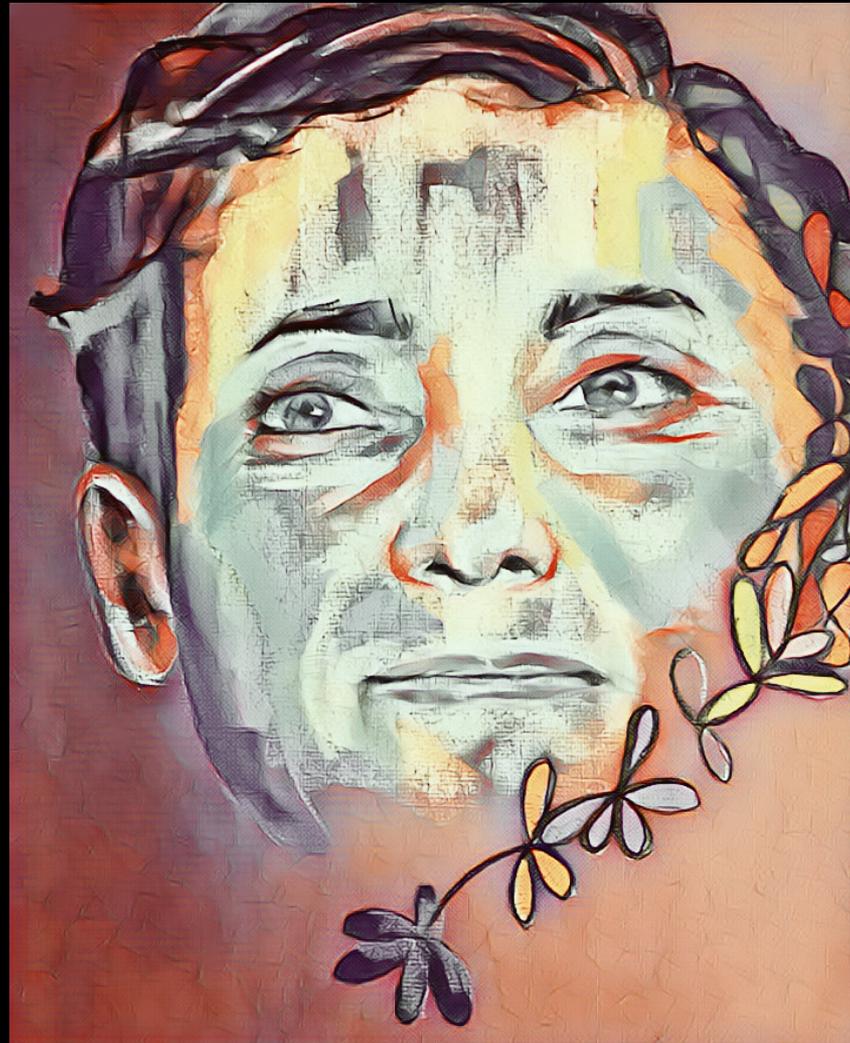


Prix junior Maryam Mirzakhani 2022





FM
JH

FONDATION MATHÉMATIQUE
JACQUES HADAMARD

Le mot de la présidente du Jury

Amandine Veber

Lauréate du prix junior Maryam Mirzakhani 2022

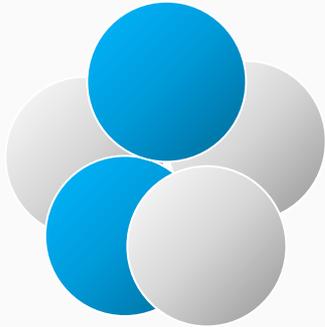
Angélique Campaniello

STAGE DE L3- LDD MATHS-PHYSIQUE (UNIVERSITÉ PARIS-SACLAY)

SOUS LA DIRECTION DE:

VITTORIO SOMÀ (CEA)

ALIX DELEPORTE (LMO)



PRIX MARYAM MIRZAKHANI JUNIOR



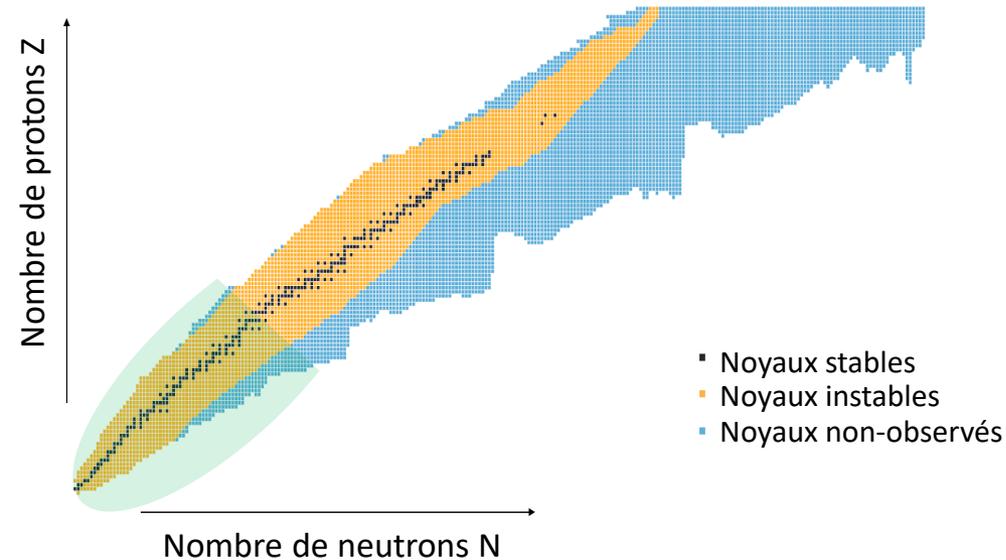
Application des techniques de factorisation de tenseurs aux calculs ab initio de structure nucléaire

Angélique Campaniello

université
PARIS-SACLAY

FACULTÉ
DES SCIENCES
D'ORSAY

Mini-introduction aux noyaux atomiques...



- N neutrons et Z protons en interaction forte
- ~260 noyaux stables
- Des milliers de noyaux instables dont ~ 3000 synthétisés et étudiés
- **Objectif:** étude expérimentale et théorique
- **Problème:** étude théorique ab initio limitée à $N+Z < 100$



Motivation

Résoudre équation de Schrödinger aux valeurs propres pour un opérateur auto-adjoint : $H\psi = E\psi$

Dans un noyau à N+Z nucléons:

$$H = T_1 + V_2 + V_3 + \dots$$

Opérateurs à A corps représenté par tenseur à 2A indices : T_{ij} V_{ijkl} V_{ijklmn}

Motivation

Résoudre équation de Schrödinger aux valeurs propres pour un opérateur auto-adjoint : $H\psi = E\psi$

Dans un noyau à N+Z nucléons:

$$H = T_1 + V_2 + V_3 + \dots$$

Opérateurs à A corps représenté par tenseur à 2A indices : T_{ij} V_{ijkl} V_{ijklmn}

Objectif: Prédire des quantité physiques:

- Énergie de liaison
- Rayon
- ...

Motivation

Résoudre équation de Schrödinger aux valeurs propres pour un opérateur auto-adjoint : $H\psi = E\psi$

Dans un noyau à N+Z nucléons:

$$H = T_1 + V_2 + V_3 + \dots$$

Opérateurs à A corps représenté par tenseur à 2A indices : T_{ij} V_{ijkl} V_{ijklmn} → état de la base de l'espace de Hilbert à 1 particule

- Objectif:** Prédire des quantité physiques:
- Énergie de liaison
 - Rayon
 - ...

Dimension du tenseur : n^{2A}

Dimension la base de l'espace de Hilbert à 1 particule
→ infinie donc tronquée à n en pratique

Problème: dimension de la base à A corps → **exponentielle!**

Interaction	Taille
1 corps	200 kB
2 corps	~1GB
3 corps	~10TB

Motivation

Résoudre équation de Schrödinger aux valeurs propres pour un opérateur auto-adjoint : $H\psi = E\psi$

Dans un noyau à N+Z nucléons:

$$H = T_1 + V_2 + V_3 + \dots$$

Opérateurs à A corps représenté par tenseur à 2A indices : T_{ij} V_{ijkl} V_{ijklmn} → état de la base de l'espace de Hilbert à 1 particule

Objectif: Prédire des quantité physiques:

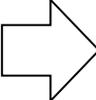
- Énergie de liaison
- Rayon
- ...

