# Identification de mélanges gaussiens via des méthodes tensorielles

Vicnesh Venedittan encadré par Rima Khouja

2 Juin 2023

# Introduction

- Loi gaussienne multivariée

# Introduction

- Loi gaussienne multivariée
- Modèle de mélange Gaussien

# Introduction

- Loi gaussienne multivariée
- Modèle de mélange Gaussien
- Tenseur

# Loi gaussienne multivariée

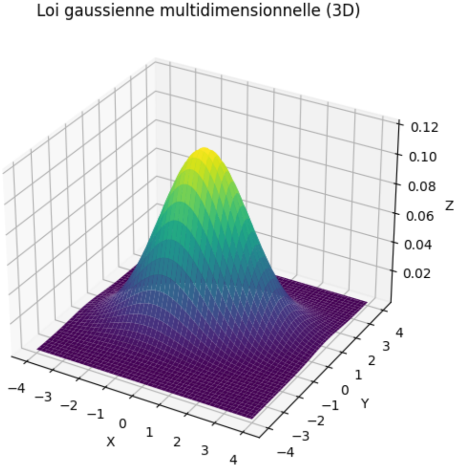


## Definition

Soit $X \in \mathbb{R}^n$ un vecteur aléatoire, $X$ suit une distribution gaussienne multivariée de paramètre $\mu \in \mathbb{R}^n$ et $\Sigma \in \mathbb{R}^{n \times n}$ définie positive, s'il admet pour densité :

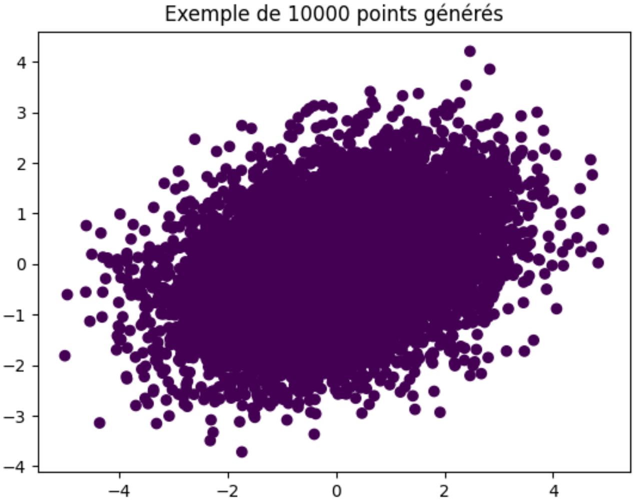
$$f(X, \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} exp(-\frac{1}{2}(X - \mu)^t \Sigma^{-1}(X - \mu))$$

# Exemple loi gaussienne multivariée

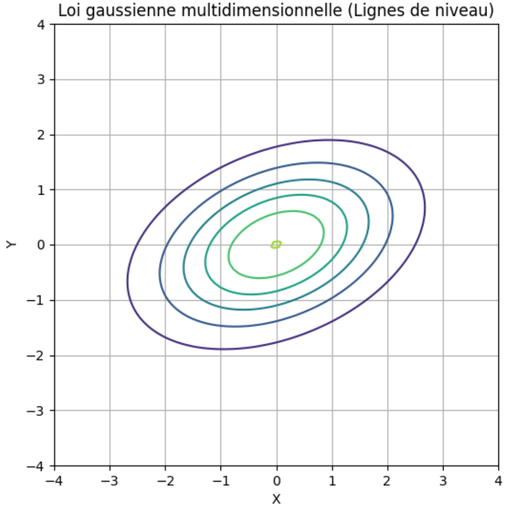


# Exemple loi gaussienne multivariée

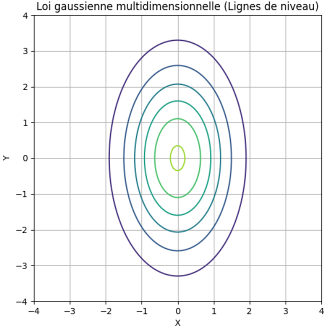


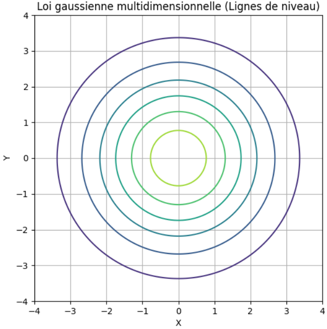


# Exemple loi gaussienne multivariée



# Remarque


Loi gaussienne multidimensionnelle (Lignes de niveau)
X
Y


Loi gaussienne multidimensionnelle (Lignes de niveau)
X
Y

# Modèle de mélange gaussien



## Definition

Soit $\lambda = (\lambda_1, ..., \lambda_m)$ tel que $\sum_{i=1}^{m} \lambda_i = 1$, puis $\mu = (\mu_1, ..., \mu_m)$ et $\Sigma = (\Sigma_1, ..., \Sigma_m)$. $X$ suit une loi de mélange gaussien à $m$ composante de dimension $p$ si elle admet pour densité :

$$f(X, \mu, \Sigma) = \sum_{i=1}^{m} \lambda_i f_i(X, \mu_i, \Sigma_i).$$

Où pour tout $i \in 1, ..., m$, $f_i$ suit une loi gaussienne multivariée de paramètre $(\mu_i, \Sigma_i)$.

