Analyse multidimensionnelle des préférences, et application aux données de vote

SIRA LINA ACHOURI, INA CAMPAN, IDA ROGIE

ENCADRANT : OLIVIER SPANJAARD

Sommaire

1. Introduction
2. Méthodes utilisées (PM & NMDS)
3. Conception et implémentation
4. Expérimentation numérique
5. Conclusion

Contexte et données

Expérimentation *Voter Autrement* (2017 et 2022) :

- Vote par note
- Jugement majoritaire
- Méthode de Borda

Format des données

VOTER AUTREMENT

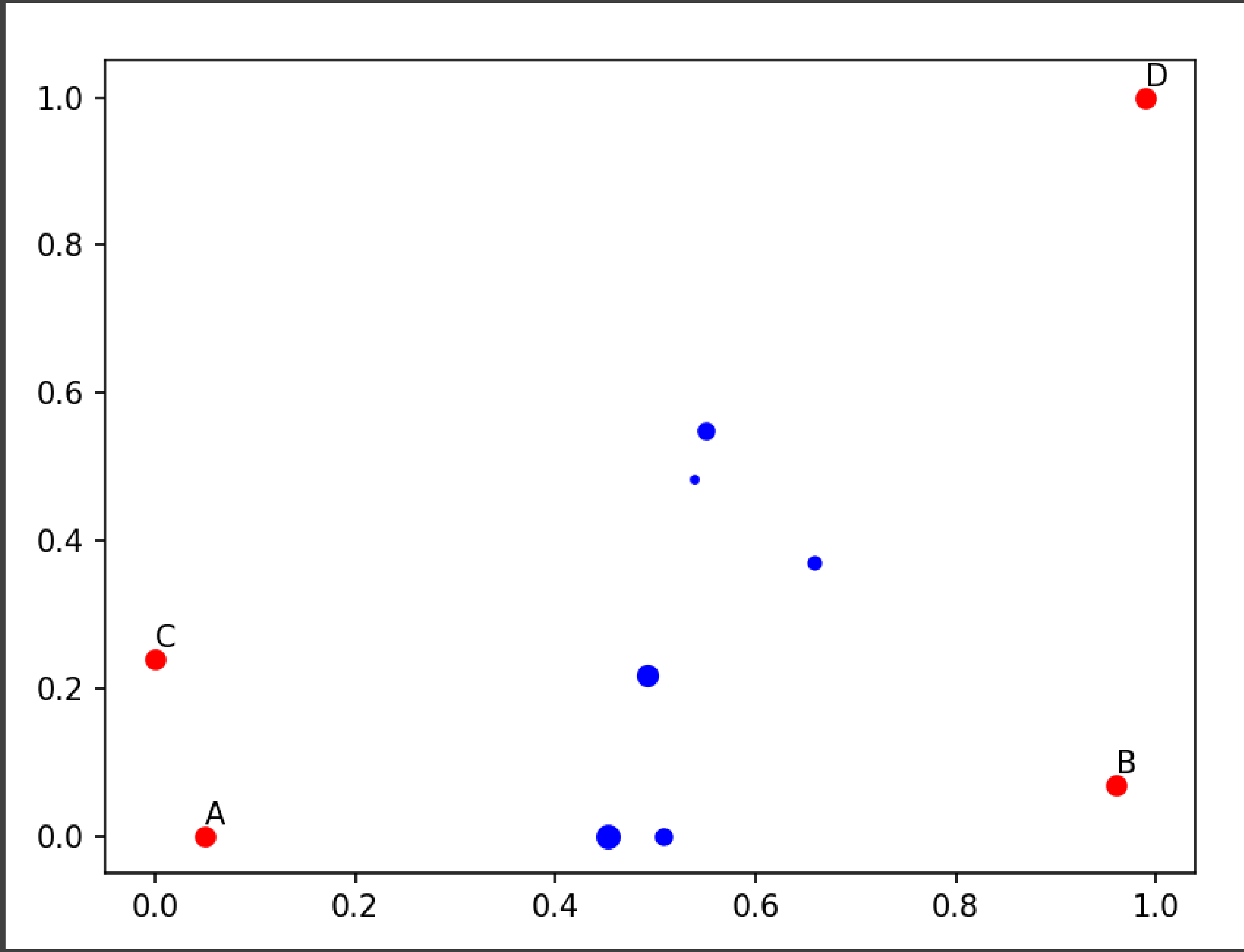
Nathalie Arthaud	François Asselineau	Jacques Cheminade	Nicolas Dupont-Aignan	François Fillon	Benoît Hamon	Jean Lassalle	Marine Le Pen	Emmanuel Macron	Jean-Luc Mélenchon	Philippe Poutou	official vote
								11			Emmanuel Macron
9									11	10	Jean-Luc Mélenchon
					10			11	9	8	Emmanuel Macron
5	2	4	3	7	11	6	1	9	10	8	Benoît Hamon