Dimension du tenseur : n^{2A}

Dimension la base de l'espace de Hilbert à 1 particule
→ infinie donc tronquée à n en pratique

Problème: dimension de la base à A corps → **exponentielle!**

Interaction	Taille
1 corps	200 kB
2 corps	~1GB
3 corps	~10TB

Solution  Compression (factorisation) de tenseurs

Décompositions

Décomposition en valeurs singulières (DVS)

→réduction de rang de matrice (tenseur 2D): **marche très bien**

→ pour $A \in \mathbb{C}^{m \times n}$ matrice de rang r , alors $A = \sum_{j=1}^r \sigma_j u_j v_j^* = UDV^*$

Décompositions

Décomposition en valeurs singulières (DVS)

→réduction de rang de matrice (tenseur 2D): **marche très bien**

→ pour $A \in \mathbb{C}^{m \times n}$ matrice de rang r , alors $A = \sum_{j=1}^r \sigma_j u_j v_j^* = UDV^*$

Norme de Frobenius:

→Pour un tenseur $V \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$:

$$\|V\|_F = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \dots \sum_{i_N=1}^{I_N} V_{i_1 i_2 \dots i_N} V_{i_1 i_2 \dots i_N}^*$$

Décompositions

Décomposition en valeurs singulières (DVS)

→réduction de rang de matrice (tenseur 2D): **marche très bien**

→ pour $A \in \mathbb{C}^{m \times n}$ matrice de rang r , alors $A = \sum_{j=1}^r \sigma_j u_j v_j^* = UDV^*$

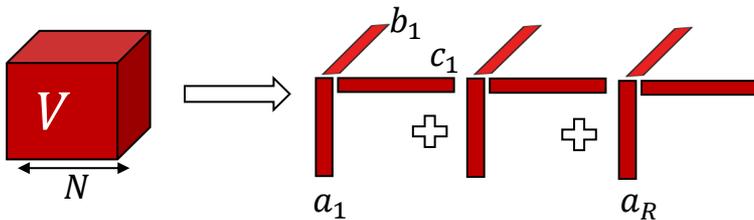
Norme de Frobenius:

→Pour un tenseur $V \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$:

$$\|V\|_F = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \dots \sum_{i_N=1}^{I_N} V_{i_1 i_2 \dots i_N} V_{i_1 i_2 \dots i_N}^*$$

Canonical Polyadic (CP)

$$V \approx \sum_{r=1}^R a_r \circ b_r \circ c_r$$



Taille après décomposition: DRN

Décompositions

Décomposition en valeurs singulières (DVS)

→réduction de rang de matrice (tenseur 2D): **marche très bien**

→ pour $A \in \mathbb{C}^{m \times n}$ matrice de rang r , alors $A = \sum_{j=1}^r \sigma_j u_j v_j^* = UDV^*$

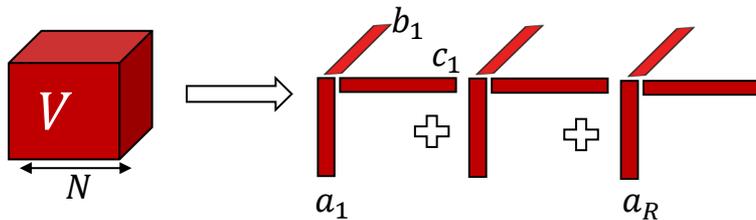
Norme de Frobenius:

→Pour un tenseur $V \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$:

$$\|V\|_F = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \dots \sum_{i_N=1}^{I_N} V_{i_1 i_2 \dots i_N} V_{i_1 i_2 \dots i_N}^*$$

Canonical Polyadic (CP)

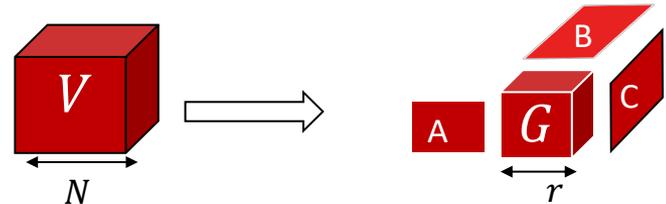
$$V \approx \sum_{r=1}^R a_r \circ b_r \circ c_r$$



Taille après décomposition: DRN

Tucker

$$V \approx G \times_1 A \times_2 B \times_3 C$$



Taille après décomposition: $r^D + DrN$

Divergence CP?

Tenseur de rang 3 $V = a_1 \circ b_1 \circ c_2 + a_1 \circ b_2 \circ c_1 + a_2 \circ b_1 \circ c_1$

Approximé par un tenseur de rang 2 $\tilde{V} = \alpha \left(a_1 + \frac{1}{\alpha} a_2 \right) \circ \left(b_1 + \frac{1}{\alpha} b_2 \right) \circ \left(c_1 + \frac{1}{\alpha} c_2 \right) - \alpha a_1 \circ b_1 \circ c_1$

$$\|V - \tilde{V}\|_F = \frac{1}{\alpha} \|a_2 \circ b_2 \circ c_1 + a_2 \circ b_1 \circ c_2 + a_1 \circ b_2 \circ c_2 + \frac{1}{\alpha} a_2 \circ b_2 \circ c_2\|_F$$

$$\begin{aligned} \tilde{V} &\rightarrow V \\ \alpha &\rightarrow +\infty \end{aligned} \quad \text{Diverge!}$$

Algorithmes

Canonical Polyadic (CP)-ALS

On cherche à minimiser $\|V - f(A, B, C)\|_F^2$



On fixe B et C pour trouver A etc

Problème: inversion de matrice

Solution: Descente de gradient avec régularisation

On cherche à minimiser

$$\gamma(A) = \|V - f(A, B, C)\|_F^2 + \underbrace{\epsilon \|A\|_F^2}_{\text{régularisation}}$$

régularisation: Si $A \gg$, on ne minimise plus $\gamma(A)$

Algorithmes

Canonical Polyadic (CP)-ALS

On cherche à minimiser $\|V - f(A, B, C)\|_F^2$

On fixe B et C pour trouver A etc

Problème: inversion de matrice

Solution: Descente de gradient avec régularisation

On cherche à minimiser

$$\gamma(A) = \|V - f(A, B, C)\|_F^2 + \underbrace{\epsilon \|A\|_F^2}_{\text{régularisation}}$$

régularisation: Si $A \gg$, on ne minimise plus $\gamma(A)$

Tucker-HOOI (ALS)

On cherche à minimiser $\|V - f(G, A, B, C)\|_F^2$

On fixe B et C pour trouver A etc

Pas de problème: matrices orthogonales

Algorithmes

Canonical Polyadic (CP)-ALS

On cherche à minimiser $\|V - f(A, B, C)\|_F^2$

On fixe B et C pour trouver A etc

Problème: inversion de matrice

Solution: Descente de gradient avec régularisation

On cherche à minimiser

$$\gamma(A) = \|V - f(A, B, C)\|_F^2 + \epsilon \|A\|_F^2$$

régularisation: Si $A \gg$, on ne minimise plus $\gamma(A)$

Tucker-HOOI (ALS)

On cherche à minimiser $\|V - f(G, A, B, C)\|_F^2$

On fixe B et C pour trouver A etc

Pas de problème: matrices orthogonales

	twoJ	twoT	MT	PI	nlj1	nlj2	nnlj1	nnlj2	nlj1x	nlj2x	nnlj1x	nnlj2x	symcoeff	inter
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.000000
1	0	2	-1	0	0	0	0	0	0	0	0	0	0	0
2	0	2	0	0	0	0	0	0	0	0	0	0	0	0
3	0	2	1	0	0	0	0	0	0	0	0	0	0	0
4	2	0	0	0	0	0	0	0	0	0	0	0	0	0
...
4203	8	2	1	0	5	5	5	5	5	5	5	5	5	5
4204	10	0	0	0	5	5	5	5	5	5	5	5	5	5
4205	10	2	-1	0	5	5	5	5	5	5	5	5	5	5
4206	10	2	0	0	5	5	5	5	5	5	5	5	5	5
4207	10	2	1	0	5	5	5	5	5	5	5	5	1	-0.000000

