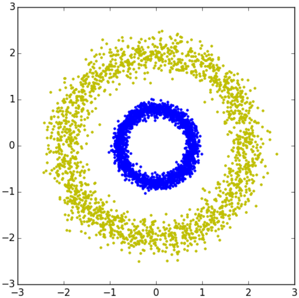
Spectral clustering : example and principle (1)

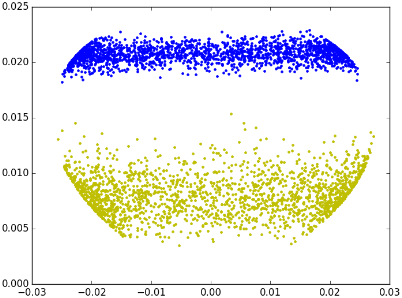
First **transform** the observations :



Spectral clustering : example and principle (1)

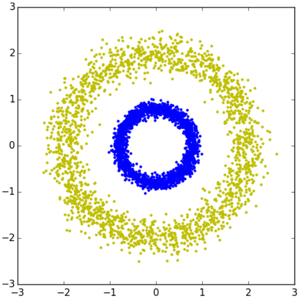
First **transform** the observations :

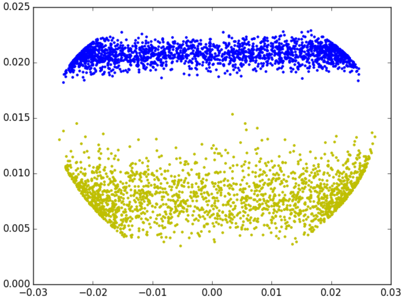




Spectral clustering : example and principle (1)

First **transform** the observations :

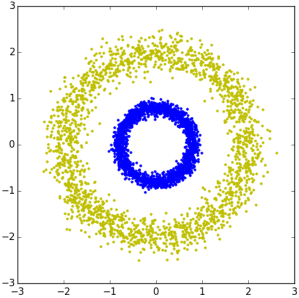


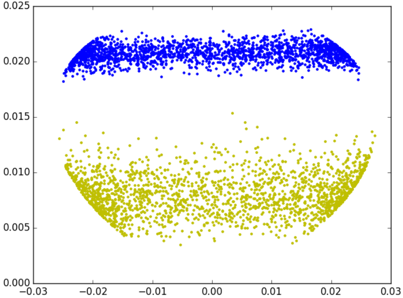


Left : $x_1, \ldots, x_n$

Spectral clustering : example and principle (1)

First **transform** the observations :

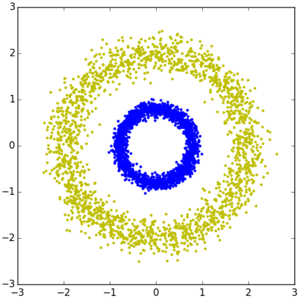


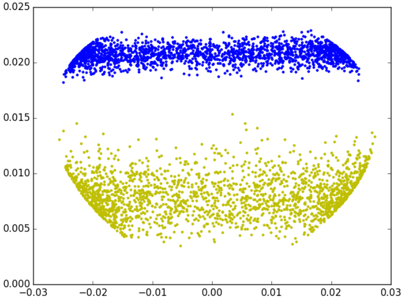


Left : $x_1, \ldots, x_n \implies K := \left[e^{-c\|x_j - x_i\|^2} \right]_{i,j=1}^{n}$

Spectral clustering : example and principle (1)

First **transform** the observations :

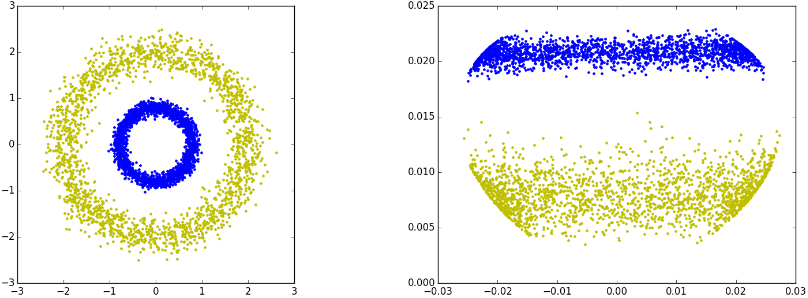




Left : $x_1, \ldots, x_n \implies K := \left[e^{-c\|x_j - x_i\|^2}\right]_{i,j=1}^{n}$ with **eigenvectors** $\vec{V}_1, \vec{V}_2, \cdots$

Spectral clustering : example and principle (1)

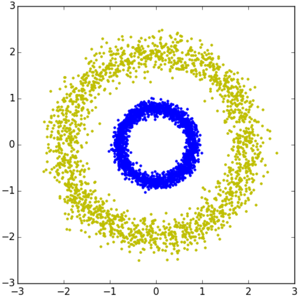
First **transform** the observations :

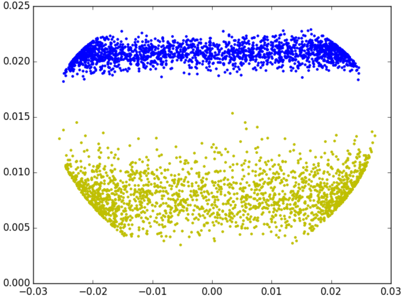


Left : $x_1, \ldots, x_n \implies K := \left[e^{-c\|x_j - x_i\|^2} \right]_{i,j=1}^{n}$ with **eigenvectors** $\vec{V}_1, \vec{V}_2, \cdots \implies$ right : $y_1, \ldots, y_n$ defined by :

Spectral clustering : example and principle (1)

First **transform** the observations :



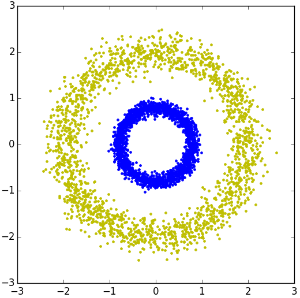


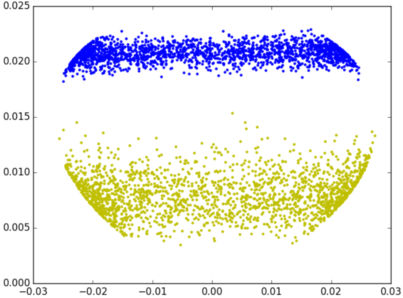
Left : $x_1, \ldots, x_n \implies K := \left[e^{-c\|x_j - x_i\|^2}\right]_{i,j=1}^{n}$ with **eigenvectors** $\vec{V}_1, \vec{V}_2, \cdots \implies$ right : $y_1, \ldots, y_n$ defined by :

$$\left[\vec{V}_2 \quad \vec{V}_1\right]$$

Spectral clustering : example and principle (1)

First **transform** the observations :

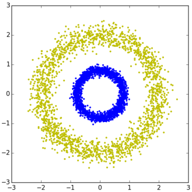




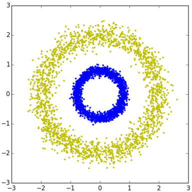
Left : $x_1, \ldots, x_n \implies K := \left[e^{-c\|x_j - x_i\|^2} \right]_{i,j=1}^{n}$ with **eigenvectors** $\vec{V}_1, \vec{V}_2, \cdots \implies$ right : $y_1, \ldots, y_n$ defined by :

$$\begin{bmatrix} \vec{V}_2 & \vec{V}_1 \end{bmatrix} =: \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

Spectral clustering : example and principle (1)

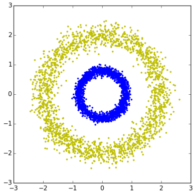


Spectral clustering : example and principle (1)



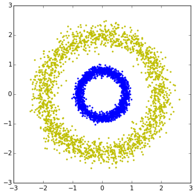
- $x_1, \ldots, x_{n_1}$: small circle

Spectral clustering : example and principle (1)



- $x_1, \ldots, x_{n_1}$: small circle
- $x_{n_1+1}, \ldots, x_{n_1+n_2}$: large circle

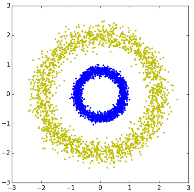
Spectral clustering : example and principle (1)



- $x_1, \ldots, x_{n_1}$: small circle
- $x_{n_1+1}, \ldots, x_{n_1+n_2}$: large circle

$\hookrightarrow$ Matrix $K := \left[e^{-c\|x_j - x_i\|^2} \right]_{i,j=1}^{n}$:

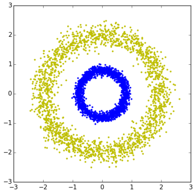
Spectral clustering : example and principle (1)



- $x_1, \ldots, x_{n_1}$: small circle
- $x_{n_1+1}, \ldots, x_{n_1+n_2}$: large circle

$\hookrightarrow$ Matrix $K := \left[e^{-c\|x_j - x_i\|^2} \right]_{i,j=1}^{n}$: $\qquad K = \begin{bmatrix} X & \varepsilon \\ \varepsilon & Y \end{bmatrix}$

Spectral clustering : example and principle (1)

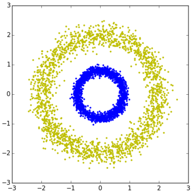


- $x_1, \ldots, x_{n_1}$: small circle
- $x_{n_1+1}, \ldots, x_{n_1+n_2}$: large circle

$\hookrightarrow$ Matrix $K := \left[e^{-c\|x_j - x_i\|^2} \right]_{i,j=1}^{n}$:

$$K = \begin{bmatrix} X & \varepsilon \\ \varepsilon & Y \end{bmatrix} \approx \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix}$$

Spectral clustering : example and principle (1)

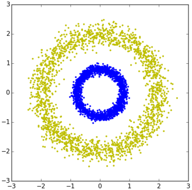


- $x_1, \ldots, x_{n_1}$: small circle
- $x_{n_1+1}, \ldots, x_{n_1+n_2}$: large circle

$\hookrightarrow$ Matrix $K := \left[e^{-c\|x_j - x_i\|^2} \right]_{i,j=1}^{n}$: $\quad K = \begin{bmatrix} X & \varepsilon \\ \varepsilon & Y \end{bmatrix} \approx \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix}$