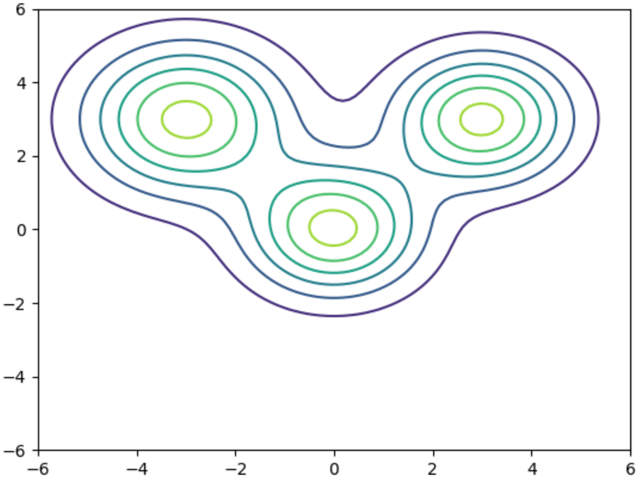
# Exemple de modèle de mélange gaussien



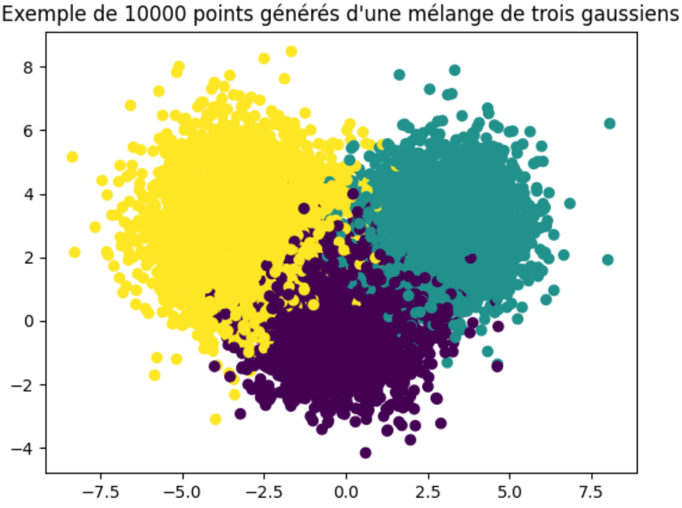
## Example

Pour $\lambda = (0.3, 0.3, 0.4)$, $\mu = \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \begin{pmatrix} 3 \\ 3 \end{pmatrix} \quad \begin{pmatrix} -3 \\ 3 \end{pmatrix} \right)$ et $\Sigma = \left( \begin{pmatrix} 1.5 & 0 \\ 0 & 1.5 \end{pmatrix} \quad \begin{pmatrix} 1.5 & 0 \\ 0 & 1.5 \end{pmatrix} \quad \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \right)$.

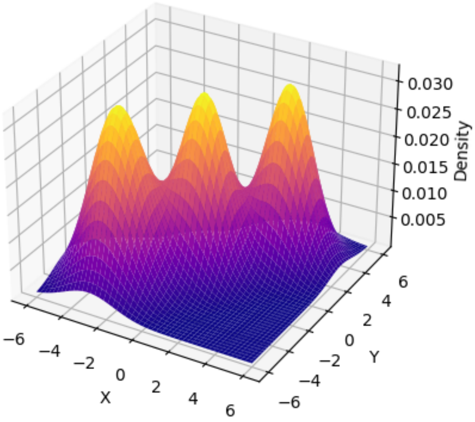
# Lignes de niveaux



# Représentation d'un mélange gaussien



# Représentation d'un mélange gaussien



# Objectif

- Un problème d'identification d'un modèle de mélange revient à estimer le paramètre $\theta = (\lambda, \mu, \Sigma)$

# Objectif



- Un problème d'identification d'un modèle de mélange revient à estimer le paramètre $\theta = (\lambda, \mu, \Sigma)$
- Algorithme espérance-maximisation (EM)