L'interaction à 2 corps en pratique

$$V_{abcd} \rightarrow V_{(nlj)_a, (nlj)_b, (nlj)_c, (nlj)_d}^{J T M T}$$

Tenseur à 4 indices pour chaque J, T et MT

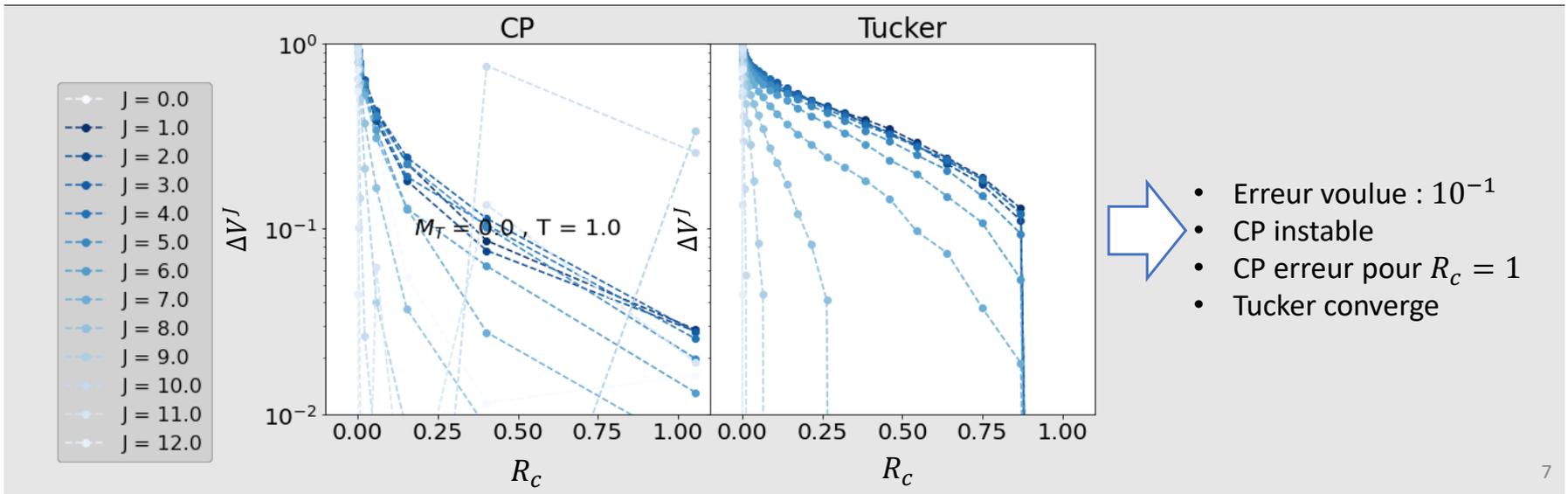
Résultats:

- Estimation de l'erreur

$$\Delta V = \frac{\|\tilde{V} - V\|_F}{\|V\|_F}$$

- Estimation du facteur de compression

$$R_c = \frac{\text{Nouvelle taille}}{\text{Ancienne taille}}$$



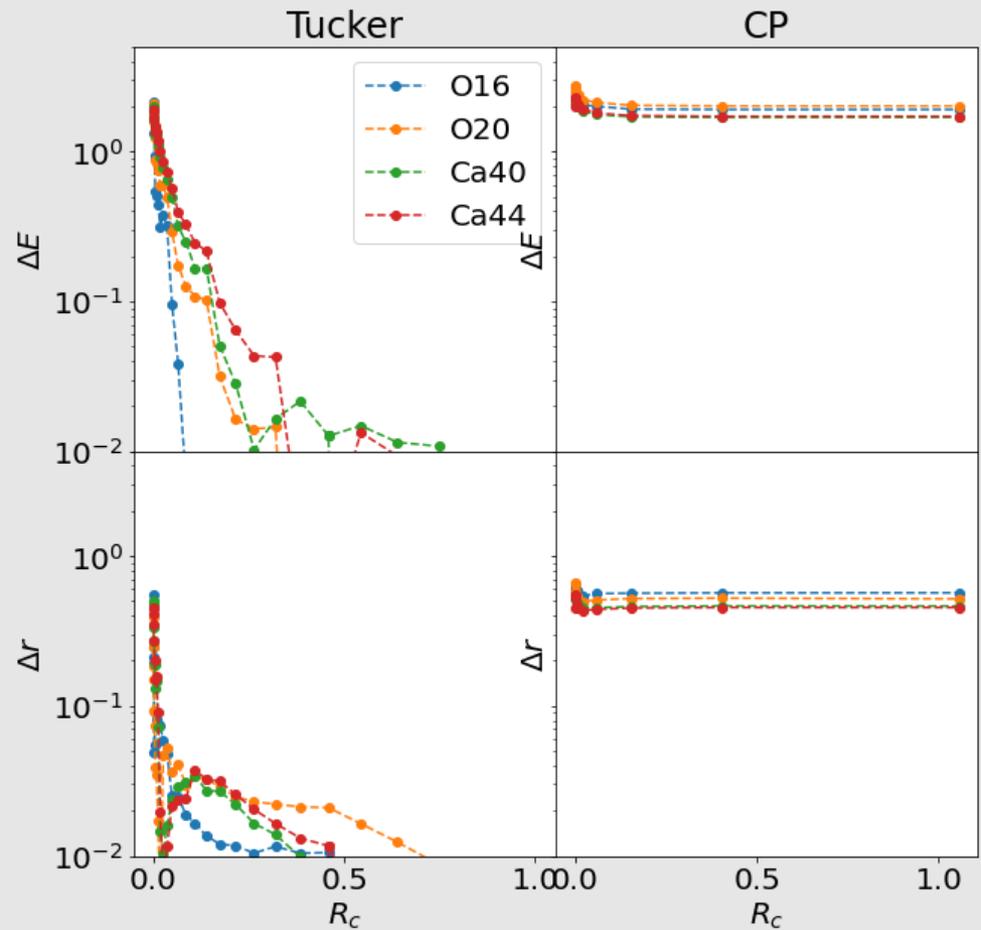
Résultats

- Observables
 - Énergie du fondamental E

$$\Delta E = \frac{\| \tilde{E} - E \|_F}{\| E \|_F}$$

- Rayon moyen r

$$\Delta r = \frac{\| \tilde{r} - r \|_F}{\| r \|_F}$$



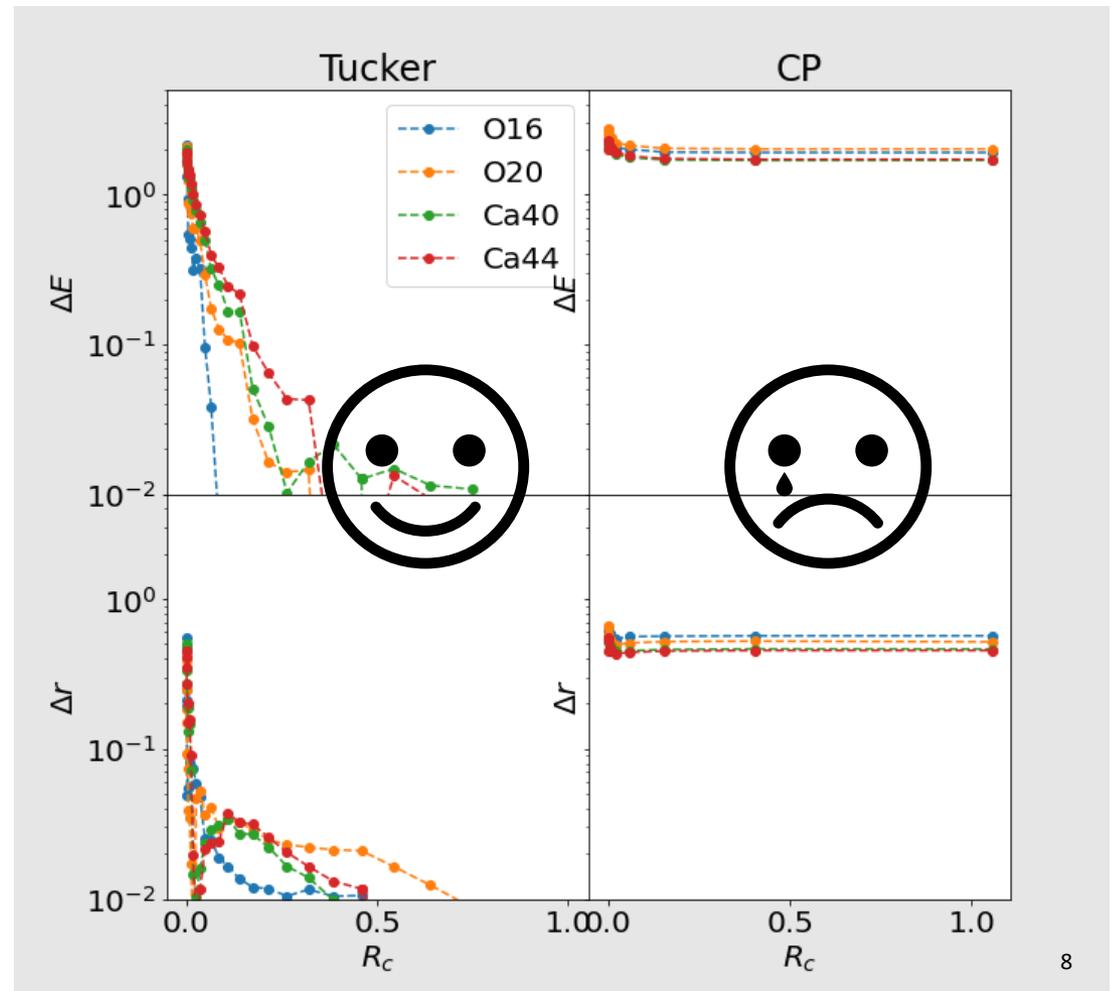
Résultats

- Observables
 - Énergie du fondamental E

$$\Delta E = \frac{\| \tilde{E} - E \|_F}{\| E \|_F}$$

- Rayon moyen r

$$\Delta r = \frac{\| \tilde{r} - r \|_F}{\| r \|_F}$$



Conclusion

- Décomposition de l'interaction à 2 corps:
 - CP converge moins bien 😞
 - Tucker: réussite! 😊
- Aller plus loin:
 - Base avec plus d'états → (dimension réaliste 10^3)
 - Tester d'autres décompositions
 - Améliorer gradient CP
 - **Décomposer l'interaction à 3 corps = élément bloquant**

Conclusion

- Décomposition de l'interaction à 2 corps:
 - CP converge moins bien 😞
 - Tucker: réussite! 😊
- Aller plus loin:
 - Base avec plus d'états → (dimension réaliste 10^3)
 - Tester d'autres décompositions
 - Améliorer gradient CP
 - Décomposer l'interaction à 3 corps = élément bloquant

**Merci pour
votre attention !**

Lauréate du prix junior Maryam Mirzakhani 2022

Audrey Antoine

Le polynôme de Jones pour les nœuds

Audrey ANTOINE

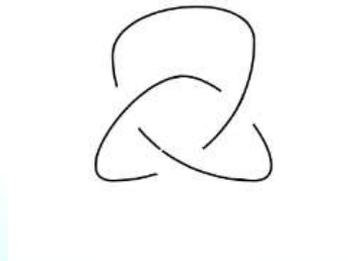
Sommaire

- Définition d'un nœud
- Représentation d'un nœud
- L'équivalence de deux nœuds au travers des mouvements de Reidemeister
- Un invariant de nœuds: le polynôme de Jones

Définitions

Définition 1. (Entrelacs) Un entrelacs L constitué de m composantes est un sous-ensemble de \mathbb{S}^3 , ou de \mathbb{R}^3 , comportant m courbes fermées simples, linéaires par morceaux et disjointes.

Définition 2. (Noeud) Un noeud est un entrelacs contenant une seule composante connexe.



Noeud de trèfle

Représentation d'un nœud

La représentation des noeuds se fait à travers des diagrammes en deux dimensions respectant les règles suivantes :

1. La courbe doit être une immersion, c'est à dire que sa dérivée ne s'annule pas. Cela lui donne une apparence lisse.
2. Deux portions de courbe ne peuvent être tangentes
3. Le diagramme ne peut contenir un point triple

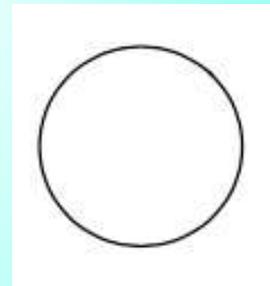
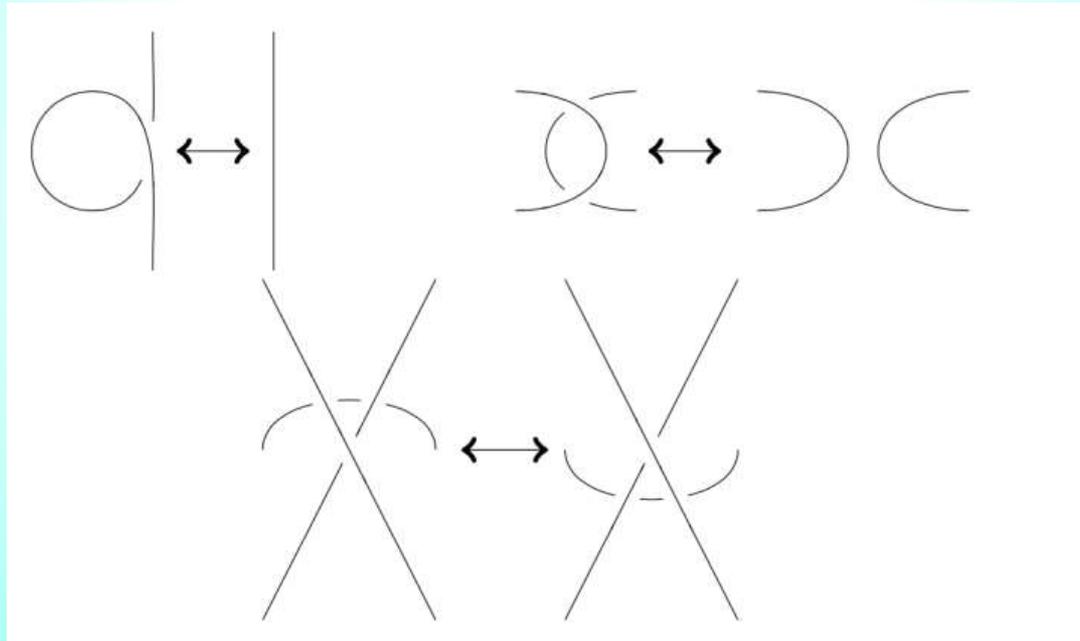


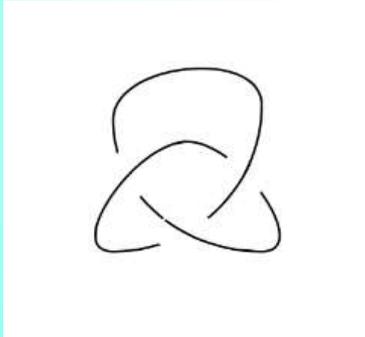
Diagramme Nul

Les mouvements de Reidemeister

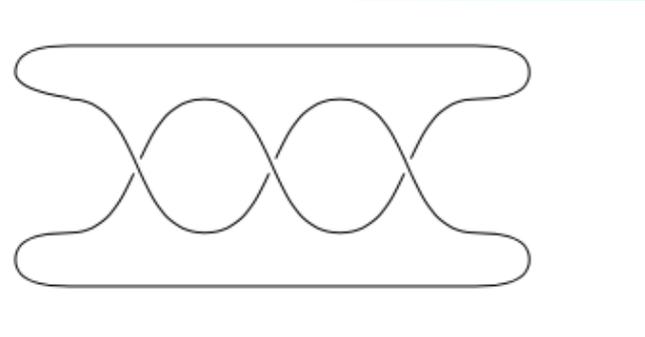
RI, RII et RIII



Théorème 1.1. *Deux diagrammes d'entrelacs représentent le même entrelacs, si et seulement si on peut passer de l'un à l'autre par une suite de mouvements de Reidemeister.*



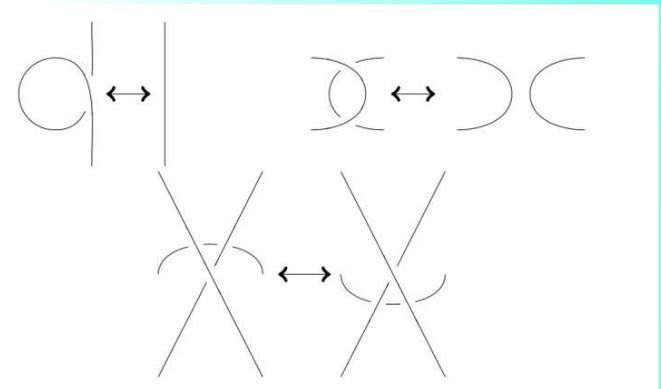
Noeud de trèfle



Noeud de trèfle sous forme de tresse

Un invariant de nœud: le polynôme de Jones

Types de croisements



Définition 3. (Entortillement) En associant 1 aux croisements positifs et -1 aux croisements négatifs, on définit l'entortillement comme l'application w qui à tout diagramme D d'un noeud associe $w(D)$ la somme de ses croisements.