3	4	7	5	1	11	10	2	9	8	6	Emmanuel Macron
			6	7	10		5	8	11	9	Jean-Luc Mélenchon
					10			11			Emmanuel Macron
				10	9			11	8		Emmanuel Macron
4	3	7	2	8	9	5	1	11	10	6	Emmanuel Macron
					9	8		11	10		Emmanuel Macron
				9	10			11			Emmanuel Macron
9					10			7	11	8	Jean-Luc Mélenchon
6		5		4	11	7		10	8	9	Benoît Hamon
				11				10			François Fillon
			10	11		9					François Fillon
7		8			10			6	11	9	Jean-Luc Mélenchon
				11				10			François Fillon
			11		8	6		10	9	7	Nicolas Dupont-Aignan
6					9	8		7	11	10	Jean-Luc Mélenchon
					10	8		7	11	9	Jean-Luc Mélenchon
					9			11	10		Emmanuel Macron
9					8				11	10	Jean-Luc Mélenchon
									11	10	Jean-Luc Mélenchon
5	4	3	7	2	11	10	1	8	9	6	Benoît Hamon
6	2	8	3	4	11	5	1	9	10	7	Benoît Hamon
				10				11			Emmanuel Macron
					11			10	8	9	Benoît Hamon
					11			10			Benoît Hamon