# Objectif



- Un problème d'identification d'un modèle de mélange revient à estimer le paramètre $\theta = (\lambda, \mu, \Sigma)$
- Algorithme espérance-maximisation (EM)
- Méthode itérative, on alterne deux étapes :

# Objectif



- Un problème d'identification d'un modèle de mélange revient à estimer le paramètre $\theta = (\lambda, \mu, \Sigma)$
- Algorithme espérance-maximisation (EM)
- Méthode itérative, on alterne deux étapes :
  - Etape E : Calculer les valeurs manquantes dans le modèle en utilisant les données observées et les paramètres actuels du modèle

# Objectif



- Un problème d'identification d'un modèle de mélange revient à estimer le paramètre $\theta = (\lambda, \mu, \Sigma)$
- Algorithme espérance-maximisation (EM)
- Méthode itérative, on alterne deux étapes :
  - Etape E : Calculer les valeurs manquantes dans le modèle en utilisant les données observées et les paramètres actuels du modèle
  - Etape M : on maximise la log-vraisemblance du modèle complété en ajustant les paramètres pour lesquels on a maintenant toutes les données.

# Objectif



- Un problème d'identification d'un modèle de mélange revient à estimer le paramètre $\theta = (\lambda, \mu, \Sigma)$
- Algorithme espérance-maximisation (EM)
- Méthode itérative, on alterne deux étapes :
  - Etape E : Calculer les valeurs manquantes dans le modèle en utilisant les données observées et les paramètres actuels du modèle
  - Etape M : on maximise la log-vraisemblance du modèle complété en ajustant les paramètres pour lesquels on a maintenant toutes les données.
- Méthode des moments tensoriels

# Tenseur

## Definition

Soit $\mathbb{R}^{n_1} \otimes \ldots \otimes \mathbb{R}^{n_d}$ l'espace vectoriel du produit tensoriel de $\mathbb{R}^{n_1}, \ldots, \mathbb{R}^{n_d}$ qu'on note $\mathcal{T}_n^d$ avec $n = (n_1, ..., n_d)$. Un tenseur $\mathcal{T}$ de $\mathcal{T}_n^d$ peut être représenté par un tableau multidimensionnel dans $\mathbb{R}^{n_1 \times \ldots \times n_d}$, par rapport à une base fixe sur $\mathbb{R}^{n_1}, \ldots, \mathbb{R}^{n_d}$ avec $\mathcal{T} = [t_{i_1,\ldots,i_d}]_{1 \leq i_1 \leq n_1, \ldots, 1 \leq i_d \leq n_d}$, où $t_{i_1,\ldots,i_d} \in \mathbb{R}$ est l'élément $(i_1, \ldots, i_d)$ du tableau. Dans ce qui suit, on considère la base canonique sur $\mathbb{R}^{n_1}, \ldots, \mathbb{R}^{n_d}$.

# Tenseur

## Definition

Soit $\mathbb{R}^{n_1} \otimes \ldots \otimes \mathbb{R}^{n_d}$ l'espace vectoriel du produit tensoriel de $\mathbb{R}^{n_1}, \ldots, \mathbb{R}^{n_d}$ qu'on note $\mathcal{T}_n^d$ avec $n = (n_1, ..., n_d)$. Un tenseur $\mathcal{T}$ de $\mathcal{T}_n^d$ peut être représenté par un tableau multidimensionnel dans $\mathbb{R}^{n_1 \times \ldots \times n_d}$, par rapport à une base fixe sur $\mathbb{R}^{n_1}, \ldots, \mathbb{R}^{n_d}$ avec $\mathcal{T} = [t_{i_1,\ldots,i_d}]_{1 \leq i_1 \leq n_1, \ldots, 1 \leq i_d \leq n_d}$, où $t_{i_1,\ldots,i_d} \in \mathbb{R}$ est l'élément $(i_1, \ldots, i_d)$ du tableau. Dans ce qui suit, on considère la base canonique sur $\mathbb{R}^{n_1}, \ldots, \mathbb{R}^{n_d}$.