Définition 4. (Crochet de Kauffman) Pour D un diagramme d'entrelacs non vide, son polynôme de Kauffman $\langle D \rangle \in \mathbb{Z}[A, B, d]$ est donné par les relations suivantes :

1. $\langle \bigcirc \rangle = 1$
2. $\langle D \sqcup \bigcirc \rangle = d \langle D \rangle$
3. $\langle \text{X} \rangle = A \langle \text{Y} \rangle + B \langle \text{Z} \rangle$

où \bigcirc est le diagramme trivial et \sqcup désigne le fait de juxtaposer deux entrelacs sans créer de nouveau croisement entre eux. Pour que le crochet de Kauffman soit invariant par le mouvement de Reidemeister II on verra qu'il faut avoir $B = A^{-1}$ et $d = -A^2 - A^{-2}$.

Définition 5. (Polynôme de Jones) On définit, pour tout diagramme K d'un noeud, son polynôme de Jones $V(K) = (-A)^{-3w(K)} \langle K \rangle$.

Lauréate du prix junior Maryam Mirzakhani 2022

Margaux Liechty

Catégories de partitions
TER réalisé avec Nailya Manatova et TERENCE Viaud, encadré par
Amaury Freslon

Margaux Liechty

31 Août 2022

1 Définitions

- Partition
- Catégorie de partitions

2 Catégories de partitions non-croisées

- Classification

Partition

Définition

Soient $k, \ell \in \mathbb{N}$. Une *partition* p de k points supérieurs et ℓ points inférieurs est une partition de l'ensemble

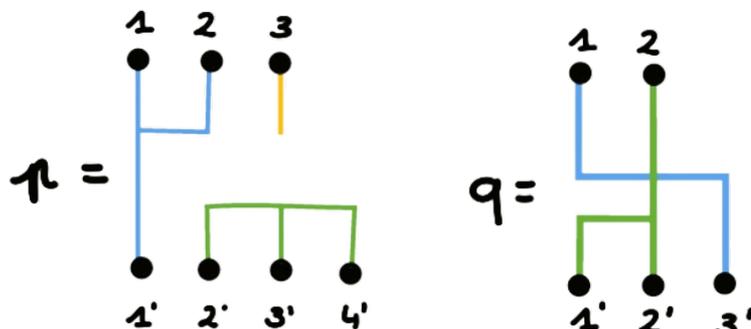
$$\{1, \dots, k\} \cup \{1, \dots, \ell\} \simeq \{1, \dots, k + \ell\}$$

i.e. une décomposition de cet ensemble en sous-ensembles non vides et deux à deux disjoints, appelés des *blocs*.

Notation : On note $\mathcal{P}(k, \ell)$ l'ensemble des partitions de k points inférieurs et ℓ points inférieurs.

On note $\mathcal{P} := \bigcup_{k, \ell \in \mathbb{N}} \mathcal{P}(k, \ell)$.

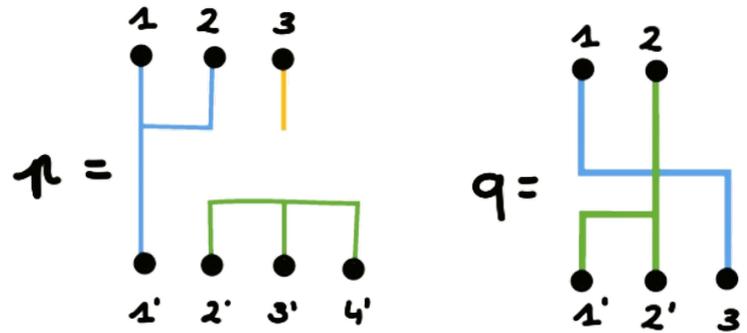
Exemples : $p = \{\{1, 2, 1'\}, \{3\}, \{2', 3', 4'\}\} \in \mathcal{P}(3, 4)$ et $q = \{\{1, 3'\}, \{2, 1', 2'\}\} \in \mathcal{P}(2, 3)$.



On appelle *singleton* un bloc qui ne contient qu'un seul point.

On appelle *intervalle* un bloc formé de points consécutifs.

Partition non-croisée



Graphiquement, une partition non-croisée est une partition dont aucun traits ne se croisent.

Définition

Une partition $p = \{B_1, \dots, B_r\}$ sur une seule ligne est dite *croisée* s'il existe $k_1 < k_2 < k_3 < k_4$ des points de p tels que

- 1 k_1 et k_3 sont dans le même bloc B_i ;
- 2 k_2 et k_4 sont dans le même bloc B_j ;
- 3 k_1 et k_2 ne sont pas dans le même bloc.

On appelle le 4-uplet $K := (k_1, k_2, k_2, k_4)$ un *croisement* ; une *partition non-croisée* est alors une partition qui ne contient pas de croisement.

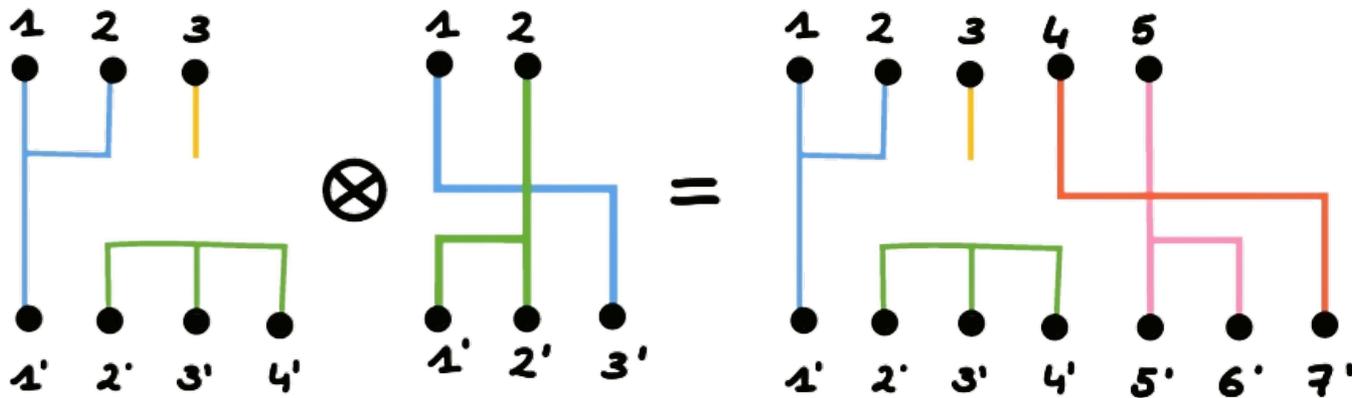
Notation : On note $\mathcal{P}^{nc} = \bigcup_{k, \ell \in \mathbb{N}} \mathcal{P}^{nc}(k, \ell)$ l'ensemble de toutes les partitions non-croisées.

Produit tensoriel

On définit maintenant quatre opérations sur les partitions que l'on appelle les *opérations de catégories*.

Produit tensoriel

Le *produit tensoriel* de deux partitions $p \in \mathcal{P}(k, \ell)$ et $q \in \mathcal{P}(k', \ell')$ est la partition $p \otimes q \in \mathcal{P}(k + k', \ell + \ell')$ obtenue en concaténant horizontalement p et q .

