

PHÉNOMÈNE DE CUTOFF POUR LES CHAÎNES DE MARKOV À COURBURE POSITIVE

[d'après J. Salez]

par Anna Ben-Hamou

Introduction

Soient \mathcal{X} un ensemble fini et P un noyau de transition sur \mathcal{X} . On supposera toujours que P est irréductible, c'est-à-dire que pour tous $x, y \in \mathcal{X}$, il existe $k \in \mathbb{N}$ tel que $P^k(x, y) > 0$. La chaîne de Markov en temps continu $(X_t)_{t \in \mathbb{R}_+}$ de noyau P est obtenue en effectuant des transitions selon P à des temps espacés par des variables indépendantes de loi exponentielle de paramètre 1. En notant, pour $t \in \mathbb{R}_+$ et pour $x, y \in \mathcal{X}$, $\mathcal{P}_t(x, y) = \mathbf{P}_x(X_t = y)$, et en remarquant que le nombre de sauts entre 0 et t est de loi de Poisson de paramètre t , on a donc

$$\mathcal{P}_t(x, y) = \sum_{k=0}^{+\infty} \frac{e^{-t} t^k}{k!} P^k(x, y) = e^{-t(I-P)}(x, y).$$

On dit que $(\mathcal{P}_t)_{t \geq 0}$ est le semigroupe associé à P . Considérer la chaîne en temps continu plutôt que discret permet en particulier de parer à tout éventuel problème de périodicité et garantit la convergence de la chaîne vers son unique probabilité stationnaire $\pi = \pi P$, au sens où, pour tous $x, y \in \mathcal{X}$, on a

$$\mathcal{P}_t(x, y) \xrightarrow[t \rightarrow +\infty]{} \pi(y).$$

Une question cruciale est celle de savoir à quelle vitesse a lieu cette convergence. Pour pouvoir quantifier l'écart entre la loi $\mathcal{P}_t(x, \cdot)$ et la loi π , il faut déjà se donner une distance sur l'ensemble des mesures de probabilité sur \mathcal{X} . Il existe de nombreux choix possibles mais le plus naturel pour le probabiliste est celui de la distance en variation totale

$$\|\mu - \pi\|_{\text{TV}} = \max_{A \subset \mathcal{X}} \{\mu(A) - \pi(A)\} = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \pi(x)|.$$

Pour obtenir un contrôle uniforme sur le point de départ, on s'intéresse à la fonction $\mathcal{D} : \mathbb{R}_+ \rightarrow [0, 1]$ définie par

$$\forall t \geq 0, \mathcal{D}(t) = \max_{x \in \mathcal{X}} \|\mathcal{P}_t(x, \cdot) - \pi\|_{\text{TV}}$$

La fonction \mathcal{D} est à valeurs dans $[0, 1]$ et décroît vers 0 quand $t \rightarrow +\infty$. On peut précisément caractériser le taux asymptotique de cette décroissance exponentielle :

$$(1) \quad \mathcal{D}(t)^{1/t} \xrightarrow[t \rightarrow \infty]{} e^{-\tau},$$

où τ correspond à la plus petite partie réelle des valeurs propres non nulles de $I - P$. Ainsi, le spectre de la matrice de transition gouverne en grande partie le profil asymptotique de la fonction \mathcal{D} . L'étude de la convergence des chaînes de Markov finies a longtemps consisté à déterminer les valeurs propres de P , la principale quantité d'intérêt étant le trou spectral, dont l'inverse est appelé temps de relaxation. Cependant, l'approximation $\mathcal{D}(t) \approx e^{-\tau t}$ n'est valable que pour un espace d'états fixé et pour de très grandes valeurs de t , donc lorsque la loi de la chaîne est déjà arbitrairement proche de la loi stationnaire. Ce que l'on aimerait plutôt comprendre, c'est le temps qu'il faut à la chaîne pour que la distance devienne plus petite qu'un certain $\varepsilon \in]0, 1[$ fixé, et comment ce temps dépend de la taille de l'espace d'états, que l'on aimerait pouvoir faire tendre elle aussi vers l'infini. En général en effet, on considère en réalité une suite de noyaux irréductibles $(P_n)_{n \geq 1}$, et donc une suite $(\mathcal{D}_n)_{n \geq 1}$ de fonctions de distance, indexées par un entier $n \geq 1$ qui correspond typiquement à une mesure de la taille de l'espace d'états. Au lieu de s'intéresser au comportement d'une chaîne fixée lorsque $t \rightarrow \infty$ comme dans (1), on va maintenant se donner une cible $\varepsilon \in]0, 1[$ pour la distance et s'intéresser au comportement asymptotique quand $n \rightarrow +\infty$ de la quantité