- En particuliers si $n_1 = n_2 = ... = n_d$ et $t_{j_1,j_2,\ldots,j_d} = t_{j_{\sigma(1)},\ldots,j_{\sigma(d)}}$, pour tout $\sigma \in \mathfrak{S}^d$, des groupes de permutation de $\{1, ..., d\}$, on dit alors que $\mathcal{T}$ est un tenseur symétrique

# Tenseur

## Definition

Soit $\mathbb{R}^{n_1} \otimes \ldots \otimes \mathbb{R}^{n_d}$ l'espace vectoriel du produit tensoriel de $\mathbb{R}^{n_1}, \ldots, \mathbb{R}^{n_d}$ qu'on note $\mathcal{T}_n^d$ avec $n = (n_1, ..., n_d)$. Un tenseur $\mathcal{T}$ de $\mathcal{T}_n^d$ peut être représenté par un tableau multidimensionnel dans $\mathbb{R}^{n_1 \times \cdots \times n_d}$, par rapport à une base fixe sur $\mathbb{R}^{n_1}, \ldots, \mathbb{R}^{n_d}$ avec $\mathcal{T} = [t_{i_1,\ldots,i_d}]_{1 \leq i_1 \leq n_1, \ldots, 1 \leq i_d \leq n_d}$, où $t_{i_1,\ldots,i_d} \in \mathbb{R}$ est l'élément $(i_1, \ldots, i_d)$ du tableau. Dans ce qui suit, on considère la base canonique sur $\mathbb{R}^{n_1}, \ldots, \mathbb{R}^{n_d}$.

- En particuliers si $n_1 = n_2 = ... = n_d$ et $t_{j_1,j_2,\ldots,j_d} = t_{j_{\sigma(1)},\ldots,j_{\sigma(d)}}$, pour tout $\sigma \in \mathfrak{S}^d$, des groupes de permutation de $\{1, ..., d\}$, on dit alors que $\mathcal{T}$ est un tenseur symétrique
- On note $\mathcal{S}_n^d$ l'espace des tenseurs symétriques d'ordre $d$ de dimension $n$.

### Definition (Opération *Sym*)

Pour tout $\mathcal{A} \in \mathcal{T}_n^d$, $(j_1, j_2, ..., j_d) \in \{1, ..., n\}^d$:

$$Sym[\mathcal{A}](j_1, j_2, ..., j_d) = \frac{1}{d!} \sum_{\sigma \in \mathfrak{S}^d} a_{j_{\sigma(1)}, \dots, j_{\sigma(d)}}.$$

## Definition (Produit scalaire tensoriel)

Soit $\mathcal{A}, \mathcal{B} \in \mathcal{T}_n^d$, on note le produit scalaire de $\mathcal{A}$ et $\mathcal{B}$ par :

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1=1}^{n} \dots \sum_{i_d=1}^{n} a_{i_1,\dots,i_d} b_{i_1,\dots,i_d}.$$

- En particulier, on a :

$$\langle \mathcal{A}, \mathcal{A} \rangle = ||\mathcal{A}||^2 = \sum_{i_1=1}^{n} \dots \sum_{i_d=1}^{n} a^2_{i_1,\dots,i_d}.$$

# Résumé de l'article : PEREIRA, KILEEL, KOLDA 2022

- Identification de mélange gaussien par analyse tensorielle.

# Résumé de l'article : PEREIRA, KILEEL, KOLDA 2022



- Identification de mélange gaussien par analyse tensorielle.
- Utiliser une méthode de moment.

# Résumé de l'article : PEREIRA, KILEEL, KOLDA 2022

- Identification de mélange gaussien par analyse tensorielle.
- Utiliser une méthode de moment.
- Une nouvelle expression du moment d'ordre $d$ tensoriel.

# Résumé de l'article : PEREIRA, KILEEL, KOLDA 2022

- Identification de mélange gaussien par analyse tensorielle.
- Utiliser une méthode de moment.
- Une nouvelle expression du moment d'ordre $d$ tensoriel.
- Transformer le problème d'identification en problème d'optimisation.

# Résumé de l'article

- On définit $\Phi : \bigcup_{d=0}^{\infty} \mathcal{T}_n^d \to \bigcup_{d=0}^{\infty} \mathbb{R}[z_1, ..., z_n]_d$ (l'ensemble des polynômes homogènes réels) tel que pour tout $\mathcal{V} \in \mathcal{T}_n^d$ et $z \in \mathbb{R}^n$

$$\Phi[\mathcal{V}](z) = \langle \mathcal{V}, z^{\otimes d} \rangle$$

# Résumé de l'article

- On définit $\Phi : \bigcup_{d=0}^{\infty} \mathcal{T}_n^d \to \bigcup_{d=0}^{\infty} \mathbb{R}[z_1, ..., z_n]_d$ (l'ensemble des polynômes homogènes réels) tel que pour tout $\mathcal{V} \in \mathcal{T}_n^d$ et $z \in \mathbb{R}^n$
$$\Phi[\mathcal{V}](z) = \langle \mathcal{V}, z^{\otimes d} \rangle$$
- L'application $\Phi$ est une application linéaire bijective pour la restriction sur $\mathcal{S}_n^d$. Pour tout $\mathcal{A} \in \mathcal{S}_n^d$:
$$\Phi[\mathcal{A}] = \Phi[Sym(\mathcal{A})].$$
En particuliers si $\mathcal{B} \in \mathcal{T}_n^d$, et $\Phi[\mathcal{A}] = \Phi[\mathcal{B}]$ alors $\mathcal{A} = Sym[\mathcal{B}]$.