$\hookrightarrow$ Maximal eigenvectors : $\approx$ either supported by $\vec{e}_1, \ldots, \vec{e}_{n_1}$ or by $\vec{e}_{n_1+1}, \ldots, \vec{e}_{n_1+n_2}$

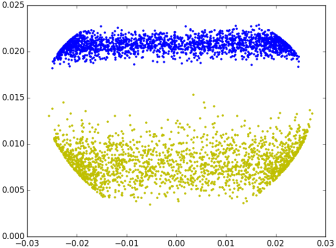
Spectral clustering : example and principle (1)



- $x_1, \ldots, x_{n_1}$: small circle
- $x_{n_1+1}, \ldots, x_{n_1+n_2}$: large circle

$\hookrightarrow$ Matrix $K := \left[e^{-c\|x_j - x_i\|^2} \right]_{i,j=1}^{n}$: $\quad K = \begin{bmatrix} X & \varepsilon \\ \varepsilon & Y \end{bmatrix} \approx \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix}$

$\hookrightarrow$ Maximal eigenvectors : $\approx$ either supported by $\vec{e}_1, \ldots, \vec{e}_{n_1}$ or by $\vec{e}_{n_1+1}, \ldots, \vec{e}_{n_1+n_2}$



Spectral clustering : principle (2)

$f(\|x_j - x_i\|) \geq 0$: **similarity** of x_i and x_j

(ex : $f(\|x_j - x_i\|) = \mathrm{e}^{-c\|x_j - x_i\|^2}$)

Spectral clustering : principle (2)

$f(\|x_j - x_i\|) \geq 0$: **similarity** of x_i and x_j

(ex : $f(\|x_j - x_i\|) = \mathrm{e}^{-c\|x_j - x_i\|^2}$)

$$L = \left[\frac{f(\|x_j - x_i\|)}{\sqrt{d_i d_j}}\right]_{i,j=1}^{n}, \quad d_i := \sum_k f(\|x_k - x_i\|)$$

Spectral clustering : principle (2)

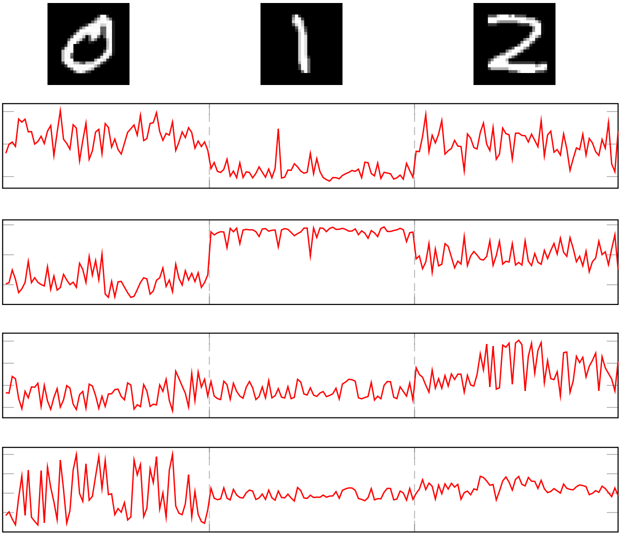
$f(\|x_j - x_i\|) \geq 0$: **similarity** of x_i and x_j

(ex : $f(\|x_j - x_i\|) = \mathrm{e}^{-c\|x_j - x_i\|^2}$)

$$L = \left[\frac{f(\|x_j - x_i\|)}{\sqrt{d_i d_j}}\right]_{i,j=1}^{n}, \quad d_i := \sum_k f(\|x_k - x_i\|)$$

Spectral clustering of $x_1, \ldots, x_n$ in k classes (2) :

L : symmetric Laplacian matrix. Replace observations $x_1, \ldots, x_n$ by the rows of the matrix of the largest eigenvectors of L and apply k-means on these (new) observations.



: Four leading eigenvectors of L for (partial) MNIST data ($n = 192$, $p = 784$, $k = 3$)

Model and Assumptions

Model and Assumptions

Gaussian mixture model :

Model and Assumptions

Gaussian mixture model :

- $x_1, \ldots, x_n \in \mathbb{R}^p$ independent

Model and Assumptions

Gaussian mixture model :

- $x_1, \ldots, x_n \in \mathbb{R}^p$ independent
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$

Model and Assumptions

Gaussian mixture model :

- $x_1, \ldots, x_n \in \mathbb{R}^p$ independent
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$
- $\mathcal{C}_a = \{i \mid x_i \sim \mathcal{N}(\mu_a, C_a)\}$,

Model and Assumptions

Gaussian mixture model :

- $x_1, \ldots, x_n \in \mathbb{R}^p$ independent
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$
- $\mathcal{C}_a = \{i \mid x_i \sim \mathcal{N}(\mu_a, C_a)\}, \quad \|C_a\| = O(1)$

Model and Assumptions

Gaussian mixture model :

- $x_1, \ldots, x_n \in \mathbb{R}^p$ independent
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$
- $\mathcal{C}_a = \{i \mid x_i \sim \mathcal{N}(\mu_a, C_a)\}, \quad \|C_a\| = O(1)$

Convergence rate : We have $n, p \gg 1$ and :

Model and Assumptions

Gaussian mixture model :

- $x_1, \ldots, x_n \in \mathbb{R}^p$ independent
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$
- $\mathcal{C}_a = \{i \mid x_i \sim \mathcal{N}(\mu_a, C_a)\}, \quad \|C_a\| = O(1)$

Convergence rate : We have $n, p \gg 1$ and :

1. **Data scaling** : $c_0 := \frac{p}{n}$ away from 0 and $+\infty$

Model and Assumptions

Gaussian mixture model :

- $x_1, \ldots, x_n \in \mathbb{R}^p$ independent
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$
- $\mathcal{C}_a = \{i \mid x_i \sim \mathcal{N}(\mu_a, C_a)\}, \quad \|C_a\| = O(1)$

Convergence rate : We have $n, p \gg 1$ and :

1. **Data scaling** : $c_0 := \frac{p}{n}$ away from 0 and $+\infty$
2. **Class scaling** : $c_a := \frac{\#\mathcal{C}_a}{n}$ away from 0 and 1

Model and Assumptions

Gaussian mixture model :

- $x_1, \ldots, x_n \in \mathbb{R}^p$ independent
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$
- $\mathcal{C}_a = \{i \mid x_i \sim \mathcal{N}(\mu_a, C_a)\}, \quad \|C_a\| = O(1)$

Convergence rate : We have $n, p \gg 1$ and :

1. **Data scaling** : $c_0 := \frac{p}{n}$ away from 0 and $+\infty$
2. **Class scaling** : $c_a := \frac{\#\mathcal{C}_a}{n}$ away from 0 and 1
3. **Mean and covariance scaling** :

Model and Assumptions

Gaussian mixture model :

- $x_1, \ldots, x_n \in \mathbb{R}^p$ independent
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$
- $\mathcal{C}_a = \{i \mid x_i \sim \mathcal{N}(\mu_a, C_a)\}, \quad \|C_a\| = O(1)$

Convergence rate : We have $n, p \gg 1$ and :

1. **Data scaling** : $c_0 := \frac{p}{n}$ away from 0 and $+\infty$
2. **Class scaling** : $c_a := \frac{\#\mathcal{C}_a}{n}$ away from 0 and 1
3. **Mean and covariance scaling** :
   Cases where **simple methods** are efficient :

Model and Assumptions

Gaussian mixture model :

- $x_1, \ldots, x_n \in \mathbb{R}^p$ independent
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$
- $\mathcal{C}_a = \{i \mid x_i \sim \mathcal{N}(\mu_a, C_a)\}, \quad \|C_a\| = O(1)$

Convergence rate : We have $n, p \gg 1$ and :

1. **Data scaling** : $c_0 := \frac{p}{n}$ away from 0 and $+\infty$
2. **Class scaling** : $c_a := \frac{\#\mathcal{C}_a}{n}$ away from 0 and 1
3. **Mean and covariance scaling** :
   Cases where **simple methods** are efficient :
   - $p \gg 1 \implies x_i - \mu_a = O(\sqrt{\operatorname{Tr} C_a}) = O(\sqrt{p})$ so

Model and Assumptions

Gaussian mixture model :

- $x_1, \dots, x_n \in \mathbb{R}^p$ independent
- k classes $\mathcal{C}_1, \dots, \mathcal{C}_k$
- $\mathcal{C}_a = \{i \mid x_i \sim \mathcal{N}(\mu_a, C_a)\}, \quad \|C_a\| = O(1)$

Convergence rate : We have $n, p \gg 1$ and :

1. **Data scaling** : $c_0 := \frac{p}{n}$ away from 0 and $+\infty$
2. **Class scaling** : $c_a := \frac{\#\mathcal{C}_a}{n}$ away from 0 and 1
3. **Mean and covariance scaling** :
   Cases where **simple methods** are efficient :
   - $p \gg 1 \implies x_i - \mu_a = O(\sqrt{\operatorname{Tr} C_a}) = O(\sqrt{p})$ so
     $\|\mu_a - \mu_b\| \gg \sqrt{p} \implies$

Model and Assumptions

Gaussian mixture model :

- $x_1, \ldots, x_n \in \mathbb{R}^p$ independent
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$
- $\mathcal{C}_a = \{i \mid x_i \sim \mathcal{N}(\mu_a, C_a)\}, \quad \|C_a\| = O(1)$

Convergence rate : We have $n, p \gg 1$ and :

1. **Data scaling** : $c_0 := \frac{p}{n}$ away from 0 and $+\infty$
2. **Class scaling** : $c_a := \frac{\#\mathcal{C}_a}{n}$ away from 0 and 1
3. **Mean and covariance scaling** :
   Cases where **simple methods** are efficient :
   - $p \gg 1 \implies x_i - \mu_a = O(\sqrt{\operatorname{Tr} C_a}) = O(\sqrt{p})$ so
     $\|\mu_a - \mu_b\| \gg \sqrt{p} \implies k$-means (possibly *well* projected) is efficient

Model and Assumptions

Gaussian mixture model :

- $x_1, \ldots, x_n \in \mathbb{R}^p$ independent
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$
- $\mathcal{C}_a = \{i \mid x_i \sim \mathcal{N}(\mu_a, C_a)\}$, $\quad \|C_a\| = O(1)$