$$t_{\text{mix}}^{(n)}(\varepsilon) = \inf\{t \geq 0, \mathcal{D}_n(t) \leq \varepsilon\},$$

appelée ε -temps de mélange.

Ce changement de paradigme, initié notamment par les travaux de Persi Diaconis et David Aldous au début des années 1980, a donné lieu à un vif regain d'intérêt pour les chaînes de Markov finies, et a permis de voir que la question des temps de mélange était en fait loin de se réduire à celle du spectre. Cela a aussi permis de découvrir un phénomène fascinant, le phénomène de cutoff, qui correspond à des situations où le premier ordre de $t_{\text{mix}}^{(n)}(\varepsilon)$ ne dépend pas de ε :

$$\forall \varepsilon \in]0, 1[, \quad \frac{t_{\text{mix}}^{(n)}(\varepsilon)}{t_{\text{mix}}^{(n)}(1 - \varepsilon)} \xrightarrow[n \rightarrow +\infty]{} 1.$$

Cela signifie que la distance à l'équilibre reste proche de 1 pendant un temps de l'ordre du temps de mélange, puis chute abruptement vers 0 en un temps bien plus court. Ce phénomène a été observé pour la première fois par Diaconis et Shahshahani (1981) pour le mélange de cartes par transpositions aléatoires. Aldous (1983) observe ce même phénomène pour la marche aléatoire sur l'hypercube $\{0, 1\}^n$.

Depuis ces exemples pionniers, le cutoff a été identifié dans des contextes très variés. Établir le cutoff de façon rigoureuse requiert une analyse très détaillée de la chaîne considérée et constitue bien souvent une tâche difficile, même dans des situations présentant beaucoup de symétrie. On sait désormais que ce phénomène est largement répandu, et

l'on conjecture qu'il est vérifié pour plusieurs grandes familles de chaînes de Markov. Cependant, il existe très peu de résultats généraux concernant le cutoff.

Identifier des conditions générales sous lesquelles le cutoff a lieu, sans forcément chercher à déterminer le temps de mélange, est très vite apparu comme une question essentielle et reste probablement l'un des problèmes les plus fondamentaux dans le domaine des temps de mélange. À cet égard, l'article *Cutoff for non-negatively curved Markov chains* (Salez, 2023a) de Justin Salez constitue une considérable avancée.

La première contribution de cet article est l'énoncé d'une borne supérieure très générale sur la fenêtre du mélange $t_{\text{mix}}^{(n)}(\varepsilon) - t_{\text{mix}}^{(n)}(1 - \varepsilon)$, en fonction d'une quantité appelée *varentropie*. De cette borne découle un critère général pour le cutoff : le critère varentropique. Si ce critère offre une compréhension nouvelle du cutoff comme résultant d'un phénomène de concentration entropique, son utilisation pratique suppose d'estimer la varentropie, qui est une quantité complexe et encore peu étudiée. La deuxième contribution de l'article est de montrer que la varentropie peut être relativement facilement estimée pour une grande famille de chaînes de Markov : les chaînes à *courbure* positive ou nulle (notion définie plus bas). Cela conduit au théorème suivant :