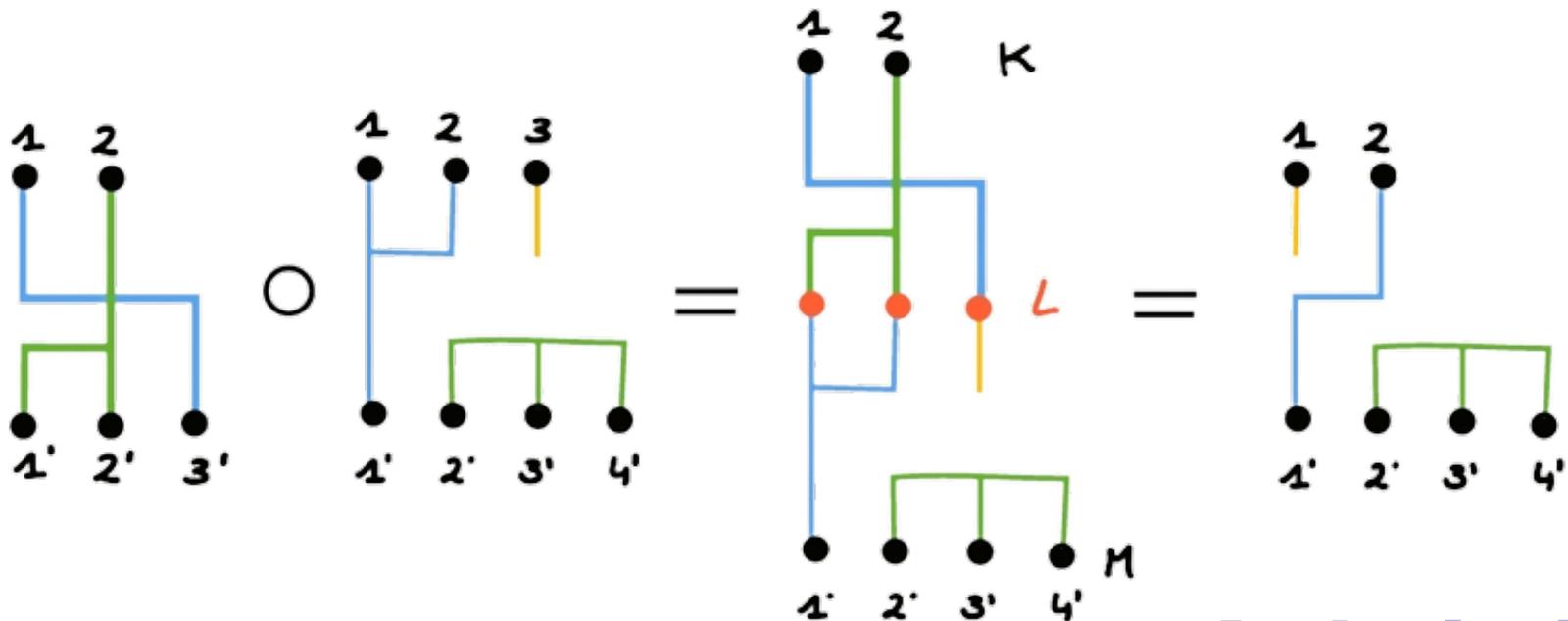


Composition

Composition

La *composition* de deux partitions $p \in \mathcal{P}(k, \ell)$ et $q \in \mathcal{P}(\ell, m)$ est la partition $p \circ q \in \mathcal{P}(k, m)$ obtenue de la manière suivante :

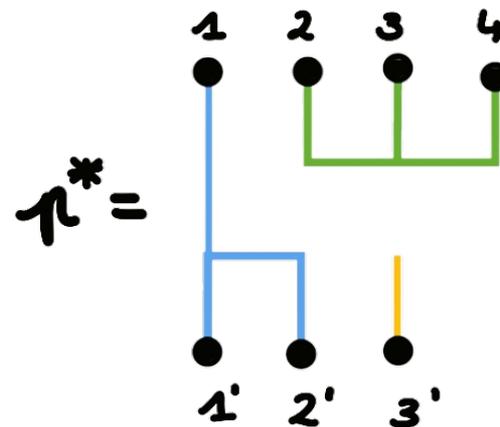
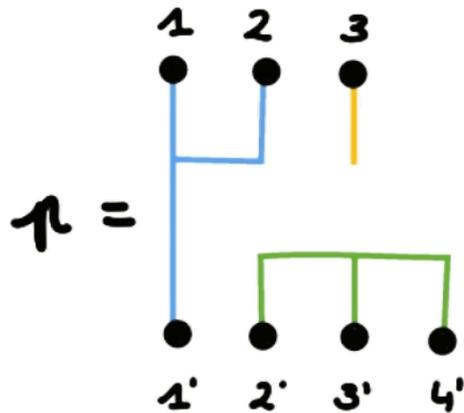
- 1 On concatène verticalement p sur q pour obtenir un graphe composé de trois lignes de points K , L et M , de cardinal respectif k , ℓ et m , tel que les points de K et L soient reliés comme dans p et les points de L et M soient reliés comme dans q ;
- 2 on localise les points de K et M appartenant aux mêmes blocs ;
- 3 on supprime la ligne L et on relie les points de K et de M qui sont dans les mêmes blocs pour obtenir une partition qui est bien dans $\mathcal{P}(k, m)$.



Involution

Involution

L'*involution* d'une partition $p \in \mathcal{P}(k, \ell)$ est la partition $p^* \in \mathcal{P}(\ell, k)$ obtenue en inversant les points inférieurs et supérieurs dans p sans changer les traits qui les relient.



Rotation

Rotation(s)

Soit $p \in \mathcal{P}(k, \ell)$ une partition. On déplace un point extrémal supérieur gauche [resp. droit] pour qu'il devienne un point extrémal inférieur gauche [resp. droit], ce qui nous donne une partition de $\mathcal{P}(k - 1, \ell + 1)$ appelée *rotation de la partition p*.

On peut aussi faire le contraire : rendre supérieur un point extrémal inférieur pour obtenir une partition de $\mathcal{P}(k + 1, \ell - 1)$.

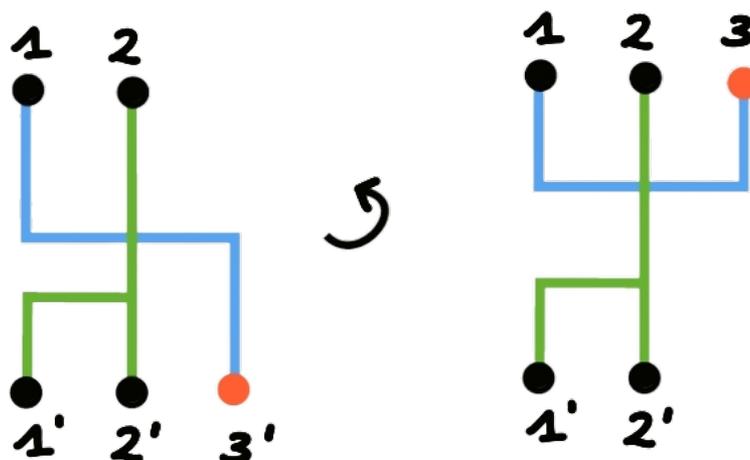


Figure – Rotation du point extrémal inférieur droit de q .

Catégorie de partitions

On peut munir l'ensemble \mathcal{P} d'une structure.

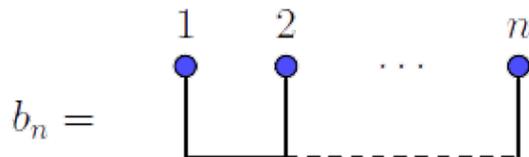
Catégorie de partitions

Une *catégorie de partitions* \mathcal{C} est une collection d'ensembles de partitions $\mathcal{C}(k, \ell) \subset \mathcal{P}(k, \ell)$ (pour tout entiers k, ℓ) stable par les quatre opérations de catégories et qui contient la partition $| \in \mathcal{P}(1, 1)$, appelée *partition identité*.

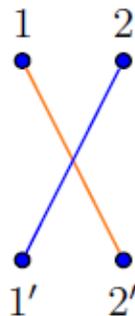
\mathcal{P} est une catégorie de partition appelée *catégorie de toutes les partitions*.

Si p_1, \dots, p_n sont des partitions, on désigne par $\langle p_1, \dots, p_n \rangle$ la plus petite catégorie de partition qui contient p_1, \dots, p_n . On dit que p_1, \dots, p_n sont des *générateurs* de $\langle p_1, \dots, p_n \rangle$.

Par exemple, si on note b_n ($n \geq 1$) l'intervalle constitué de n points :



on peut montrer que $\mathcal{P} = \langle b_3, X \rangle$, avec $X \in \mathcal{P}(2, 2)$.



Catégories de partitions non-croisées

Définition

Une *catégorie de partitions non-croisées* est une catégorie de partitions dont tous les $\mathcal{C}(k, \ell)$ sont dans \mathcal{P}^{nc} .