PREFLIB

```
# FILE NAME: test.soc
# TITLE: Test Data
# DESCRIPTION:
# DATA TYPE: soc
# MODIFICATION TYPE: original
# RELATES TO:
# RELATED FILES:
# PUBLICATION DATE: 2023-06-05
# MODIFICATION DATE: 2023-06-05
# NUMBER ALTERNATIVES: 4
# NUMBER VOTERS: 35
# NUMBER UNIQUE ORDERS: 7
# ALTERNATIVE NAME 1: A
# ALTERNATIVE NAME 2: B
# ALTERNATIVE NAME 3: C
# ALTERNATIVE NAME 4: D
10: 1,2,3,4
5: 2,1,3,4
8: 2,3,1,4
3: 2,3,4,1
1: 3,2,4,1
5: 3,4,2,1
3: 4,3,2,1
```

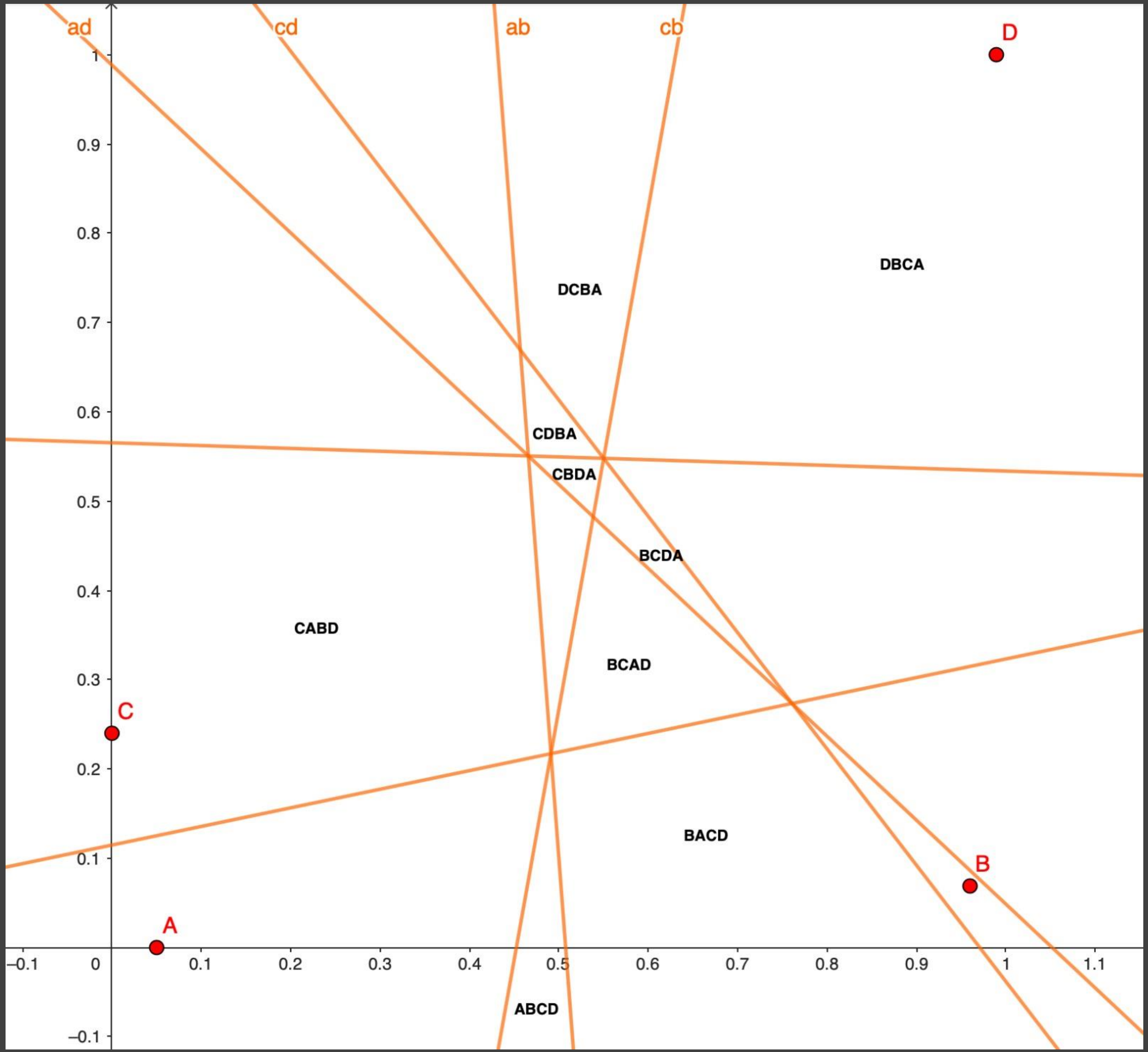
Problématique

La préférence d'un votant est décroissante avec sa distance aux candidats.



Problématique

On obtient plusieurs zones qui correspondent chacune à une préférence particulière.



Programmation mathématique (PM)

Les variables et les contraintes sont :

$0 \leq c_i \leq 1, \forall c \in C, \forall \mathrm{i} \in \{0,1\}$

$0 \leq v_i \leq 1, \forall v \in V, \forall \mathrm{i} \in \{0,1\}$

$d_c^v = \sqrt{\sum_{i=0}^{n-1}(c_i - v_i)^2}, \forall c \in C, \forall v \in V$ est la distance euclidienne.