THÉORÈME 0.1. — *Soit $(P_n)_{n \geq 1}$ une suite de noyaux irréductibles à courbure positive ou nulle tels que*

$$\forall x, y \in \mathcal{X}_n, P_n(x, y) > 0 \Leftrightarrow P_n(y, x) > 0.$$

Soit $t_{\text{rel}}^{(n)}$ le temps de relaxation de P_n et soit $\delta^{(n)}$ la plus petite entrée non nulle de P_n . Si pour tout $\varepsilon \in]0, 1[$, on a

$$t_{\text{mix}}^{(n)}(\varepsilon) \gg \left(\log \left(\frac{1}{\delta^{(n)}} \right) t_{\text{rel}}^{(n)} \right)^2,$$

alors la suite $(P_n)_{n \geq 1}$ présente un cutoff.

Le but de ce texte est de faire comprendre à des personnes qui ne sont pas expertes en temps de mélange en quoi ce résultat constitue une avancée majeure. On commence par rappeler les liens entre temps de relaxation et temps de mélange en Section 1. La Section 2 est ensuite consacrée au phénomène de cutoff et à la question de l'identification des mécanismes sous-jacents. En Section 3, on présente le critère varentropique de Salez (2023a). Finalement, en Section 4, on présente les implications de ce critère pour les chaînes à courbure positive.

1. Temps de relaxation et temps de mélange

Revenons temporairement au cas d'un seul noyau irréductible P et notons \mathcal{D} la fonction distance associée. Par la propriété de sous-multiplicativité $\mathcal{D}(t+s) \leq 2\mathcal{D}(t)\mathcal{D}(s)$

et le lemme de Fekete, le terme de droite dans (1) correspond en fait à l'infimum de l'ensemble $\{(2\mathcal{D}(t))^{1/t}, t > 0\}$. Ainsi, pour tout $t > 0$, on a

$$\mathcal{D}(t) \geq \frac{1}{2}e^{-\tau t},$$

et pour tout $\varepsilon \in]0, 1[$,

$$(2) \quad t_{\text{mix}}(\varepsilon) \geq \frac{1}{\tau} \log \left(\frac{1}{2\varepsilon} \right),$$

où $t_{\text{mix}}(\varepsilon) = \inf\{t \geq 0, \mathcal{D}(t) \leq \varepsilon\}$. Dans l'autre direction, on aimerait pouvoir dire qu'en prenant t de l'ordre de $1/\tau$, on rend la distance $\mathcal{D}(t)$ petite. Comme nous le verrons, cela est très souvent faux. On peut néanmoins obtenir une borne supérieure spectrale sur $t_{\text{mix}}(\varepsilon)$ de la façon suivante. Soit γ la constante de Poincaré de P , i.e. la plus grande constante telle que pour toute fonction $f: \mathcal{X} \rightarrow \mathbb{R}$, on a

$$\gamma \mathbf{Var}(f) \leq \mathcal{E}(f, f),$$

où $\mathbf{Var}(f)$ correspond à la variance de f sous la loi π :

$$\mathbf{Var}(f) = \frac{1}{2} \sum_{x, y \in \mathcal{X}} \pi(x)\pi(y) (f(y) - f(x))^2,$$

et où $\mathcal{E}(f, g)$ correspond à la forme de Dirichlet de P :

$$\mathcal{E}(f, g) = \langle (I - P)f, g \rangle_{\pi},$$

avec $\langle \cdot, \cdot \rangle_{\pi}$ le produit scalaire dans $\ell^2(\pi)$. En particulier, on a

$$\mathcal{E}(f, f) = \frac{1}{2} \sum_{x, y \in \mathcal{X}} \pi(x)P(x, y) (f(y) - f(x))^2,$$

ce qui peut être interprété comme une variance locale de f sous l'action de P . Notons que dans le cas réversible où P est égal à son adjoint P^* dans $\ell^2(\pi)$, on a

$$\gamma = \tau.$$

Dans le cas général, la quantité $1 - \gamma$ peut être définie, de façon équivalente, comme la deuxième plus grande valeur propre du noyau réversible $\frac{P+P^*}{2}$. C'est pourquoi γ est souvent appelé le *trou spectral* : c'est l'écart entre les deux plus grandes valeurs propres de $\frac{P+P^*}{2}$. Le temps de relaxation est alors défini comme l'inverse du trou spectral :