Convergence rate : We have $n, p \gg 1$ and :

1. **Data scaling** : $c_0 := \frac{p}{n}$ away from 0 and $+\infty$
2. **Class scaling** : $c_a := \frac{\#\mathcal{C}_a}{n}$ away from 0 and 1
3. **Mean and covariance scaling** :
   Cases where **simple methods** are efficient :
   - $p \gg 1 \implies x_i - \mu_a = O(\sqrt{\operatorname{Tr} C_a}) = O(\sqrt{p})$ so
     $\|\mu_a - \mu_b\| \gg \sqrt{p} \implies$ k-means (possibly *well* projected) is efficient
   - $\|\mu_a\| \ll \sqrt{p}$ et $|\operatorname{Tr} C_a - \operatorname{Tr} C_b| \gg \sqrt{p} \quad \implies$

Model and Assumptions

Gaussian mixture model :

- $x_1, \ldots, x_n \in \mathbb{R}^p$ independent
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$
- $\mathcal{C}_a = \{i \mid x_i \sim \mathcal{N}(\mu_a, C_a)\}, \quad \|C_a\| = O(1)$

Convergence rate : We have $n, p \gg 1$ and :

1. **Data scaling** : $c_0 := \frac{p}{n}$ away from 0 and $+\infty$
2. **Class scaling** : $c_a := \frac{\#\mathcal{C}_a}{n}$ away from 0 and 1
3. **Mean and covariance scaling** :
   Cases where **simple methods** are efficient :
   - $p \gg 1 \implies x_i - \mu_a = O(\sqrt{\operatorname{Tr} C_a}) = O(\sqrt{p})$ so
     $\|\mu_a - \mu_b\| \gg \sqrt{p} \implies k$-means (possibly *well* projected) is efficient
   - $\|\mu_a\| \ll \sqrt{p}$ et $|\operatorname{Tr} C_a - \operatorname{Tr} C_b| \gg \sqrt{p} \implies k$-means on $\|x_i\|$ is efficient

Model and Assumptions

Gaussian mixture model :

- $x_1, \ldots, x_n \in \mathbb{R}^p$ independent
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$
- $\mathcal{C}_a = \{x \mid x \sim \mathcal{N}(\mu_a, C_a)\}, \quad \|C_a\| = O(1)$

Convergence rate : We have $n, p \gg 1$ and :

1. **Data scaling** : $c_0 := \frac{p}{n}$ away from 0 and $+\infty$
2. **Class scaling** : $c_a := \frac{\#\mathcal{C}_a}{n}$ away from 0 and 1
3. **Mean scaling** :

Model and Assumptions

Gaussian mixture model :

- $x_1, \ldots, x_n \in \mathbb{R}^p$ independent
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$
- $\mathcal{C}_a = \{x \mid x \sim \mathcal{N}(\mu_a, C_a)\}, \quad \|C_a\| = O(1)$

Convergence rate : We have $n, p \gg 1$ and :

1. **Data scaling** : $c_0 := \frac{p}{n}$ away from 0 and $+\infty$
2. **Class scaling** : $c_a := \frac{\#\mathcal{C}_a}{n}$ away from 0 and 1
3. **Mean scaling** : with $\sum_{a=1}^k c_a \mu_a = \vec{0}$,

Model and Assumptions

Gaussian mixture model :

- $x_1, \ldots, x_n \in \mathbb{R}^p$ independent
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$
- $\mathcal{C}_a = \{x \mid x \sim \mathcal{N}(\mu_a, C_a)\}, \quad \|C_a\| = O(1)$

Convergence rate : We have $n, p \gg 1$ and :

1. **Data scaling** : $c_0 := \frac{p}{n}$ away from 0 and $+\infty$
2. **Class scaling** : $c_a := \frac{\#\mathcal{C}_a}{n}$ away from 0 and 1
3. **Mean scaling** : with $\sum_{a=1}^{k} c_a \mu_a = \vec{0}$, $\|\mu_a\| = O(1)$

Model and Assumptions

Gaussian mixture model :

- $x_1, \ldots, x_n \in \mathbb{R}^p$ independent
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$
- $\mathcal{C}_a = \{x \mid x \sim \mathcal{N}(\mu_a, C_a)\}, \quad \|C_a\| = O(1)$

Convergence rate : We have $n, p \gg 1$ and :

1. **Data scaling** : $c_0 := \frac{p}{n}$ away from 0 and $+\infty$
2. **Class scaling** : $c_a := \frac{\#\mathcal{C}_a}{n}$ away from 0 and 1
3. **Mean scaling** : with $\sum_{a=1}^k c_a \mu_a = \vec{0}$, $\|\mu_a\| = O(1)$
4. **Covariance scaling** :

Model and Assumptions

Gaussian mixture model :

- $x_1, \ldots, x_n \in \mathbb{R}^p$ independent
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$
- $\mathcal{C}_a = \{x \mid x \sim \mathcal{N}(\mu_a, C_a)\}, \quad \|C_a\| = O(1)$

Convergence rate : We have $n, p \gg 1$ and :

1. **Data scaling** : $c_0 := \frac{p}{n}$ away from 0 and $+\infty$
2. **Class scaling** : $c_a := \frac{\#\mathcal{C}_a}{n}$ away from 0 and 1
3. **Mean scaling** : with $\sum_{a=1}^k c_a \mu_a = \vec{0}$, $\|\mu_a\| = O(1)$
4. **Covariance scaling** : with $C^\circ := \sum_{a=1}^k c_a C_a$ and $C_a^\circ := C_a - C^\circ$,

Model and Assumptions

Gaussian mixture model :

- $x_1, \ldots, x_n \in \mathbb{R}^p$ independent
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$
- $\mathcal{C}_a = \{x \mid x \sim \mathcal{N}(\mu_a, C_a)\}$, $\|C_a\| = O(1)$

Convergence rate : We have $n, p \gg 1$ and :

1. **Data scaling** : $c_0 := \frac{p}{n}$ away from 0 and $+\infty$
2. **Class scaling** : $c_a := \frac{\#\mathcal{C}_a}{n}$ away from 0 and 1
3. **Mean scaling** : with $\sum_{a=1}^k c_a \mu_a = \vec{0}$, $\|\mu_a\| = O(1)$
4. **Covariance scaling** : with $C^\circ := \sum_{a=1}^k c_a C_a$ and $C_a^\circ := C_a - C^\circ$, we have

$$\|C_a\| = O(1), \quad \operatorname{Tr} C_a^\circ = O(\sqrt{p})$$

Model and Assumptions

Gaussian mixture model :

- $x_1, \ldots, x_n \in \mathbb{R}^p$ independent
- k classes $\mathcal{C}_1, \ldots, \mathcal{C}_k$
- $\mathcal{C}_a = \{x \mid x \sim \mathcal{N}(\mu_a, C_a)\}, \quad \|C_a\| = O(1)$

Convergence rate : We have $n, p \gg 1$ and :

1. **Data scaling** : $c_0 := \frac{p}{n}$ away from 0 and $+\infty$
2. **Class scaling** : $c_a := \frac{\#\mathcal{C}_a}{n}$ away from 0 and 1
3. **Mean scaling** : with $\sum_{a=1}^k c_a \mu_a = \vec{0}$, $\|\mu_a\| = O(1)$
4. **Covariance scaling** : with $C^\circ := \sum_{a=1}^k c_a C_a$ and $C_a^\circ := C_a - C^\circ$, we have

$$\|C_a\| = O(1), \quad \operatorname{Tr} C_a^\circ = O(\sqrt{p})$$

Then $\frac{1}{p}\|x_j - x_i\|^2 \approx \tau := \frac{2}{p} \operatorname{Tr} C^\circ$

Matrix of interest

We study the normalized **Laplacian** matrix :

$$L = nD^{-\frac{1}{2}}KD^{-\frac{1}{2}}$$

with

$$K = \left[f\left(\frac{1}{p}\|x_j - x_i\|^2\right)\right]_{i,j=1}^{n},$$

$$D = \mathrm{diag}(d_i, 1 \leq i \leq n), \qquad d_i = \sum_j K_{ij}$$

Objectives

We want to derive, for each leading eigenvector $\vec{V}$ and each class $\mathcal{C}_a$ $(a = 1, \ldots, k)$:

Objectives

We want to derive, for each leading eigenvector $\vec{V}$ and each class $\mathcal{C}_a$ $(a = 1, \ldots, k)$:

- *Class-wise eigenvector means* :

$$\alpha_a(\vec{V}) := \frac{1}{\#\mathcal{C}_a} \left\langle \vec{V}, 1_{\mathcal{C}_a} \right\rangle$$

Objectives

We want to derive, for each leading eigenvector $\vec{V}$ and each class $\mathcal{C}_a$ $(a = 1, \ldots, k)$:

- *Class-wise eigenvector means* :

$$\alpha_a(\vec{V}) := \frac{1}{\#\mathcal{C}_a} \left\langle \vec{V}, 1_{\mathcal{C}_a} \right\rangle$$

- *Class-wise eigenvector fluctuations* :

$$\left\| \mathrm{diag}(1_{\mathcal{C}_a}) \left(\vec{V} - \alpha_a(\vec{V}) 1_{\mathcal{C}_a} \right) \right\|$$

Objectives

We want to derive, for each leading eigenvector $\vec{V}$ and each class $\mathcal{C}_a$ $(a = 1, \ldots, k)$:

- *Class-wise eigenvector means* :

$$\alpha_a(\vec{V}) := \frac{1}{\#\mathcal{C}_a} \left\langle \vec{V}, 1_{\mathcal{C}_a} \right\rangle$$

- *Class-wise eigenvector fluctuations* :