$0 \leq \sigma^v_{(c,c')}, \forall v \in V, \forall (c, c') \in C^2$, tels que $c > c'$ pour le votant v

$d_c^v \leq d_{c'}^v + \sigma^v_{(c,c')}, \forall v \in V, \forall (c, c') \in C^2$ avec $c > c'$

La fonction objectif est :

$$\min\left(\sum_{v \in V} \sum_{(c,c') \in L^v} \sigma^v_{(c,c')}\right)$$

Non-metric multidimensional scaling (NMDS)

Les variables sont :

n : nombre de candidats

m : nombre de votants

M_{m+n} : matrice de dissimilarités symétrique

d_{ij} : distance euclidienne entre chaque paire (candidat, votant)

D_{ij} : rang du candidat j dans la préférence du votant i

Pour convertir sur une même échelle, on fait une régression linéaire : $\hat{d}_{ij} = \alpha + \beta D_{ij}$.

La valeur obtenue $\hat{d}_{ij}$ est la distance convertie en rang.

On cherche à minimiser la fonction :

$$\text{Stress} = \sqrt{\frac{\Sigma_{i,j}\left(d_{ij} - \hat{d}_{ij}\right)^2}{\Sigma_{i,j} d_{\mathrm{ij}}^2}}$$

Cette fonction mesure à quel point les rangs obtenus sont proches des vraies préférences en entrée.

Evaluation des méthodes

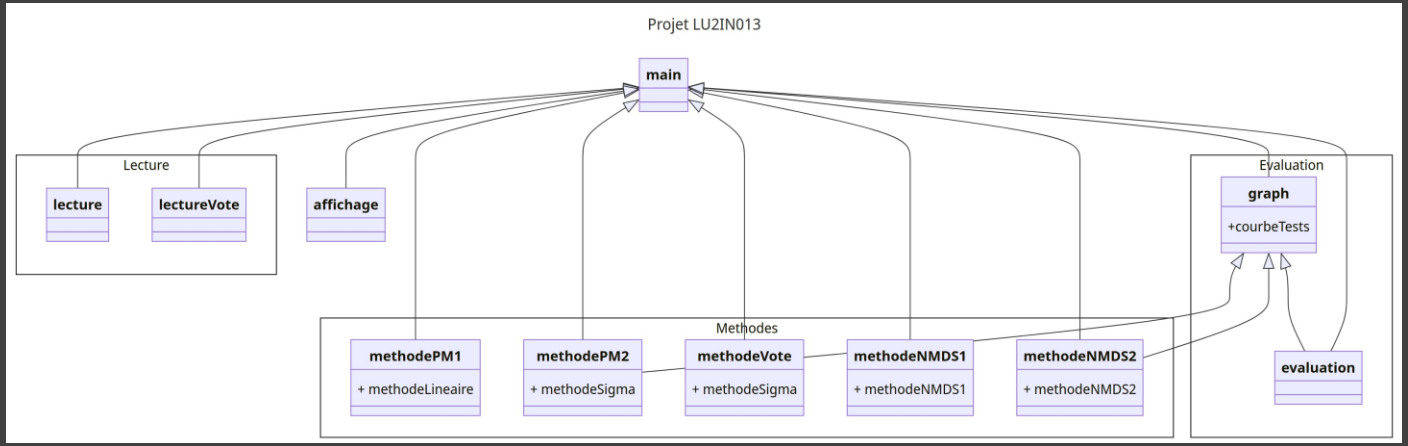
Tau de Kendall : $\tau = \frac{2}{mn(n-1)} \sum_{i=1}^{m} \mathrm{d}(R_i, R_i')$

- Proximité entre deux rangements
- Calcul du nombre de désaccords
- À valeur entre 0 (parfait respect des préférences en entrée) et 1