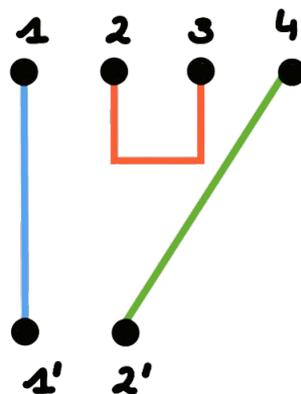
Par exemple l'ensemble de toutes les partitions non-croisées \mathcal{P}^{nc} est une catégorie de partitions non-croisées (c'est la plus grande au sens de l'inclusion).

Voici un autre exemple,

2-partition

On appelle *2-partition*, une partition dont tous les blocs contiennent exactement deux points.

\mathcal{P}_2^{nc} qui est l'ensemble de toutes les 2-partitions non-croisées est une catégorie de partitions non-croisées. C'est la plus petite catégorie de partitions non-croisées au sens de l'inclusion car $\mathcal{P}_2^{nc} = \langle | \rangle$.



Classification

Théorème (Banica-Speicher, Weber)

Il y a exactement sept catégories de partitions non-croisées.

- $\mathcal{P}^{nc} = \langle b_3 \rangle$,
- $\mathcal{P}_2^{nc} = \langle | \rangle$,
- $\mathcal{P}_{1,2}^{nc} = \langle \cdot \rangle$,
- $\mathcal{P}_{pair}^{nc} = \langle b_4 \rangle$,
- $\mathcal{P}^{nc'} = \langle b_4, : \rangle = \langle b_3 \otimes \cdot \rangle$,
- $\mathcal{P}_{1,2}'^{nc} = \langle p^\bullet \rangle$,
- $\mathcal{P}_{1,2}^{nc\#} = \langle : \rangle = \langle \cdot \otimes \cdot \rangle$.

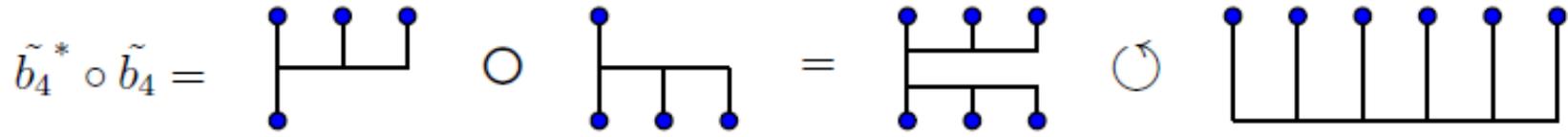
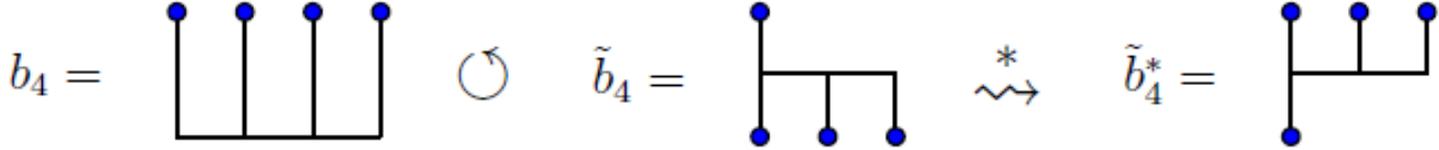
Avec p^\bullet la partition suivante



Démonstration de $\mathcal{P}_{pair}^{nc} = \langle b_4 \rangle$

- 1 On montre que $\mathcal{P}_{pair}^{nc} \subset \langle b_4 \rangle$.
- 2 On montre que \mathcal{P}_{pair}^{nc} est une catégorie de partitions.
- 3 Comme $b_4 \in \mathcal{P}_{pair}^{nc}$, et que \mathcal{P}_{pair}^{nc} est une catégorie de partitions, l'égalité suit.

Point 1.



Point 2.

Lemme

Soit $p \in \mathcal{P}(n)$ une partition non-croisée sur une seule ligne. Alors p possède un *intervalle*.

Seule la stabilité par composition a besoin d'être explicitée. Soit p et q dans $\mathcal{P}_{\text{pair}}^{nc}$ de taille compatible avec p sur une seule ligne. On procède par récurrence sur le nombre de points de p .

- Initialisation évidente.
- On suppose la propriété vraie pour ℓ fixé, $p' \circ q' \in \mathcal{P}_{\text{pair}}^{nc}$ pour tout $p' \in \mathcal{P}_{2\ell}^{nc}$ et q' compatible. Si $p \in \mathcal{P}_{2\ell+2}^{nc}$, alors il possède un intervalle par le lemme. En particulier, il a au moins deux points consécutifs dans un même bloc.

Si ces deux points consécutifs sont reliés dans q , on peut les supprimer dans p et dans q tout en gardant le même résultat de composition, ce qui montre par hypothèse de récurrence que $p \circ q \in \mathcal{P}_{\text{pair}}^{nc}$.

Si les deux points sont dans des blocs différents B_1 et B_2 dans q , on peut supprimer les deux points de B_1 et B_2 et réunir les deux blocs, ce qui donnera un bloc de cardinal pair dans q . Le résultat de la composition reste inchangé, et par hypothèse de récurrence, la composition est dans $\mathcal{P}_{\text{pair}}^{nc}$.

Lauréate du prix junior Maryam Mirzakhani 2022

Alizée Geffroy

Inférence bayésienne pour la phylolinguistique

Par Alizée Geffroy

Encadrée par Robin Ryder

π

Le stage

- › Stage de fin de M1 - Mai-Juin 2022
- › Organisme d'accueil : CEREMADE-Université Paris-Dauphine

Supervisé par **Robin Ryder** (CEREMADE) et **Gilles Blanchard** (Université Paris-Saclay)

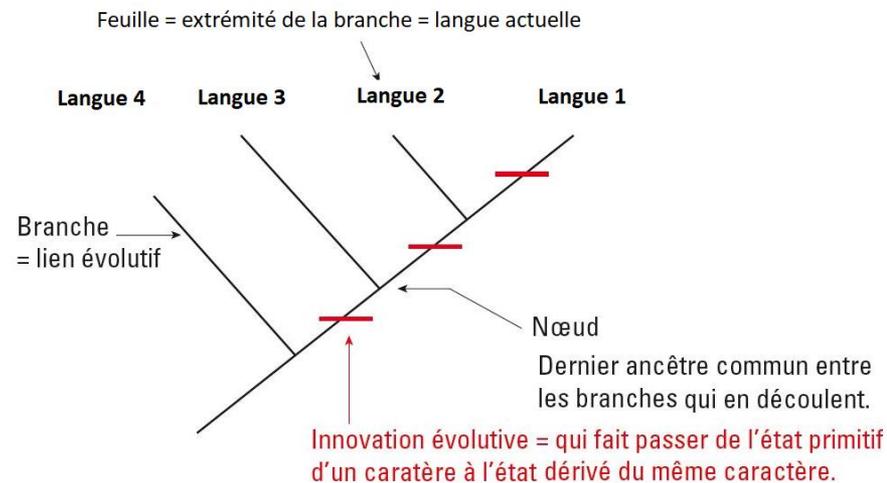


Sommaire

1. Introduction
2. Modèle génératif
3. Convergence méthode MCMC
4. BEAST et fiabilité
5. Conclusion

1. Introduction

› Phylolinguistique : phylogénie + données linguistiques



R.R. Bouckaert, C. Bower, and Q.D. Atkinson. “The origin and expansion of Pama-Nyungan languages across Australia”. In: *Nature ecology evolution* 2 (Apr. 2018), pp. 741–749. DOI: 10.1038/s41559-018-0489-3.

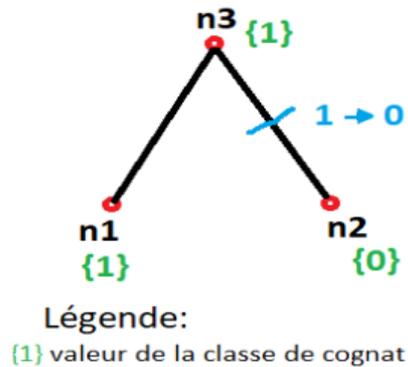
- › Phylogénie de 302 langues Pama-Nyungan australiennes
- › But: inférer l’origine temporelle et spatiale



- › Comment infère-t-on des arbres phylogéniques à partir de données linguistiques ?
- › A quels point ces arbres sont-ils robustes et fiables ?