# Résumé de l'article

- On définit $\Phi : \bigcup_{d=0}^{\infty} \mathcal{T}_n^d \to \bigcup_{d=0}^{\infty} \mathbb{R}[z_1, \ldots, z_n]_d$ (l'ensemble des polynômes homogènes réels) tel que pour tout $\mathcal{V} \in \mathcal{T}_n^d$ et $z \in \mathbb{R}^n$
$$\Phi[\mathcal{V}](z) = \langle \mathcal{V}, z^{\otimes d} \rangle$$
- L'application $\Phi$ est une application linéaire bijective pour la restriction sur $\mathcal{S}_n^d$. Pour tout $\mathcal{A} \in \mathcal{S}_n^d$:
$$\Phi[\mathcal{A}] = \Phi[Sym(\mathcal{A})].$$
En particuliers si $\mathcal{B} \in \mathcal{T}_n^d$, et $\Phi[\mathcal{A}] = \Phi[\mathcal{B}]$ alors $\mathcal{A} = Sym[\mathcal{B}]$.
- Pour tout vecteur $v \in \mathbb{R}^n$ et pour toutes matrices $M \in \mathbb{R}^{n \times n}$, les polynômes homogènes $\Phi[v], \Phi[M]$ s'écrivent :
$$\Phi[v](z) = v^T z \text{ et } \Phi[M](z) = z^T Mz.$$

# Résumé de l'article

- On définit $\Phi : \bigcup_{d=0}^{\infty} \mathcal{T}_n^d \to \bigcup_{d=0}^{\infty} \mathbb{R}[z_1, ..., z_n]_d$ (l'ensemble des polynômes homogènes réels) tel que pour tout $\mathcal{V} \in \mathcal{T}_n^d$ et $z \in \mathbb{R}^n$
$$\Phi[\mathcal{V}](z) = \langle \mathcal{V}, z^{\otimes d} \rangle$$
- L'application $\Phi$ est une application linéaire bijective pour la restriction sur $\mathcal{S}_n^d$. Pour tout $\mathcal{A} \in \mathcal{S}_n^d$:
$$\Phi[\mathcal{A}] = \Phi[Sym(\mathcal{A})].$$
En particuliers si $\mathcal{B} \in \mathcal{T}_n^d$, et $\Phi[\mathcal{A}] = \Phi[\mathcal{B}]$ alors $\mathcal{A} = Sym[\mathcal{B}]$.
- Pour tout vecteur $v \in \mathbb{R}^n$ et pour toutes matrices $M \in \mathbb{R}^{n \times n}$, les polynômes homogènes $\Phi[v], \Phi[M]$ s'écrivent :
$$\Phi[v](z) = v^T z \text{ et } \Phi[M](z) = z^T M z.$$
- Pour tout $\mathcal{A} \in \mathcal{T}_n^{d_1}, \mathcal{B} \in \mathcal{T}_n^{d_2}$, on a :
$$\Phi[\mathcal{A} \otimes \mathcal{B}] = \Phi[\mathcal{A}]\Phi[\mathcal{B}].$$

# Pereira, Kileel, Kolda, 2022



**Théorème 1 : Pereira, Kileel, Kolda, 2022**

Soit $X \sim \mathcal{N}(\mu, \Sigma)$, le moment d'ordre $d$ de $X$ est $\mathcal{M}^{(d)} = \mathbb{E}(X^{\otimes d})$ plus précisément :

$$\mathcal{M}^{(d)} = \sum_{k=0}^{\lfloor d/2 \rfloor} C_{d,k} Sym(\mu^{\otimes d-2k} \otimes \Sigma^{\otimes k}),$$

avec : $C_{d,k} = \binom{d}{2k} \frac{2k!}{k!2^k}$.

### Théorème 2 : Pereira, Kileel, Kolda, 2022

Si $X \sim \sum_{j=1}^{m} \lambda_j \mathcal{N}(\mu_j, \Sigma_j)$, le moment d'ordre $d$ de $X$ est $\mathcal{M}^{(d)} = \mathbb{E}(X^{\otimes d})$ plus précisément :

$$\mathcal{M}^{(d)} = \sum_{j=1}^{m} \sum_{k=0}^{\lfloor d/2 \rfloor} \lambda_j C_{d,k} Sym(\mu^{\otimes d-2k} \otimes \Sigma^{\otimes k}),$$

avec : $C_{d,k} = \binom{d}{2k} \frac{2k!}{k!2^k}$.