$$t_{\text{rel}} = \frac{1}{\gamma}.$$

Soit maintenant μ_0 une distribution initiale sur \mathcal{X} et notons $\mu_t = \mu_0 \mathcal{P}_t$ la distribution au temps $t \geq 0$, et $f_t = \frac{\mu_t}{\pi}$ la densité de μ_t par rapport à la distribution stationnaire π . Par l'inégalité de Cauchy–Schwarz, on a

$$(3) \quad 4\|\mu_t - \pi\|_{\text{TV}}^2 \leq \sum_{x \in \mathcal{X}} \pi(x) \left(\frac{\mu_t(x)}{\pi(x)} - 1 \right)^2 = \mathbf{Var}(f_t).$$

Or un simple calcul montre que

$$\partial_t \mathbf{Var}(f_t) = -2\mathcal{E}(f_t, f_t).$$

Par définition du trou spectral, on a alors $\partial_t \mathbf{Var}(f_t) \leq -2\gamma \mathbf{Var}(f_t)$, ce qui donne

$$\mathbf{Var}(f_t) \leq \mathbf{Var}(f_0) e^{-2\gamma t}.$$

En revenant à (3) et en utilisant le fait que

$$\mathbf{Var}(f_0) \leq \sum_{x \in \mathcal{X}} \pi(x) f_0(x)^2 = \sum_{x \in \mathcal{X}} \mu_0(x) \frac{\mu_0(x)}{\pi(x)} \leq \left\| \frac{\mu_0}{\pi} \right\|_\infty,$$

avec $\left\| \frac{\mu_0}{\pi} \right\|_\infty = \max_{x \in \mathcal{X}} \frac{\mu_0(x)}{\pi(x)}$, on obtient

$$(4) \quad \|\mu_t - \pi\|_{\text{TV}} \leq \frac{1}{2} e^{-\gamma t} \sqrt{\left\| \frac{\mu_0}{\pi} \right\|_\infty}.$$

En prenant le maximum sur μ_0 (obtenu pour μ_0 concentrée sur un point de masse stationnaire minimale), on a

$$(5) \quad \mathcal{D}(t) \leq \frac{1}{2\sqrt{\pi_\star}} e^{-\gamma t},$$

avec $\pi_\star = \min_{x \in \mathcal{X}} \pi(x)$, soit encore

$$(6) \quad t_{\text{mix}}(\varepsilon) \leq t_{\text{rel}} \log \left(\frac{1}{2\varepsilon\sqrt{\pi_\star}} \right).$$

Comme $\pi_\star \leq \frac{1}{|\mathcal{X}|}$, le terme de droite dans cette inégalité est d'ordre plus grand que t_{rel} . Cela dit, savoir estimer le temps de mélange à un facteur logarithmique près en utilisant seulement le trou spectral n'en est pas moins impressionnant, et il existe de nombreuses situations pour lesquelles ce facteur est nécessaire.

Ci-dessous, on détaille deux exemples, illustrant des comportements de mélange très différents : la marche aléatoire sur le cercle et la marche aléatoire sur l'hypercube.

Exemple 1.1 (Marche aléatoire sur le cercle). — Soient $\mathcal{X}_n = \mathbb{Z}/n\mathbb{Z}$, l'ensemble des entiers modulo n , et

$$P_n(x, y) = \begin{cases} \frac{1}{2} & \text{si } y \equiv x \pm 1 \pmod{n}, \\ 0 & \text{sinon.} \end{cases}$$

Cela correspond à faire une marche aléatoire symétrique sur un cercle à n nœuds numérotés de 0 à $n - 1$. Il s'agit d'un noyau réversible (pour la mesure uniforme) dont les valeurs propres sont égales à $\cos(2\pi j/n)$, $j = 0, \dots, n - 1$. En particulier,

$$t_{\text{rel}}^{(n)} = \frac{1}{1 - \cos(2\pi/n)} \underset{n \rightarrow +\infty}{\sim} \frac{n^2}{2\pi^2}.$$