Conception & Implémentation

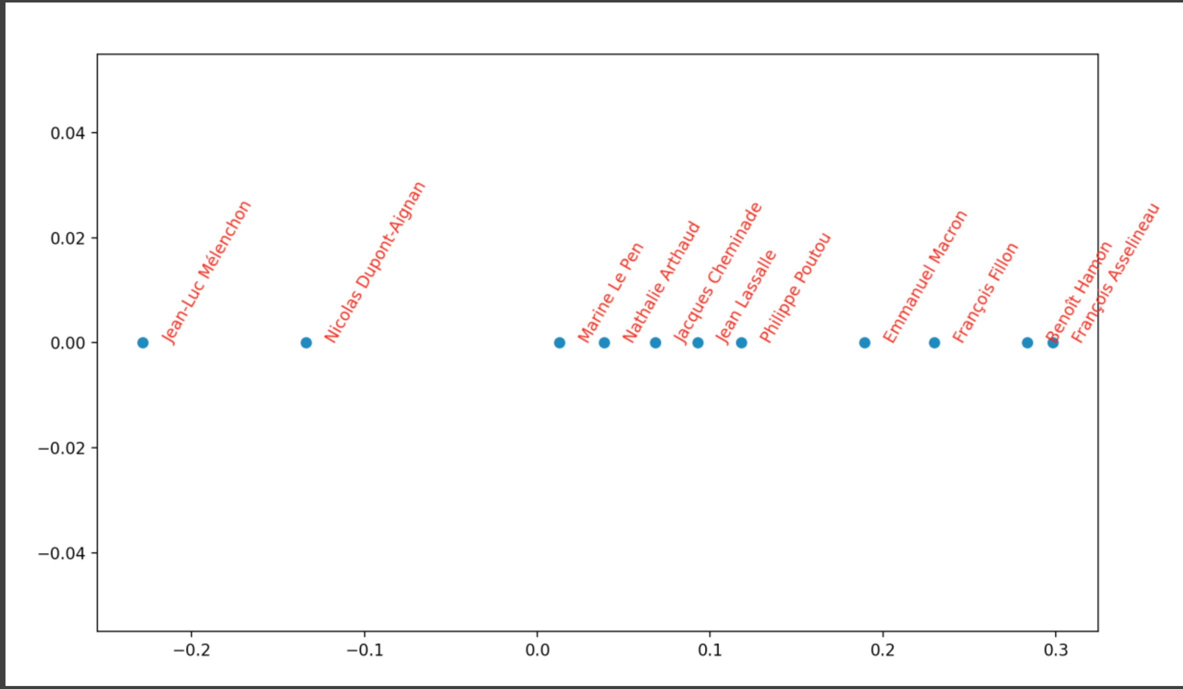
- Gurobi : API pour la programmation mathématique
- Bibliothèque python scikit-learn : pour NMDS (matrice de dissimilarité)