# La méthode du moment proposée

- On considère un échantillon $(x_1, ..., x_p)$ de réalisation d'un mélange gaussien de paramètre $\theta = (\lambda, \mu, \Sigma)$

# La méthode du moment proposée



- On considère un échantillon $(x_1, ..., x_p)$ de réalisation d'un mélange gaussien de paramètre $\theta = (\lambda, \mu, \Sigma)$
- On veut estimer le paramètre $\theta = (\lambda, \mu, \Sigma)$

# La méthode du moment proposée

- On considère un échantillon $(x_1, ..., x_p)$ de réalisation d'un mélange gaussien de paramètre $\theta = (\lambda, \mu, \Sigma)$
- On veut estimer le paramètre $\theta = (\lambda, \mu, \Sigma)$
- On veut approcher le moment empirique $\frac{1}{p} \sum_{i=1}^{p} x_i^{\otimes d}$ et le moment théorique $\mathcal{M}^{(d)}$, cela revient à résoudre le problème d'optimisation

# La méthode du moment proposée



- On considère un échantillon $(x_1, ..., x_p)$ de réalisation d'un mélange gaussien de paramètre $\theta = (\lambda, \mu, \Sigma)$
- On veut estimer le paramètre $\theta = (\lambda, \mu, \Sigma)$
- On veut approcher le moment empirique $\frac{1}{p}\sum_{i=1}^{p} x_i^{\otimes d}$ et le moment théorique $\mathcal{M}^{(d)}$, cela revient à résoudre le problème d'optimisation

$$min_\theta F(\theta) = ||\mathcal{M}^{(d)} - \frac{1}{p}\sum_{i=1}^{p} x_i^{\otimes d}||^2$$

# La méthode du moment proposée

- $F(\theta) = ||\mathcal{M}^{(d)} - \frac{1}{p}\sum_{i=1}^{p} x_i^{\otimes d}||^2 =$

# La méthode du moment proposée

- $F(\theta) = ||\mathcal{M}^{(d)} - \frac{1}{p}\sum_{i=1}^{p} x_i^{\otimes d}||^2 =$
- $\left\|\mathcal{M}^{(d)}\right\|^2 - 2 \times \frac{1}{p}\sum_{i=1}^{p}\langle \mathcal{M}^{(d)}, x_i^{\otimes d}\rangle$

# La méthode du moment proposée



- $F(\theta) = ||\mathcal{M}^{(d)} - \frac{1}{p}\sum_{i=1}^{p} x_i^{\otimes d}||^2 =$
- $\left\|\mathcal{M}^{(d)}\right\|^2 - 2 \times \frac{1}{p}\sum_{i=1}^{p}\langle \mathcal{M}^{(d)}, x_i^{\otimes d}\rangle$
- $F_1(\theta) = \left\|\mathcal{M}^{(d)}\right\|^2$ et $F_2(\theta) = \frac{1}{p}\sum_{i=1}^{p}\langle \mathcal{M}^{(d)}, x_i^{\otimes d}\rangle$

# La méthode du moment proposée

- Après avoir décomposé notre problème d'optimisation en 2 problèmes d'optimisation plus simple, nous allons calculer le gradient de $F_1$ et de $F_2$.

# La méthode du moment proposée



- Après avoir décomposé notre problème d'optimisation en 2 problèmes d'optimisation plus simple, nous allons calculer le gradient de $F_1$ et de $F_2$.
- L'article propose une expression des gradients explicites en fonction de $\theta$.

# La méthode du moment proposée



- Après avoir décomposé notre problème d'optimisation en 2 problèmes d'optimisation plus simple, nous allons calculer le gradient de $F_1$ et de $F_2$.
- L'article propose une expression des gradients explicites en fonction de $\theta$.
- Ces gradients peuvent être utilisé pour implémenter n'importe quelle méthode du 1er ordre de gradient.

# La méthode du moment proposée



- Après avoir décomposé notre problème d'optimisation en 2 problèmes d'optimisation plus simple, nous allons calculer le gradient de $F_1$ et de $F_2$.
- L'article propose une expression des gradients explicites en fonction de $\theta$.
- Ces gradients peuvent être utilisé pour implémenter n'importe quelle méthode du 1er ordre de gradient.
- On choisit la méthode du gradient à pas fixe.