De plus, on peut facilement se convaincre que le temps de mélange est lui aussi d'ordre n^2 : par le théorème central limite, le support de la marche au temps t correspond à un segment d'ordre \sqrt{t} . Pour atteindre la distribution uniforme sur tout le cercle, il faut donc

prendre $t \approx n^2$. En utilisant la décomposition spectrale de P_n (ou bien un raffinement local du théorème central limite), on peut être bien plus précis : pour tout $\alpha > 0$,

$$\mathcal{D}_n(\alpha n^2) \xrightarrow{n \rightarrow +\infty} \varphi(\alpha) = \int_0^1 \left| \sum_{k=1}^{+\infty} e^{-2\alpha\pi^2 k^2} \cos(2\pi k u) \right| du.$$

Ainsi, à l'échelle n^2 , le profil limite de \mathcal{D}_n est donné par la fonction φ , qui décroît continûment de 1 vers 0. Autrement dit, pour tout $\varepsilon \in]0, 1[$,

$$t_{\text{mix}}^{(n)}(\varepsilon) \underset{n \rightarrow +\infty}{\sim} \varphi^{-1}(\varepsilon) n^2.$$

Exemple 1.2 (Marche aléatoire sur l'hypercube). — Soient $\mathcal{X}_n = \{0, 1\}^n$, l'hypercube discret en dimension n , et

$$P_n(x, y) = \begin{cases} \frac{1}{n} & \text{si } \sum_{i=1}^n |y_i - x_i| = 1, \\ 0 & \text{sinon.} \end{cases}$$

Il s'agit là aussi d'un noyau réversible, et les valeurs propres sont données par $1 - \frac{2k}{n}$, $k = 0, \dots, n$ (avec multiplicité $\binom{n}{k}$ respectivement). On a donc

$$t_{\text{rel}}^{(n)} \underset{n \rightarrow +\infty}{\sim} \frac{n}{2}.$$

Pour déterminer le temps de mélange, remarquons que, comme pour le cas du cercle, la distance ne dépend pas du point de départ. On peut donc supposer que l'on part du point $(0, \dots, 0)$. Dans ce cas, la loi de la chaîne au temps t est celle d'un vecteur i.i.d. de variables de Bernoulli, dont le paramètre correspond à la probabilité qu'une variable de loi de Poisson de paramètre t/n soit impaire, soit $\frac{1}{2}(1 - e^{-2t/n})$. On obtient alors

$$\mathcal{D}_n(t) = \frac{1}{2^{n+1}} \sum_{k=0}^n \binom{n}{k} \left| 1 - (1 - e^{-2t/n})^k (1 + e^{-2t/n})^{n-k} \right|,$$

et l'on observe que pour tout $\alpha > 0$,

$$\mathcal{D}_n(\alpha n \log n) \xrightarrow{n \rightarrow +\infty} \begin{cases} 1 & \text{si } \alpha < \frac{1}{4}, \\ 0 & \text{si } \alpha > \frac{1}{4}. \end{cases}$$

Autrement dit, pour tout $\varepsilon \in]0, 1[$,

$$t_{\text{mix}}^{(n)}(\varepsilon) \underset{n \rightarrow +\infty}{\sim} \frac{1}{4} n \log n.$$

Contrairement au cas du cercle, le temps de mélange est ici bien plus grand que le temps de relaxation. De plus, la convergence a lieu de façon extrêmement abrupte : à l'échelle $n \log n$, la distance converge vers une fonction indicatrice et le premier ordre du temps de mélange ne dépend pas de la cible ε . C'est le phénomène de cutoff.