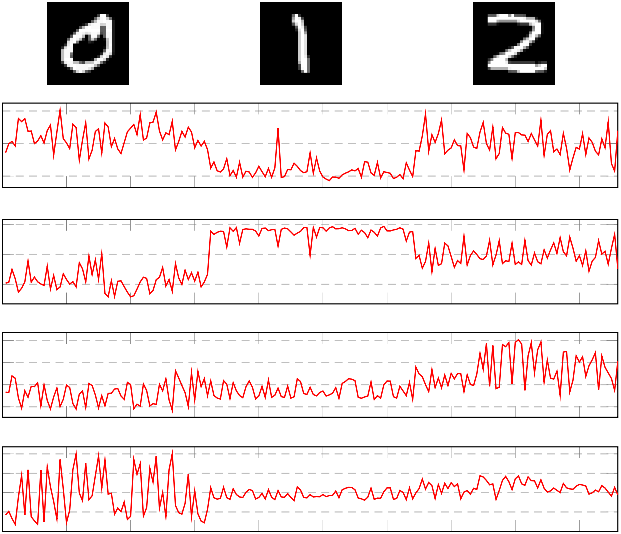
$$\left\| \mathrm{diag}(1_{\mathcal{C}_a}) \left(\vec{V} - \alpha_a(\vec{V}) 1_{\mathcal{C}_a} \right) \right\|$$

- *Class-wise cross correlations* :

$$\left\langle \left(\vec{V} - \alpha_a(\vec{V}) 1_{\mathcal{C}_a} \right), \mathrm{diag}(1_{\mathcal{C}_a}) \left(\vec{W} - \alpha_a(\vec{W}) 1_{\mathcal{C}_a} \right) \right\rangle$$

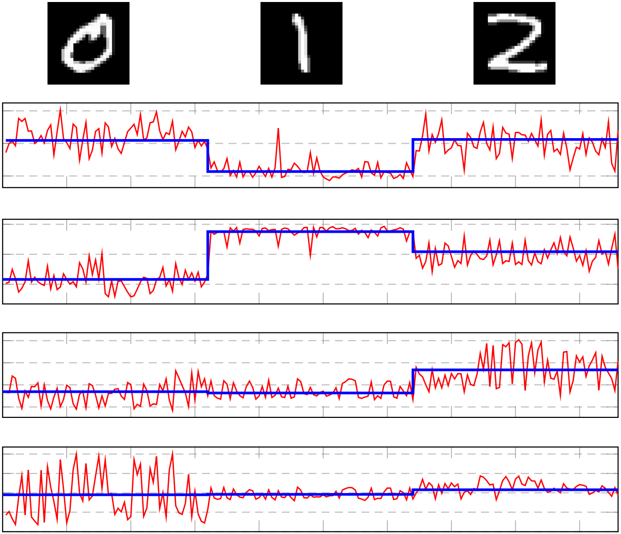
for $\vec{W}$ another leading eigenvector

Class-wise eigenvector means and fluctuations



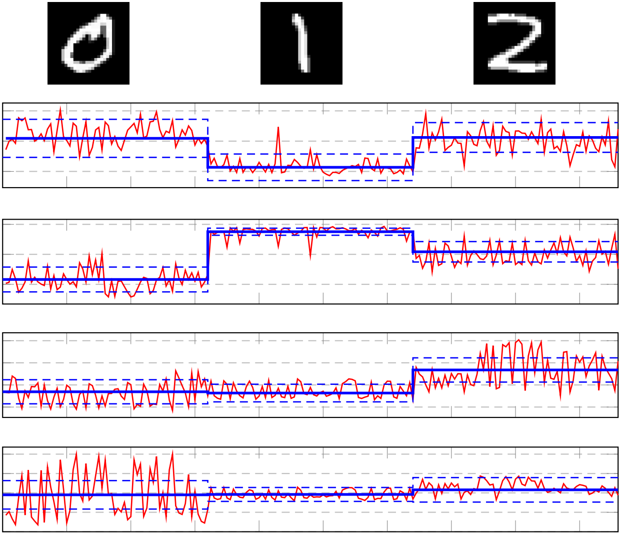
: MNIST data : four leading eigenvectors of L (**red**) and theoretical findings (**blue**).

Class-wise eigenvector means and fluctuations



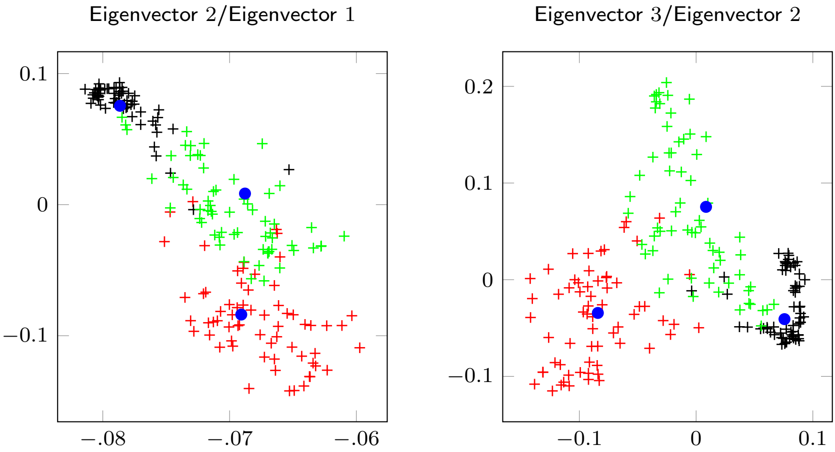
: MNIST data : four leading eigenvectors of L (**red**) and theoretical findings (**blue**).

Class-wise eigenvector means and fluctuations



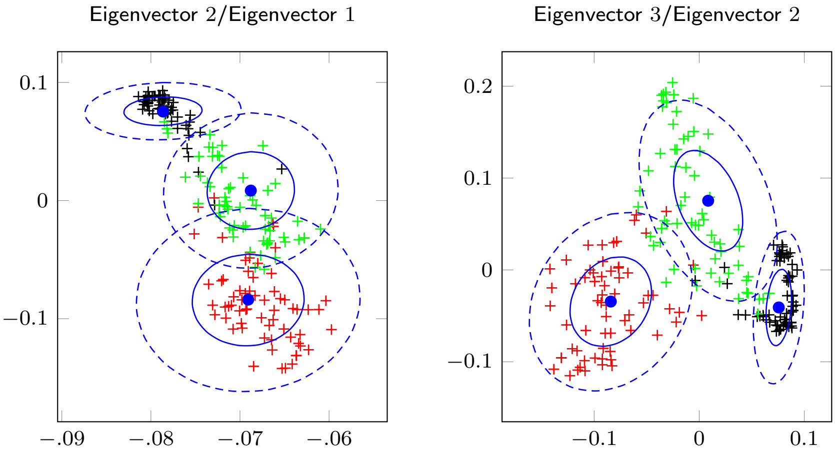
: MNIST data : four leading eigenvectors of L (**red**) and theoretical findings (**blue**).

Class-wise means, fluctuations and cross correlations



: 2D representation of eigenvectors of L, for the MNIST dataset. Theoretical means and 1- and 2-standard deviations in **blue**. Class 1 in **red**, Class 2 in **black**, Class 3 in **green**.

Class-wise means, fluctuations and cross correlations



: 2D representation of eigenvectors of L, for the MNIST dataset. Theoretical means and 1- and 2-standard deviations in **blue**. Class 1 in **red**, Class 2 in **black**, Class 3 in **green**.

Matrix of interest

We study the normalized **Laplacian** matrix :

$$L = nD^{-\frac{1}{2}} K D^{-\frac{1}{2}}$$

with

$$K = \left[f\left(\frac{1}{p} \|x_j - x_i\|^2 \right) \right]_{i,j=1}^{n},$$

$$D = \mathrm{diag}(d_i, 1 \leq i \leq n), \qquad d_i = \sum_j K_{ij}$$

Eigenvectors

Dominant Eigenvector : $(\sqrt{d_1}, \ldots, \sqrt{d_n})^\mathsf{T}$

Eigenvectors

Dominant Eigenvector : $(\sqrt{d_1}, \ldots, \sqrt{d_n})^\mathsf{T}$

Proposition (Eigenvector $D^{\frac{1}{2}} 1_n$)

We have

$$\frac{D^{\frac{1}{2}} 1_n}{\sqrt{1_n^\mathsf{T} D 1_n}} = \frac{1_n}{\sqrt{n}} + \frac{1}{n\sqrt{c_0}} \frac{f'(\tau)}{2f(\tau)} \left[\{t_a 1_{\mathcal{C}_a}\}_{a=1}^k + \operatorname{diag}\left\{ \sqrt{\frac{2}{p} \operatorname{Tr}(C_a^2)} 1_{\mathcal{C}_a} \right\}_{a=1}^k \varphi \right] + o(n^{-1})$$

with $t_a := \frac{1}{\sqrt{p}} \operatorname{Tr} C_a^\circ$ *(*$a = 1, \ldots, k$*) and* $\varphi \sim \mathcal{N}(0, I_n)$*.*

Eigenvectors

Dominant Eigenvector : $(\sqrt{d_1}, \ldots, \sqrt{d_n})^\mathsf{T}$

Proposition (Eigenvector $D^{\frac{1}{2}}1_n$)

We have

$$\frac{D^{\frac{1}{2}}1_n}{\sqrt{1_n^\mathsf{T} D 1_n}} = \frac{1_n}{\sqrt{n}} + \frac{1}{n\sqrt{c_0}} \frac{f'(\tau)}{2f(\tau)} \left[\{t_a 1_{\mathcal{C}_a}\}_{a=1}^k + \operatorname{diag}\left\{ \sqrt{\frac{2}{p}\operatorname{Tr}(C_a^2)} 1_{\mathcal{C}_a} \right\}_{a=1}^k \varphi \right] + o(n^{-1})$$

with $t_a := \frac{1}{\sqrt{p}} \operatorname{Tr} C_a^\circ$ *(*$a = 1, \ldots, k$*) and* $\varphi \sim \mathcal{N}(0, I_n)$.