- Manipulation de structures : matrices, listes, dictionnaires

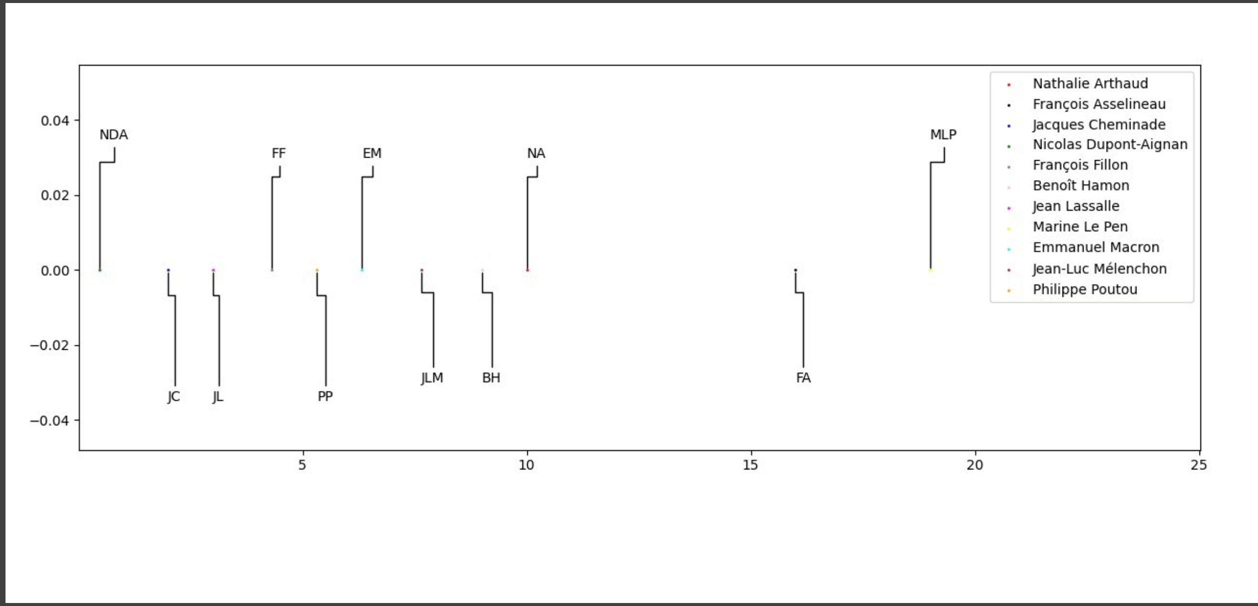


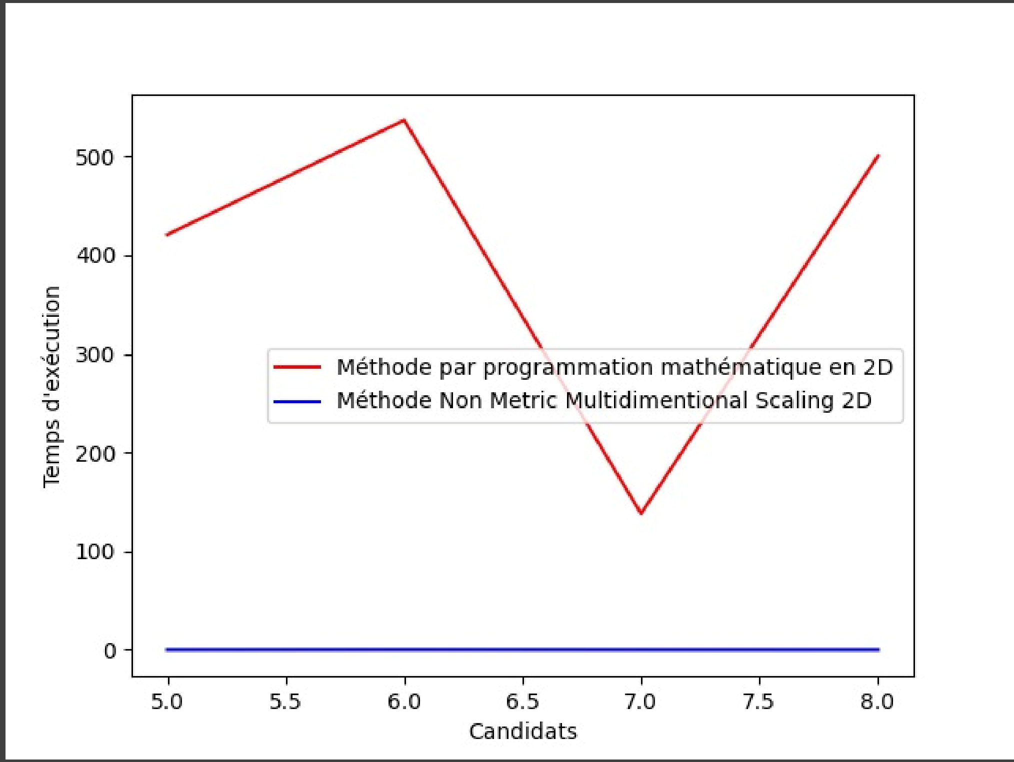
Données

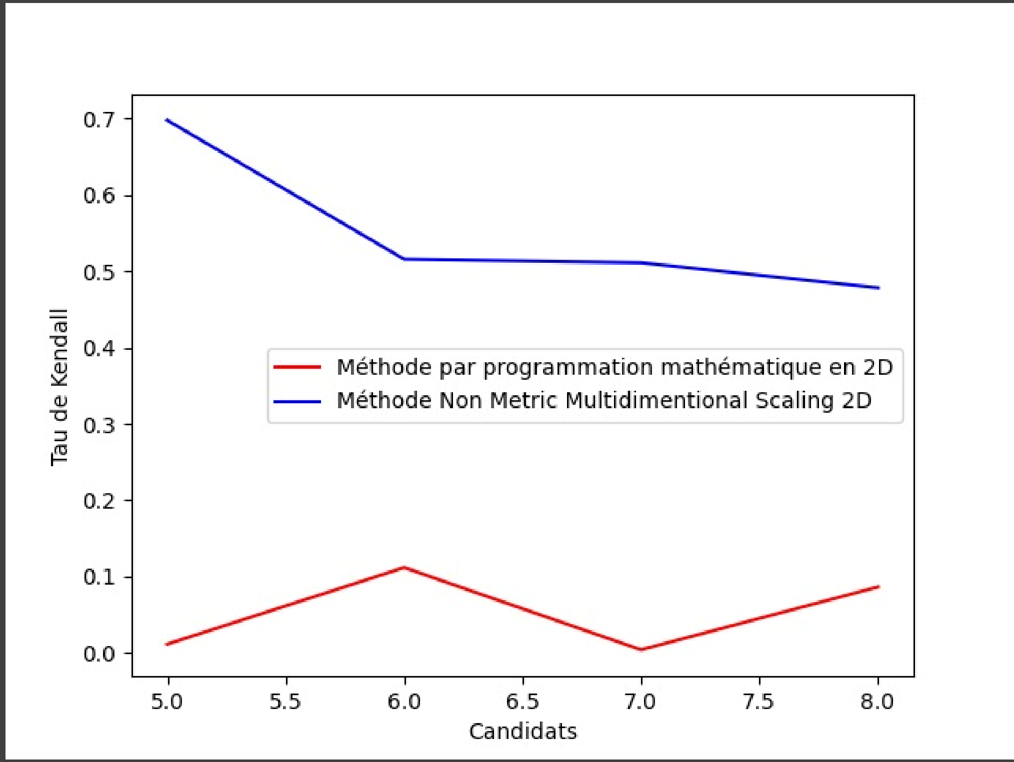
NON-METRIC MULTIDIMENTIONAL SCALING



PROGRAMMATION MATHÉMATIQUE