2. Principe du modèle génératif

- › De la racine aux feuilles
- › Choix du modèle de substitution



$$0 \begin{array}{c} \xrightarrow{\alpha} \\ \xleftarrow{\beta} \end{array} 1$$

$$\begin{array}{c} 0 \quad 1 \\ 1 \begin{bmatrix} - & \alpha \\ \beta & - \end{bmatrix} \end{array}$$

(a) Modèle CTMC

$$\begin{array}{ccc} 0_s & \begin{array}{c} \xrightarrow{\alpha_s} \\ \xleftarrow{\beta_s} \end{array} & 1_s \\ \begin{array}{c} \uparrow S \\ \downarrow S \end{array} & & \begin{array}{c} \uparrow S \\ \downarrow S \end{array} \\ 0_f & \begin{array}{c} \xrightarrow{\alpha_f} \\ \xleftarrow{\beta_f} \end{array} & 1_f \end{array}$$

$$\begin{array}{c} 0_s \quad 1_s \quad 0_f \quad 1_f \\ \begin{array}{c} 0_s \\ 1_s \\ 0_f \\ 1_f \end{array} \begin{bmatrix} - & \alpha_s & S & 0 \\ \beta_s & - & 0 & S \\ S & 0 & - & \alpha_f \\ 0 & S & \beta_f & - \end{bmatrix} \end{array}$$

(b) Modèle du Covarion

3. Méthode MCMC

$$I_r = \int r(\theta) p(\theta) d\theta = \mathbb{E}_p[r(\theta)]$$

Chaîne de Markov

Soit une chaîne de Markov $(X_n), n \in N$, de matrice de transition

- Irréductible
- Récurrente positive
- Apériodique

= ERGODIQUE

Théorème ergodique

Soit X_n une chaîne de Markov ergodique de matrice de transition $(P_{a,b})$.

› Alors pour tous $a, b \in E$ où E est un espace d'état discret

$$\lim_{n \rightarrow \infty} P_{a,b}^{(n)} = \lim_{n \rightarrow \infty} P_a(X_n = b) = \pi_b$$

› Et pour toute fonction intégrable f :

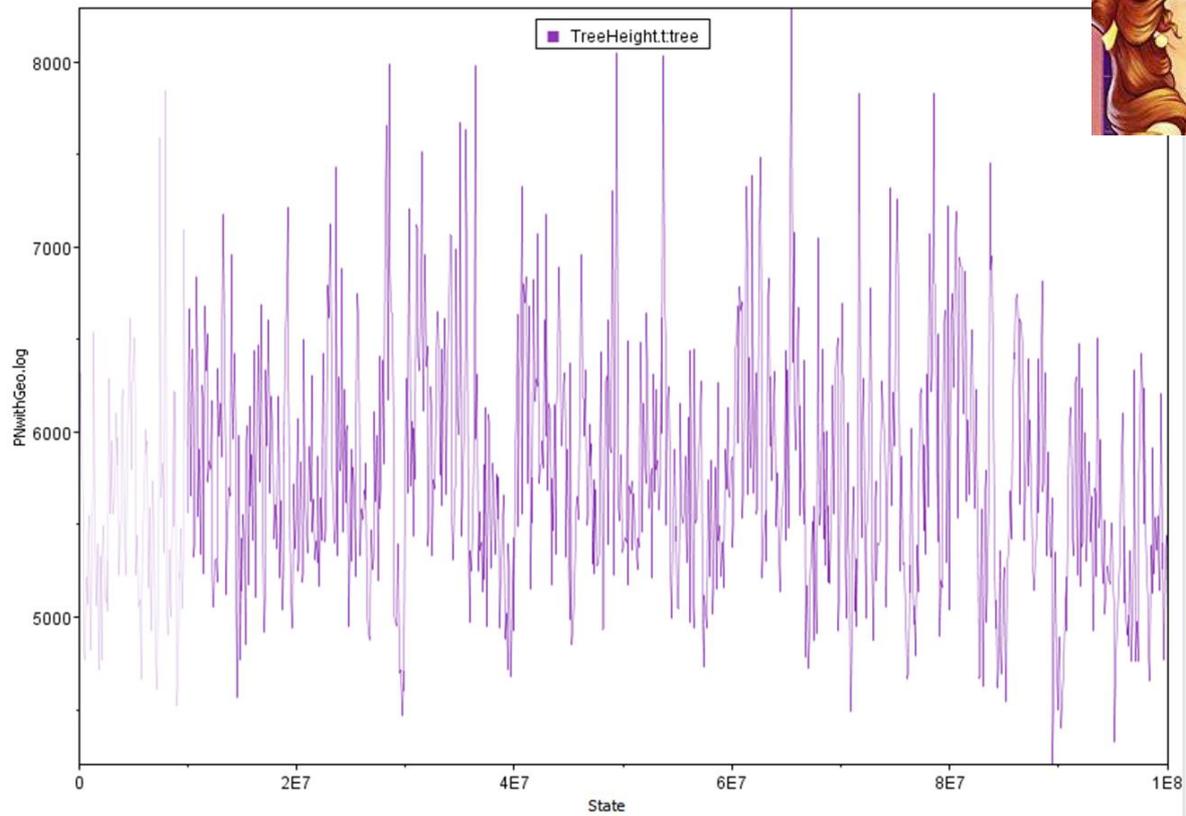
$$\frac{\sum_{n=0}^N f(X_n)}{N} \xrightarrow{\mathbb{P}} \mathbb{E}_\pi[f]$$

π

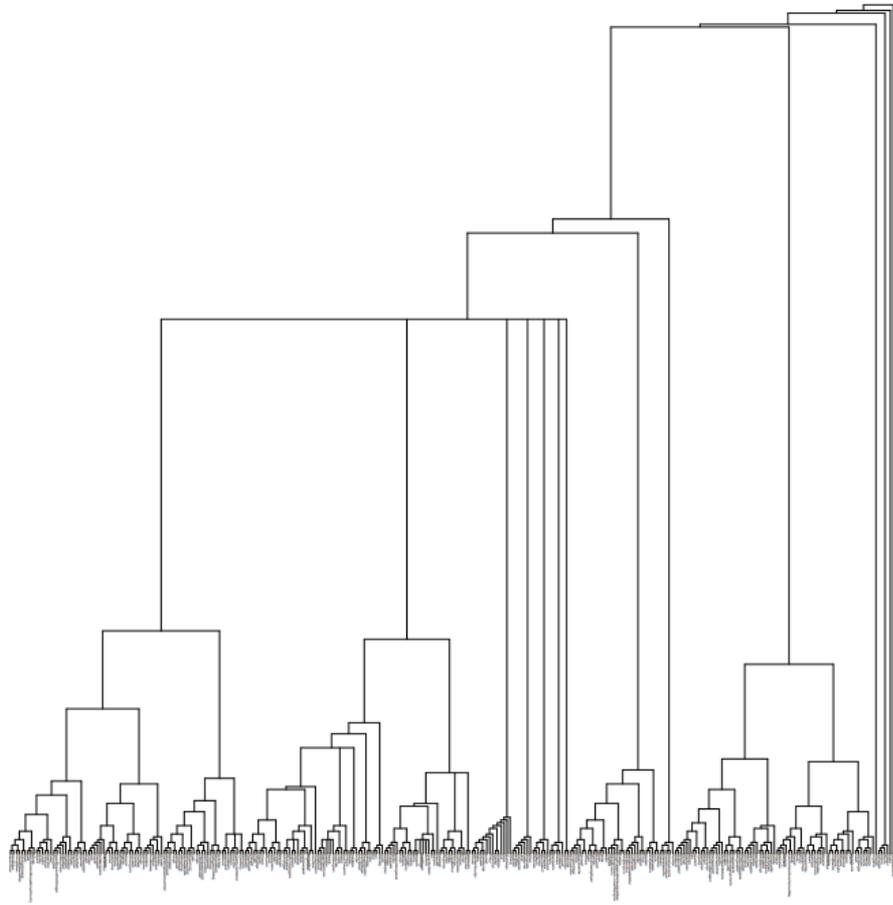
Estimateur de Montecarlo

$$\hat{I} = \frac{\sum_{n=0}^N h(X_n)}{N}$$

4. BEAST pour la fiabilité



5. Conclusion : Les arbres consensus



Remise des prix et des diplômes

Isabelle Jasinowski
(Dir. Op. de la FMJH)