To sum up :

- structure of $D^{\frac{1}{2}}1_n$: block-wise constant + noise

Eigenvectors

Dominant Eigenvector : $(\sqrt{d_1}, \ldots, \sqrt{d_n})^\mathsf{T}$

Proposition (Eigenvector $D^{\frac{1}{2}} 1_n$)

We have

$$\frac{D^{\frac{1}{2}} 1_n}{\sqrt{1_n^\mathsf{T} D 1_n}} = \frac{1_n}{\sqrt{n}} + \frac{1}{n\sqrt{c_0}} \frac{f'(\tau)}{2f(\tau)} \left[\{t_a 1_{\mathcal{C}_a}\}_{a=1}^k + \operatorname{diag}\left\{ \sqrt{\frac{2}{p} \operatorname{Tr}(C_a^2)} 1_{\mathcal{C}_a} \right\}_{a=1}^k \varphi \right] + o(n^{-1})$$

with $t_a := \frac{1}{\sqrt{p}} \operatorname{Tr} C_a^\circ$ *(*$a = 1, \ldots, k$*) and* $\varphi \sim \mathcal{N}(0, I_n)$.

To sum up :

- structure of $D^{\frac{1}{2}} 1_n$: block-wise constant + noise
- information about the classes depending on the numbers $\operatorname{Tr} C_a^\circ = \operatorname{Tr} C_a - \operatorname{Tr} C^\circ$

Eigenvectors

Dominant Eigenvector : $(\sqrt{d_1}, \ldots, \sqrt{d_n})^\mathsf{T}$

Proposition (Eigenvector $D^{\frac{1}{2}} 1_n$)

We have

$$\frac{D^{\frac{1}{2}} 1_n}{\sqrt{1_n^\mathsf{T} D 1_n}} = \frac{1_n}{\sqrt{n}} + \frac{1}{n\sqrt{c_0}} \frac{f'(\tau)}{2f(\tau)} \left[\{t_a 1_{\mathcal{C}_a}\}_{a=1}^k + \operatorname{diag}\left\{ \sqrt{\frac{2}{p} \operatorname{Tr}(C_a^2)} 1_{\mathcal{C}_a} \right\}_{a=1}^k \varphi \right] + o(n^{-1})$$

with $t_a := \frac{1}{\sqrt{p}} \operatorname{Tr} C_a^\circ$ *(*$a = 1, \ldots, k$*) and* $\varphi \sim \mathcal{N}(0, I_n)$*.*

To sum up :

- structure of $D^{\frac{1}{2}} 1_n$: block-wise constant + noise
- information about the classes depending on the numbers $\operatorname{Tr} C_a^\circ = \operatorname{Tr} C_a - \operatorname{Tr} C^\circ$
- **next (and main) step :** study the projected normalized Laplacian :

$$L' = nD^{-\frac{1}{2}} K D^{-\frac{1}{2}} - n \frac{D^{\frac{1}{2}} 1_n 1_n^\mathsf{T} D^{\frac{1}{2}}}{1_n^\mathsf{T} D 1_n}.$$

Random Matrix Equivalent for L'

Theorem (Random Matrix Equivalent)

As $n, p \to \infty$, in operator norm, $\left\| L' - \hat{L}' \right\| \xrightarrow{a.s.} 0$, where

$$\hat{L}' = PW^{\mathsf{T}}WP + \chi,$$

Random Matrix Equivalent for L'

Theorem (Random Matrix Equivalent)

As $n, p \to \infty$, in operator norm, $\left\| L' - \hat{L}' \right\| \xrightarrow{a.s.} 0$, where

$$\hat{L}' = PW^{\mathsf{T}}WP + \chi,$$

with P orthogonal projection onto $\{x_1 + \cdots + x_n = 0\}$, $W = [w_1, \ldots, w_n] \in \mathbb{R}^{p \times n}$ Gaussian ($x_i = \mu_a + p^{1/2} w_i$) and

χ is matrix with rank $\leq 2k + 4$

*(**spiked model**) depending on :*

Random Matrix Equivalent for L'

Theorem (Random Matrix Equivalent)

As $n, p \to \infty$, in operator norm, $\left\| L' - \hat{L}' \right\| \xrightarrow{a.s.} 0$, where

$$\hat{L}' = PW^{\mathsf{T}}WP + \chi,$$

with P orthogonal projection onto $\{x_1 + \cdots + x_n = 0\}$,
$W = [w_1, \ldots, w_n] \in \mathbb{R}^{p \times n}$ Gaussian ($x_i = \mu_a + p^{1/2} w_i$) and

χ is matrix with rank $\leq 2k + 4$

*(**spiked model**) depending on :*

- *the class structure*

Random Matrix Equivalent for L'

Theorem (Random Matrix Equivalent)

As $n, p \to \infty$, in operator norm, $\left\|L' - \hat{L}'\right\| \xrightarrow{a.s.} 0$, where

$$\hat{L}' = PW^\mathsf{T}WP + \chi,$$

with P orthogonal projection onto $\{x_1 + \cdots + x_n = 0\}$,
$W = [w_1, \ldots, w_n] \in \mathbb{R}^{p \times n}$ Gaussian ($x_i = \mu_a + p^{1/2}w_i$) and

$$\chi \text{ is matrix with rank} \leq 2k + 4$$

*(**spiked model**) depending on :*

- *the class structure*
- *the function f through the numbers $f(\tau)$, $f'(\tau)$, $f''(\tau)$, $\tau = \frac{2}{p}\operatorname{Tr} C^\circ$*

Random Matrix Equivalent for L'

Theorem (Random Matrix Equivalent)

As $n, p \to \infty$, in operator norm, $\left\| L' - \hat{L}' \right\| \xrightarrow{a.s.} 0$, where

$$\hat{L}' = PW^{\mathsf{T}}WP + \chi,$$

with P orthogonal projection onto $\{x_1 + \cdots + x_n = 0\}$,
$W = [w_1, \ldots, w_n] \in \mathbb{R}^{p \times n}$ Gaussian ($x_i = \mu_a + p^{1/2} w_i$) and

χ is matrix with rank $\leq 2k + 4$

*(**spiked model**) depending on :*

- *the class structure*
- *the function f through the numbers $f(\tau)$, $f'(\tau)$, $f''(\tau)$, $\tau = \frac{2}{p} \operatorname{Tr} C^{\circ}$*
- *the means μ_a ($a = 1, \ldots, k$)*

Random Matrix Equivalent for L'

Theorem (Random Matrix Equivalent)

As $n, p \to \infty$, in operator norm, $\left\|L' - \hat{L}'\right\| \xrightarrow{a.s.} 0$, where

$$\hat{L}' = PW^\mathsf{T}WP + \chi,$$

with P orthogonal projection onto $\{x_1 + \cdots + x_n = 0\}$,
$W = [w_1, \ldots, w_n] \in \mathbb{R}^{p \times n}$ Gaussian ($x_i = \mu_a + p^{1/2}w_i$) and

χ is matrix with rank $\leq 2k + 4$

*(**spiked model**) depending on :*

- *the class structure*
- *the function f through the numbers $f(\tau)$, $f'(\tau)$, $f''(\tau)$, $\tau = \frac{2}{p}\operatorname{Tr} C^\circ$*
- *the means μ_a ($a = 1, \ldots, k$)*
- *the traces $t_a = \frac{1}{\sqrt{p}}\operatorname{Tr} C_a^\circ$ ($a = 1, \ldots, k$)*

Random Matrix Equivalent for L'

Theorem (Random Matrix Equivalent)

As $n, p \to \infty$, in operator norm, $\left\| L' - \hat{L}' \right\| \xrightarrow{a.s.} 0$, where

$$\hat{L}' = PW^{\mathsf{T}}WP + \chi,$$

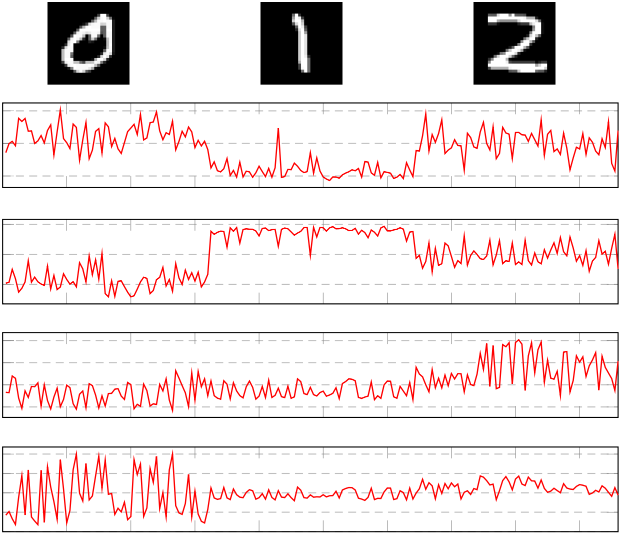
with P orthogonal projection onto $\{x_1 + \cdots + x_n = 0\}$,
$W = [w_1, \ldots, w_n] \in \mathbb{R}^{p \times n}$ Gaussian ($x_i = \mu_a + p^{1/2} w_i$) and

$$\chi \text{ is matrix with rank} \leq 2k + 4$$

*(**spiked model**) depending on :*

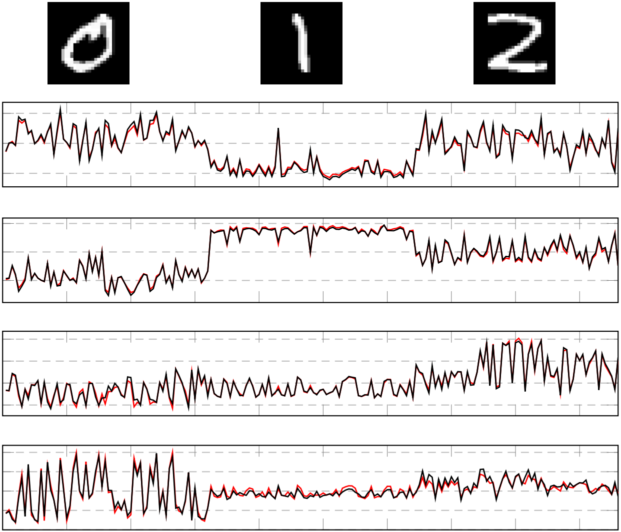
- *the class structure*
- *the function f through the numbers $f(\tau)$, $f'(\tau)$, $f''(\tau)$, $\tau = \frac{2}{p} \operatorname{Tr} C^\circ$*
- *the means μ_a ($a = 1, \ldots, k$)*
- *the traces $t_a = \frac{1}{\sqrt{p}} \operatorname{Tr} C_a^\circ$ ($a = 1, \ldots, k$)*
- *the cross-traces $T_{a,b} := \frac{1}{p} \operatorname{Tr} C_a^\circ C_b^\circ$ ($a, b = 1, \ldots, k$)*