## Méthode du gradient à pas fixe

On se donne un point initial $x_0$ et un seuil de tolérance $\epsilon > 0$. L'algorithme du gradient à pas fixe $\alpha$. L'itération $k+1$ est donné par :

- Calcul de : $\bigtriangledown f(x_k)$
- Test d'arrêt : $|| \bigtriangledown f(x_k)|| < \epsilon$
- Nouvelle itération : $x_{k+1} = x_k - \alpha \bigtriangledown f(x_k)$

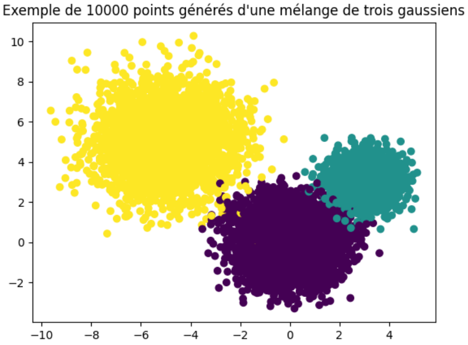
# Exemple

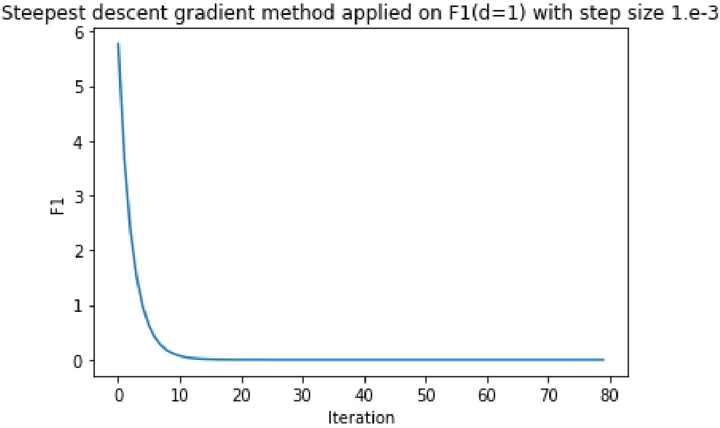
## Example

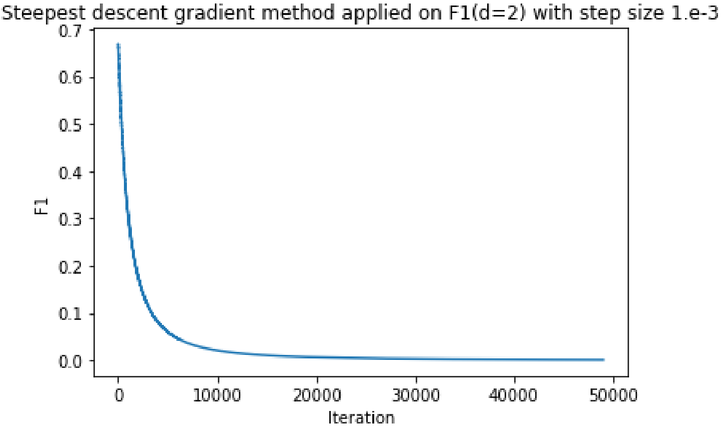
On initialise notre paramètre $\theta$ avec :

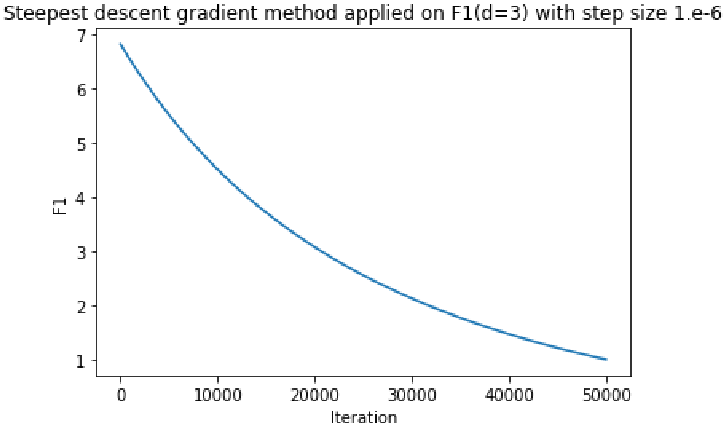
$$\lambda = (0.5, 0.2, 0.3),\ \mu = \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 3 \\ 3 \end{pmatrix} \begin{pmatrix} -5 \\ 5 \end{pmatrix} \right)$$