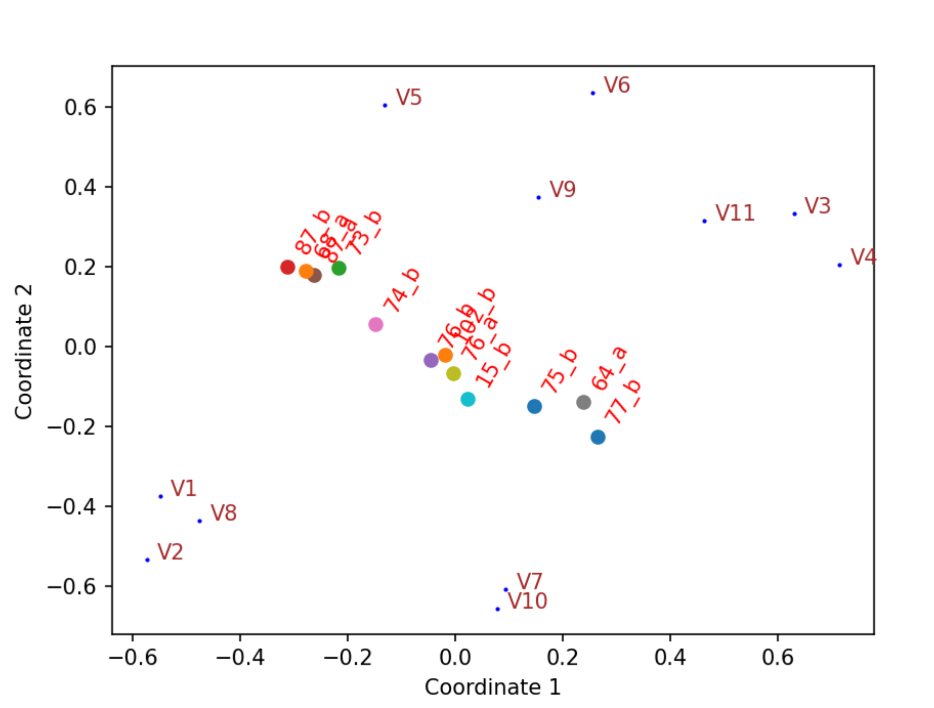




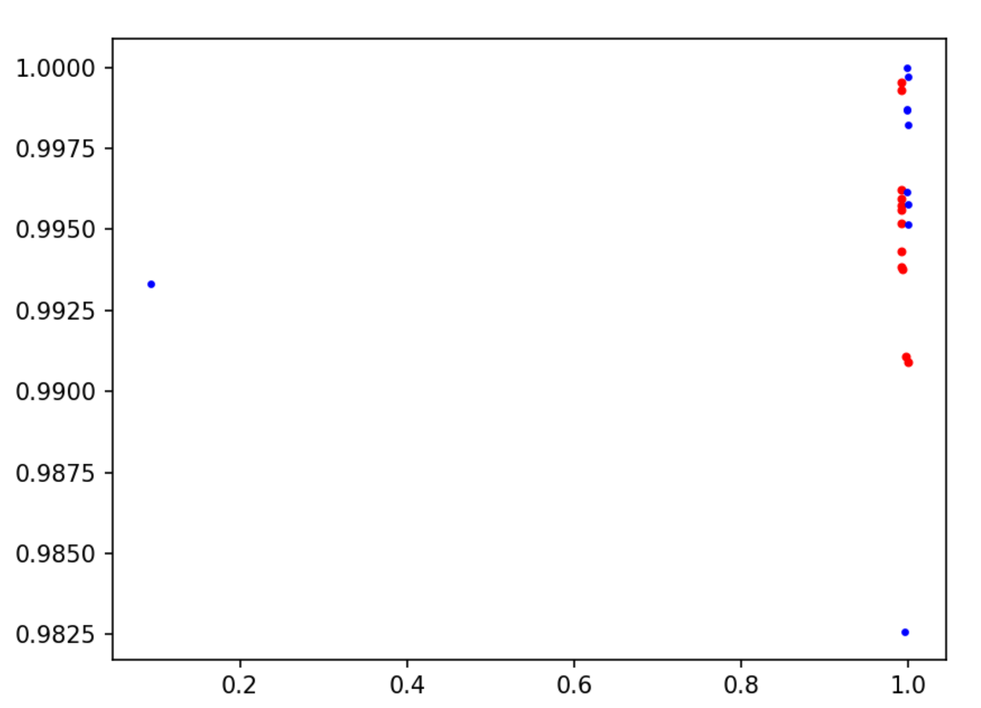
Comparaison entre les deux méthodes

Exemples de dégénérescence

NON-METRIC MULTIDIMENTIONAL SCALING



PROGRAMMATION MATHÉMATIQUE



Conclusion

SYNTHÈSE

- Deux approches différentes : une programmation mathématique et un algorithme de gradient (NMDS)
- Dégénérescence des deux côtés, temps de calculs différents

DIRECTION DE RECHERCHE

- Autre approche : minimiser le nombre de désaccords comme objectif
- Impact sur la complexité
- Gérer la dégénérescence en dimension 2 pour la programmation mathématique