Equivalence between L and $\hat{L}$: eigenvectors



: MNIST data : four leading eigenvectors of L (**red**), versus $\hat{L}$ (**black**)

Equivalence between L and $\hat{L}$: eigenvectors



: MNIST data : four leading eigenvectors of L (**red**), versus $\hat{L}$ (**black**)

Towards the eigenvectors of $\hat{L}' = PW^\top WP + \chi$

Towards the eigenvectors of $\hat{L}' = PW^\mathsf{T}WP + \chi$

1. Study eigenvalue distribution (and its support $\mathcal{S}$) of $PW^\mathsf{T}WP$

Towards the eigenvectors of $\hat{L}' = PW^{\mathsf{T}}WP + \chi$

1. Study eigenvalue distribution (and its support $\mathcal{S}$) of $PW^{\mathsf{T}}WP$

2. Isolated eigenvalues of $\hat{L}'$: solve, for $z \notin \mathcal{S}$,

$$\det\left(PW^{\mathsf{T}}WP + \chi - z\right) \;=\; 0$$

Towards the eigenvectors of $\hat{L}' = PW^{\mathsf{T}}WP + \chi$

1. Study eigenvalue distribution (and its support $\mathcal{S}$) of $PW^{\mathsf{T}}WP$

2. Isolated eigenvalues of $\hat{L}'$: solve, for $z \notin \mathcal{S}$,

$$\det\left(PW^{\mathsf{T}}WP + \chi - z\right) = 0$$

turning this $n \times n$ determinant to a smaller one :

$$\det\left(PW^{\mathsf{T}}WP + \chi - zI_n\right)$$

$$= \det\left(PW^{\mathsf{T}}WP - z\right)\det\underbrace{\left(1 + \left(PW^{\mathsf{T}}WP - z\right)^{-1}\chi\right)}$$

Towards the eigenvectors of $\hat{L}' = PW^\mathsf{T}WP + \chi$

1. Study eigenvalue distribution (and its support $\mathcal{S}$) of $PW^\mathsf{T}WP$

2. Isolated eigenvalues of $\hat{L}'$: solve, for $z \notin \mathcal{S}$,

$$\det\left(PW^\mathsf{T}WP + \chi - z\right) \;=\; 0$$

turning this $n \times n$ determinant to a smaller one :

$$\det\left(PW^\mathsf{T}WP + \chi - zI_n\right)$$

$$= \det\left(PW^\mathsf{T}WP - z\right) \det \underbrace{\left(1 + \left(PW^\mathsf{T}WP - z\right)^{-1}\chi\right)}_{\text{matrix with small co-rank}}$$

Towards the eigenvectors of $\hat{L}' = PW^\mathsf{T}WP + \chi$

1. Study eigenvalue distribution (and its support $\mathcal{S}$) of $PW^\mathsf{T}WP$

2. Isolated eigenvalues of $\hat{L}'$: solve, for $z \notin \mathcal{S}$,

$$\det\left(PW^\mathsf{T}WP + \chi - z\right) \;=\; 0$$

turning this $n \times n$ determinant to a smaller one :

$$\det\left(PW^\mathsf{T}WP + \chi - zI_n\right)$$

$$= \det\left(PW^\mathsf{T}WP - z\right)\det\underbrace{\left(1 + \left(PW^\mathsf{T}WP - z\right)^{-1}\chi\right)}_{\text{matrix with small co-rank}}$$

3. Study the eigenvectors thanks to the Cauchy Formula :

$$\text{Spectral Projection of } \hat{L}' \text{ on } I \;=\; \frac{1}{2\mathrm{i}\pi}\oint_{\gamma_I}(z - \hat{L})^{-1}\mathrm{d}z$$

Step 1 : eigenvalue distribution of $PW^\top WP$

Step 1 : eigenvalue distribution of $PW^\intercal WP$

- Random measure $\mu_n = n^{-1} \sum_{i=1}^n \delta_{\lambda_i(PW^\intercal WP)}$

Step 1 : eigenvalue distribution of $PW^\mathsf{T}WP$

- Random measure $\mu_n = n^{-1} \sum_{i=1}^n \delta_{\lambda_i(PW^\mathsf{T}WP)}$ studied through its **Stieltjes transform**

$$S_{\mu_n}(z) \;=\; \int_{\lambda\in\mathbb{R}} \frac{\mu_n(\mathrm{d}\lambda)}{\lambda - z} \;=\; \frac{1}{n}\operatorname{Tr}(Q_z) \qquad \text{for } Q_z = (PW^\mathsf{T}WP - z)^{-1}$$

Step 1 : eigenvalue distribution of $PW^\mathsf{T}WP$

- Random measure $\mu_n = n^{-1} \sum_{i=1}^n \delta_{\lambda_i(PW^\mathsf{T}WP)}$ studied through its **Stieltjes transform**

$$S_{\mu_n}(z) \; = \; \int_{\lambda\in\mathbb{R}} \frac{\mu_n(\mathrm{d}\lambda)}{\lambda - z} \; = \; \frac{1}{n}\,\mathrm{Tr}(Q_z) \qquad \text{for } Q_z = (PW^\mathsf{T}WP - z)^{-1}$$

- $Q_z = z^{-1}PW^\mathsf{T}WPQ_z - z^{-1}I_n$ + Stein Formula for Gaussian variables $\mathbb{E}[Xf(X)] = \sigma^2\mathbb{E}[f'(X)]$

Step 1 : eigenvalue distribution of $PW^\top WP$

- Random measure $\mu_n = n^{-1}\sum_{i=1}^n \delta_{\lambda_i(PW^\top WP)}$ studied through its **Stieltjes transform**

$$S_{\mu_n}(z) = \int_{\lambda\in\mathbb{R}} \frac{\mu_n(\mathrm{d}\lambda)}{\lambda - z} = \frac{1}{n}\operatorname{Tr}(Q_z) \qquad \text{for } Q_z = (PW^\top WP - z)^{-1}$$

- $Q_z = z^{-1}PW^\top WPQ_z - z^{-1}I_n$ + Stein Formula for Gaussian variables $\mathbb{E}[Xf(X)] = \sigma^2\mathbb{E}[f'(X)] \Longrightarrow$ Loop equations for Q_z

Step 1 : eigenvalue distribution of $PW^\mathsf{T}WP$

- Random measure $\mu_n = n^{-1}\sum_{i=1}^n \delta_{\lambda_i(PW^\mathsf{T}WP)}$ studied through its **Stieltjes transform**

$$S_{\mu_n}(z) \;=\; \int_{\lambda\in\mathbb{R}} \frac{\mu_n(\mathrm{d}\lambda)}{\lambda - z} \;=\; \frac{1}{n}\,\mathrm{Tr}(Q_z) \qquad \text{for } Q_z = (PW^\mathsf{T}WP - z)^{-1}$$

- $Q_z = z^{-1}PW^\mathsf{T}WPQ_z - z^{-1}I_n$ + Stein Formula for Gaussian variables $\mathbb{E}[Xf(X)] = \sigma^2\mathbb{E}[f'(X)] \Longrightarrow$ Loop equations for Q_z $\Longrightarrow$ fixed point characterization of a deterministic equivalent :

$$Q_z \;=\; \frac{p}{n}\,\mathrm{diag}\left\{g_a(z)1_{n_a}\right\}_{a=1}^k + o(1)$$

Step 1 : eigenvalue distribution of $PW^\mathsf{T}WP$

- Random measure $\mu_n = n^{-1}\sum_{i=1}^n \delta_{\lambda_i(PW^\mathsf{T}WP)}$ studied through its **Stieltjes transform**

$$S_{\mu_n}(z) \; = \; \int_{\lambda\in\mathbb{R}} \frac{\mu_n(\mathrm{d}\lambda)}{\lambda - z} \; = \; \frac{1}{n}\operatorname{Tr}(Q_z) \qquad \text{for } Q_z = (PW^\mathsf{T}WP - z)^{-1}$$

- $Q_z = z^{-1}PW^\mathsf{T}WPQ_z - z^{-1}I_n$ + Stein Formula for Gaussian variables $\mathbb{E}[Xf(X)] = \sigma^2\mathbb{E}[f'(X)] \Longrightarrow$ Loop equations for Q_z $\Longrightarrow$ fixed point characterization of a deterministic equivalent :

$$Q_z \; = \; \frac{p}{n}\operatorname{diag}\left\{g_a(z)1_{n_a}\right\}_{a=1}^k + o(1)$$

for $(g_a(z))_{a=1,\ldots,k} \in \mathbb{C}^k$ solution of

$$g_a(z) \; = \; \frac{1}{\frac{1}{n}\operatorname{tr} C_a\left(I_p + \sum_{b=1}^k \frac{n_b}{n} g_b(z) C_b\right)^{-1} - pz/n} \qquad (a = 1,\ldots,k)$$

Step 1 : eigenvalue distribution of $PW^{\mathsf{T}}WP$

- Random measure $\mu_n = n^{-1}\sum_{i=1}^{n}\delta_{\lambda_i(PW^{\mathsf{T}}WP)}$ studied through its **Stieltjes transform**

$$S_{\mu_n}(z) \;=\; \int_{\lambda\in\mathbb{R}} \frac{\mu_n(\mathrm{d}\lambda)}{\lambda - z} \;=\; \frac{1}{n}\,\mathrm{Tr}(Q_z) \qquad \text{for } Q_z = (PW^{\mathsf{T}}WP - z)^{-1}$$

- $Q_z = z^{-1}PW^{\mathsf{T}}WPQ_z - z^{-1}I_n$ + Stein Formula for Gaussian variables $\mathbb{E}[Xf(X)] = \sigma^2\mathbb{E}[f'(X)] \Longrightarrow$ Loop equations for $Q_z \Longrightarrow$ fixed point characterization of a deterministic equivalent :

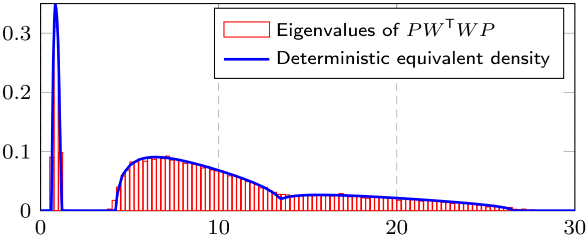
$$Q_z \;=\; \frac{p}{n}\,\mathrm{diag}\left\{g_a(z)1_{n_a}\right\}_{a=1}^{k} + o(1)$$

for $(g_a(z))_{a=1,\ldots,k}\in\mathbb{C}^k$ solution of

$$g_a(z) \;=\; \frac{1}{\frac{1}{n}\,\mathrm{tr}\,C_a\left(I_p + \sum_{b=1}^{k}\frac{n_b}{n}g_b(z)C_b\right)^{-1} - pz/n} \qquad (a = 1,\ldots,k)$$

$\Longrightarrow$ deterministic equivalent of μ_n

Eigenvalue distribution of $PW^{\mathsf{T}}WP$



: Eigenvalues of $PW^{\mathsf{T}}WP$ (across $1\,000$ realizations) versus deterministic equivalent density, $n = 32$, $p = 256$, $k = 3$.