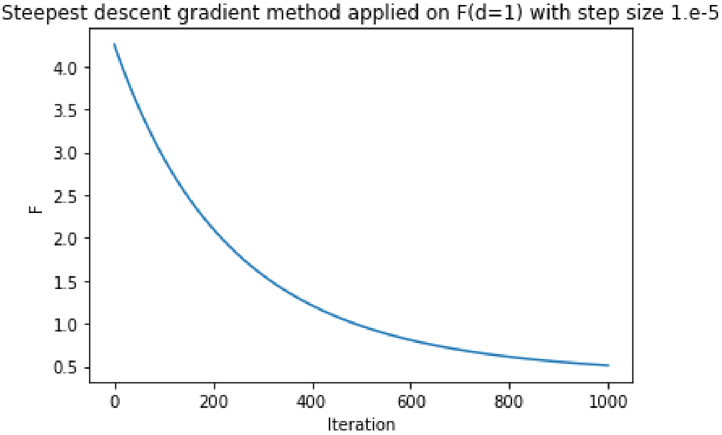
$$\Sigma = \left( \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \right).$$

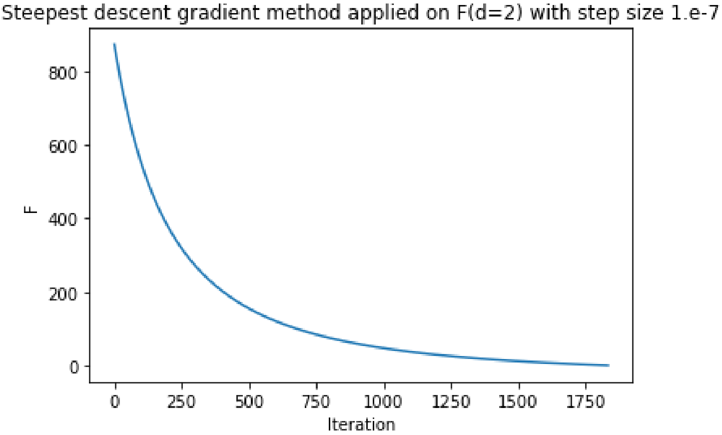



Steepest descent gradient method applied on F1(d=1) with step size 1.e-3
F1
Iteration
0
1
2
3
4
5
6
0
10
20
30
40
50
60
70
80

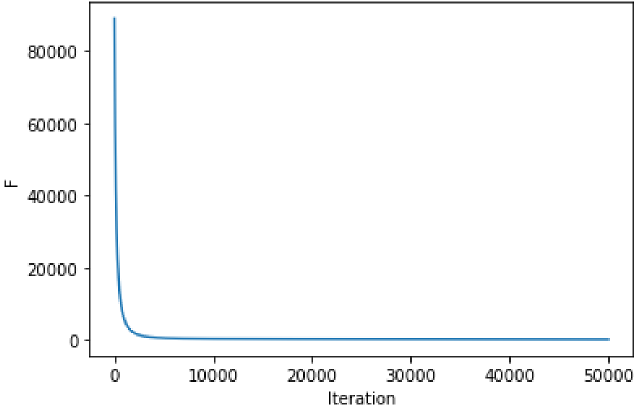

Steepest descent gradient method applied on F1(d=2) with step size 1.e-3
F1
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
10000
20000
30000
40000
50000
Iteration


Steepest descent gradient method applied on F1(d=3) with step size 1.e-6
F1
1
2
3
4
5
6
7
0
10000
20000
30000
40000
50000
Iteration


Steepest descent gradient method applied on F(d=1) with step size 1.e-5
F
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0
200
400
600
800
1000
Iteration


Steepest descent gradient method applied on F(d=2) with step size 1.e-7
F
800
600
400
200
0
0
250
500
750
1000
1250
1500
1750
Iteration

Steepest descent gradient method applied on F(d=3) with step size 1.e-9



# Conclusion



- Dans l'optique d'estimer le paramètre d'un modèle de mélange gaussien, les méthodes les plus utilisées sont :

# Conclusion



- Dans l'optique d'estimer le paramètre d'un modèle de mélange gaussien, les méthodes les plus utilisées sont :
- l'algorithme EM et la méthode des moments

# Conclusion



- Dans l'optique d'estimer le paramètre d'un modèle de mélange gaussien, les méthodes les plus utilisées sont :
- l'algorithme EM et la méthode des moments
- L'article résumé, présente une méthode algorithmique pour estimer le paramètre revenant alors à un problème d'optimisation.

# Conclusion



- Dans l'optique d'estimer le paramètre d'un modèle de mélange gaussien, les méthodes les plus utilisées sont :
- l'algorithme EM et la méthode des moments
- L'article résumé, présente une méthode algorithmique pour estimer le paramètre revenant alors à un problème d'optimisation.
- En générant un jeu de données de paramètre $\theta$ d'un modèle de mélange gaussien, on valide expérimentalement les calculs de cet article avec une méthode du gradient à pas fixe.

Merci pour votre attention !