Deterministic equivalent computed throught its Stieltjes transform :

$$S(z) \;=\; \frac{p}{n}\sum_{a=1}^{k}\frac{n_a}{n}g_a(z)$$

and the formula

$$\mathrm{density}(x) = \lim_{\eta\downarrow 0}\frac{1}{\pi}\Im(S(x+\mathrm{i}\eta))$$

Step 2 : isolated eigenvalues of $\hat{L}'$

$$\hat{L}' = PW^{\mathsf{T}}WP + \chi = PW^{\mathsf{T}}WP + UBU^{\mathsf{T}}, \qquad B : r \times r$$

Step 2 : isolated eigenvalues of $\hat{L}'$

$$\hat{L}' = PW^\top WP + \chi = PW^\top WP + UBU^\top, \qquad B : r \times r$$

$\det(I_n + XY) = \det(I_r + YX)$, (true even for rectangular matrices X, Y !)

Step 2 : isolated eigenvalues of $\hat{L}'$

$$\hat{L}' = PW^\top WP + \chi = PW^\top WP + UBU^\top, \qquad B : r \times r$$

$\det(I_n + XY) = \det(I_r + YX)$, (true even for rectangular matrices X, Y !)

$$\det(z - \hat{L}') \quad =$$

Step 2 : isolated eigenvalues of $\hat{L}'$

$$\hat{L}' = PW^\top WP + \chi = PW^\top WP + UBU^\top, \qquad B : r \times r$$

$\det(I_n + XY) = \det(I_r + YX)$, (true even for rectangular matrices X, Y !)

$$\det(z - \hat{L}') = \det(I_n - (z - PW^\top WP)^{-1}\chi)\det(z - PW^\top WP)$$

Step 2 : isolated eigenvalues of $\hat{L}'$

$$\hat{L}' = PW^\mathsf{T}WP + \chi = PW^\mathsf{T}WP + UBU^\mathsf{T}, \qquad B : r \times r$$

$\det(I_n + XY) = \det(I_r + YX)$, (true even for rectangular matrices X, Y !)

$$\begin{aligned} \det(z - \hat{L}') &= \det(I_n - (z - PW^\mathsf{T}WP)^{-1}\chi)\det(z - PW^\mathsf{T}WP) \\ &= \det(I_r - U^\mathsf{T}(z - PW^\mathsf{T}WP)^{-1}BU)\det(z - PW^\mathsf{T}WP) \end{aligned}$$

Step 2 : isolated eigenvalues of $\hat{L}'$

$$\hat{L}' = PW^\mathsf{T}WP + \chi = PW^\mathsf{T}WP + UBU^\mathsf{T}, \qquad B : r \times r$$

$\det(I_n + XY) = \det(I_r + YX)$, (true even for rectangular matrices X, Y !)

$$\begin{aligned}\det(z - \hat{L}') &= \det(I_n - (z - PW^\mathsf{T}WP)^{-1}\chi)\det(z - PW^\mathsf{T}WP) \\ &= \det(I_r - U^\mathsf{T}(z - PW^\mathsf{T}WP)^{-1}BU)\det(z - PW^\mathsf{T}WP)\end{aligned}$$

$\hookrightarrow$ key-tool : the $r \times r$ matrix $U^\mathsf{T}(z - PW^\mathsf{T}WP)^{-1}BU$

Step 2 : isolated eigenvalues of $\hat{L}'$

As

$$(z - PW^{\mathsf{T}}WP)^{-1} = -\frac{p}{n}\operatorname{diag}\left\{g_a(z)1_{n_a}\right\}_{a=1}^{k} + o(1)$$

for $(g_a(z))_{a=1,\ldots,k} \in \mathbb{C}^k$ solution of fixed point equation

$$g_a(z) = \frac{1}{\frac{1}{n}\operatorname{tr} C_a\left(I_p + \sum_{b=1}^{k}\frac{n_b}{n}g_b(z)C_b\right)^{-1} - pz/n} \qquad (a = 1,\ldots,k),$$

we get :

Step 2 : isolated eigenvalues of $\hat{L}'$

As

$$(z - PW^{\mathsf{T}}WP)^{-1} = -\frac{p}{n}\operatorname{diag}\{g_a(z)1_{n_a}\}_{a=1}^{k} + o(1)$$

for $(g_a(z))_{a=1,\ldots,k} \in \mathbb{C}^k$ solution of fixed point equation

$$g_a(z) = \frac{1}{\frac{1}{n}\operatorname{tr} C_a \left(I_p + \sum_{b=1}^{k} \frac{n_b}{n} g_b(z) C_b\right)^{-1} - pz/n} \qquad (a = 1, \ldots, k),$$

we get :

Theorem

There is a (complicated but human) function $F(z)$ such that (up to some technical hypotheses) the isolated eigenvalues of $\hat{L}'$ are the roots of

$$F(z) = 0.$$

Step 3 : isolated eigenvectors of $\hat{L}'$

We want to derive, for each such eigenvector $\vec{V}$ and each $a = 1, \ldots, k$:

Step 3 : isolated eigenvectors of $\hat{L}'$

We want to derive, for each such eigenvector $\vec{V}$ and each $a = 1, \ldots, k$:

- *Class-wise eigenvector means* :

$$\alpha_a(\vec{V}) := \frac{1}{\#\mathcal{C}_a} \left\langle \vec{V}, 1_{\mathcal{C}_a} \right\rangle$$

Step 3 : isolated eigenvectors of $\hat{L}'$

We want to derive, for each such eigenvector $\vec{V}$ and each $a = 1, \ldots, k$:

- *Class-wise eigenvector means* :

$$\alpha_a(\vec{V}) := \frac{1}{\#\mathcal{C}_a} \left\langle \vec{V}, 1_{\mathcal{C}_a} \right\rangle$$

- *Class-wise eigenvector fluctuations* :

$$\left\| \mathrm{diag}(1_{\mathcal{C}_a}) \left(\vec{V} - \alpha_a(\vec{V}) 1_{\mathcal{C}_a} \right) \right\|$$

Step 3 : isolated eigenvectors of $\hat{L}'$

We want to derive, for each such eigenvector $\vec{V}$ and each $a = 1, \ldots, k$:

- *Class-wise eigenvector means* :

$$\alpha_a(\vec{V}) := \frac{1}{\#\mathcal{C}_a} \left\langle \vec{V}, 1_{\mathcal{C}_a} \right\rangle$$

- *Class-wise eigenvector fluctuations* :

$$\left\| \mathrm{diag}(1_{\mathcal{C}_a}) \left(\vec{V} - \alpha_a(\vec{V}) 1_{\mathcal{C}_a} \right) \right\|$$

- *Class-wise cross correlations* :

$$\left\langle \left(\vec{V} - \alpha_a(\vec{V}) 1_{\mathcal{C}_a} \right), \mathrm{diag}(1_{\mathcal{C}_a}) \left(\vec{W} - \alpha_a(\vec{W}) 1_{\mathcal{C}_a} \right) \right\rangle$$

  for $\vec{W}$ another leading eigenvector

Step 3 : isolated eigenvectors of $\hat{L}'$

We want to derive, for each such eigenvector $\vec{V}$ and each $a = 1, \ldots, k$:

- *Class-wise eigenvector means* :

$$\alpha_a(\vec{V}) := \frac{1}{\#\mathcal{C}_a} \left\langle \vec{V}, 1_{\mathcal{C}_a} \right\rangle$$

- *Class-wise eigenvector fluctuations* :

$$\left\| \operatorname{diag}(1_{\mathcal{C}_a}) \left(\vec{V} - \alpha_a(\vec{V}) 1_{\mathcal{C}_a} \right) \right\|$$

- *Class-wise cross correlations* :

$$\left\langle \left(\vec{V} - \alpha_a(\vec{V}) 1_{\mathcal{C}_a} \right), \operatorname{diag}(1_{\mathcal{C}_a}) \left(\vec{W} - \alpha_a(\vec{W}) 1_{\mathcal{C}_a} \right) \right\rangle$$

  for $\vec{W}$ another leading eigenvector

$\hookrightarrow$ one needs, for $\Pi = \vec{V}\vec{V}^\mathsf{T}$ and $\Pi' = \vec{W}\vec{W}^\mathsf{T}$, the numbers

$$p^{-1} J^\mathsf{T} \Pi J \qquad ; \qquad p^{-1} J^\mathsf{T} \Pi \operatorname{diag}(1_{\mathcal{C}_a}) \Pi' J \qquad (1 \leq a \leq k)$$

for $J = [1_{\mathcal{C}_1} \cdots 1_{\mathcal{C}_k}] \in \mathbb{R}^{n \times k}$.

Step 3 : isolated eigenvectors of $\hat{L}'$

- Cauchy Formula :

$$\text{Spectral Projection of } \hat{L}' \text{ on } I \; = \; -\frac{1}{2\mathrm{i}\pi}\oint_{\gamma_I}(\hat{L}' - z)^{-1}\mathrm{d}z$$

Step 3 : isolated eigenvectors of $\hat{L}'$

- Cauchy Formula :

$$\text{Spectral Projection of } \hat{L}' \text{ on } I \;=\; -\frac{1}{2\mathrm{i}\pi}\oint_{\gamma_I} (\hat{L}' - z)^{-1}\mathrm{d}z$$

- Woodbury matrix identity : for $Q_z = (PW^\mathsf{T}WP - z)^{-1}$,

$$(\hat{L}' - z)^{-1} \;=\; Q_z - Q_z U(B^{-1} + U^\mathsf{T} Q_z U)^{-1} U^\mathsf{T} Q_z$$

Step 3 : isolated eigenvectors of $\hat{L}'$

- Cauchy Formula :

$$\text{Spectral Projection of } \hat{L}' \text{ on } I \;=\; -\frac{1}{2\mathrm{i}\pi}\oint_{\gamma_I}(\hat{L}'-z)^{-1}\mathrm{d}z$$

- Woodbury matrix identity : for $Q_z = (PW^\mathsf{T}WP - z)^{-1}$,

$$(\hat{L}'-z)^{-1} \;=\; Q_z - Q_zU(B^{-1}+U^\mathsf{T}Q_zU)^{-1}U^\mathsf{T}Q_z$$

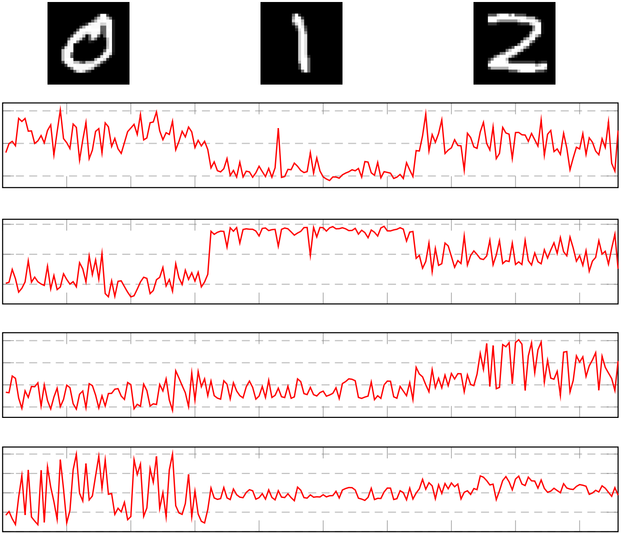
Theorem

The deterministic equivalents of the $k\times k$ matrices

$$p^{-1}J^\mathsf{T}\Pi J \qquad ; \qquad p^{-1}J^\mathsf{T}\Pi\,\mathrm{diag}(1_{\mathcal{C}_a})\Pi' J \qquad (1\le a\le k)$$

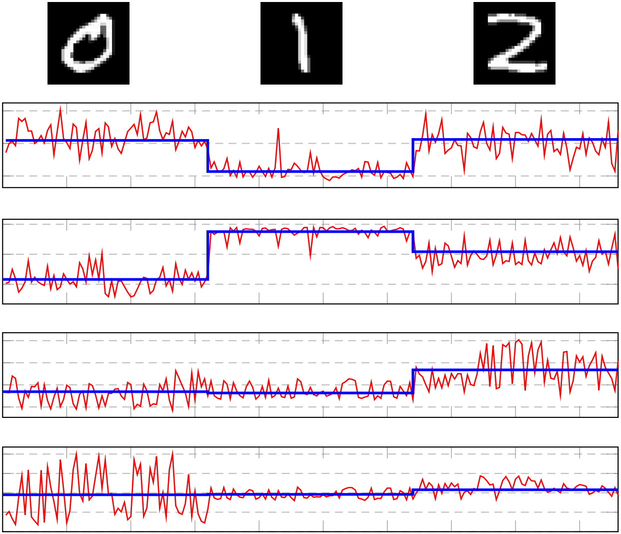
can be computed thanks to the parameters and the solutions of the fixed point equations $g_a(z)$, $a=1,\ldots,k$.

Class-wise eigenvector means and fluctuations



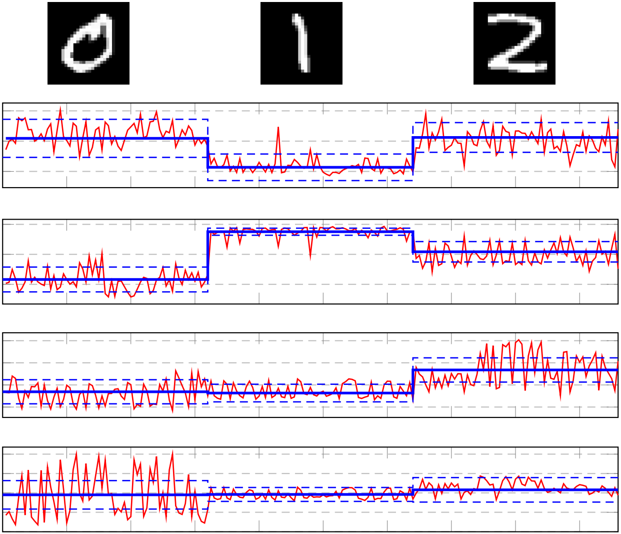
: MNIST data : four leading eigenvectors of L (**red**) and theoretical findings (**blue**).

Class-wise eigenvector means and fluctuations



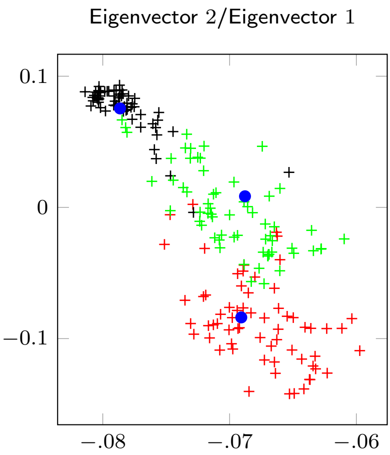
: MNIST data : four leading eigenvectors of L (**red**) and theoretical findings (**blue**).

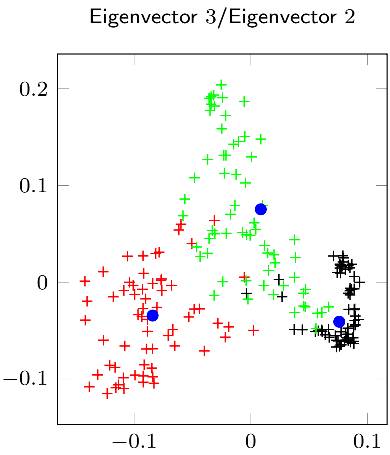
Class-wise eigenvector means and fluctuations



: MNIST data : four leading eigenvectors of L (**red**) and theoretical findings (**blue**).

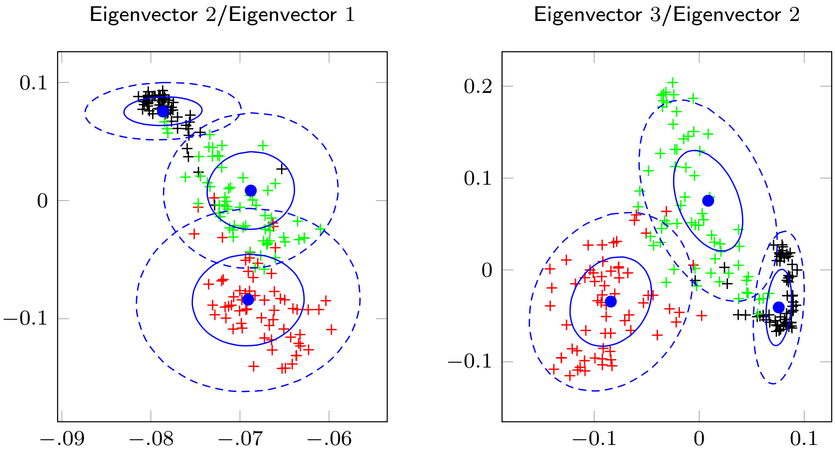
Class-wise means, fluctuations and cross correlations





: 2D representation of eigenvectors of L, for the MNIST dataset. Theoretical means and 1- and 2-standard deviations in **blue**. Class 1 in **red**, Class 2 in **black**, Class 3 in **green**.

Class-wise means, fluctuations and cross correlations



: 2D representation of eigenvectors of L, for the MNIST dataset. Theoretical means and 1- and 2-standard deviations in **blue**. Class 1 in **red**, Class 2 in **black**, Class 3 in **green**.

Concluding Remarks

Summing up :

Concluding Remarks

Summing up :

- Although Gaussian-based, adequately mimics real world examples

Concluding Remarks

Summing up :

- Although Gaussian-based, adequately mimics real world examples
- Noticeable Results :

Concluding Remarks

Summing up :

- Although Gaussian-based, adequately mimics real world examples
- Noticeable Results :
  - importance of derivatives of f at τ

Concluding Remarks

Summing up :

- Although Gaussian-based, adequately mimics real world examples
- Noticeable Results :
  - importance of derivatives of f at τ
  - choice of $f(\tau), f'(\tau), f''(\tau)$ determines importance of means, covariances

Concluding Remarks

Summing up :

- Although Gaussian-based, adequately mimics real world examples
- Noticeable Results :
  - importance of derivatives of f at τ
  - choice of $f(\tau), f'(\tau), f''(\tau)$ determines importance of means, covariances
  - eigenvector may or may not contain information (upon separability condition !)

Concluding Remarks

Summing up :

- Although Gaussian-based, adequately mimics real world examples
- Noticeable Results :
  - importance of derivatives of f at τ
  - choice of $f(\tau), f'(\tau), f''(\tau)$ determines importance of means, covariances
  - eigenvector may or may not contain information (upon separability condition !)
  - number of isolated eigenvalues not obvious

Concluding Remarks

Summing up :

- Although Gaussian-based, adequately mimics real world examples
- Noticeable Results :
  - importance of derivatives of f at τ
  - choice of $f(\tau), f'(\tau), f''(\tau)$ determines importance of means, covariances
  - eigenvector may or may not contain information (upon separability condition !)
  - number of isolated eigenvalues not obvious

Perspectives :

- (joint) class-wise eigenvector fluctuations
- implications to spectral clustering performance
- algorithm comparison
- ideally, (data-driven) algorithm improvement.