2. Le phénomène de cutoff

Depuis sa découverte au début des années 1980 à travers les exemples du battage de cartes par transpositions aléatoires (Diaconis et Shahshahani, 1981) et de la marche aléatoire sur l’hypercube (Aldous, 1983), le phénomène de cutoff a été observé pour de nombreuses chaînes dans des contextes très variés : pour de nombreux mélanges de cartes comme le fameux *riffle shuffle* (Bayer et Diaconis, 1992) et de manière générale pour de nombreuses marches aléatoires sur des groupes finis (Diaconis, 1996 ; Chen et Saloff-Coste, 2008 ; Saloff-Coste, 2004), pour divers systèmes de particules en interaction comme la dynamique de Glauber sur le modèle d’Ising à haute température (Levin, Luczak et Peres, 2010 ; Lubetzky et Sly, 2013), ou encore les processus d’exclusion symétrique (Lacoin, 2016) et asymétrique (Labbé et Lacoin, 2019), cette liste étant bien loin d’être exhaustive. À chaque fois, cela demande une analyse très précise de la chaîne considérée : le temps de mélange doit être très précisément estimé pour pouvoir voir que le pré-facteur ne dépend pas de ε .

Une question qui s’est très vite posée est celle de savoir si une chaîne présente ou non un cutoff, sans forcément avoir à déterminer le pré-facteur exact devant le temps de mélange. C’est une question essentielle d’un point de vue à la fois théorique et pratique :

- d’un point de vue théorique, il s’agit de comprendre *pourquoi* le cutoff a lieu et quels sont les mécanismes sous-jacents à l’apparition de ce phénomène ;
- d’un point de vue pratique, les chaînes de Markov sont très utilisées pour la simulation de variables aléatoires (c’est l’approche dite MCMC : *Monte Carlo Markov Chains*). Si une chaîne présente un cutoff, il est alors très important de le savoir car cela signifie qu’il existe un temps critique avant lequel il est très dangereux de s’arrêter (la distance serait encore très proche de 1), mais après lequel il est inutile d’aller (la distance est déjà très proche de 0).

2.1. La condition produit

Dans le cas réversible, on peut facilement voir qu’une condition nécessaire au cutoff est

$$(7) \quad t_{\text{rel}}^{(n)} \ll t_{\text{mix}}^{(n)},$$

avec $t_{\text{mix}}^{(n)} = t_{\text{mix}}^{(n)}(1/4)$. En effet, si une suite de chaînes présente le cutoff, alors, par l’inégalité (2), on a, pour tout $\varepsilon \in]0, 1[$,

$$\frac{t_{\text{mix}}^{(n)}}{t_{\text{rel}}^{(n)}} \underset{n \rightarrow \infty}{\sim} \frac{t_{\text{mix}}^{(n)}(\varepsilon)}{t_{\text{rel}}^{(n)}} \geq \log \left(\frac{1}{2\varepsilon} \right).$$

En faisant tendre ε vers 0, on voit que $t_{\text{mix}}^{(n)} \gg t_{\text{rel}}^{(n)}$.

En 2004, lors d’un workshop consacré aux temps de mélange, Peres (2004) a conjecturé que la condition (7), appelée *condition produit*, était aussi suffisante pour de grandes familles de chaînes. Cela a été vérifié pour les chaînes de vie et de mort (Ding, Lubetzky et Peres, 2010), et plus généralement les marches aléatoires sur les arbres (Basu, Hermon

et Peres, 2017). Mais l'espoir que cela soit vrai pour toutes les chaînes réversibles a vite été perdu. On peut même transformer n'importe quelle chaîne réversible avec cutoff en une chaîne sans cutoff qui est toujours réversible et qui vérifie toujours la condition (7), comme le montre l'exemple suivant.

Exemple 2.1. — Soit $(P_n)_{n \geq 1}$ une suite de noyaux réversibles présentant un cutoff. En particulier, cette suite vérifie $t_{\text{rel}}^{(n)} \ll t_{\text{mix}}^{(n)}$. Soit maintenant

$$\tilde{P}_n = (1 - \theta_n)P_n + \theta_n \Pi_n,$$

où Π_n est la matrice dont toutes les lignes sont égales à la distribution stationnaire π_n , et où la suite $(\theta_n)_{n \geq 1}$ vérifie

$$t_{\text{rel}}^{(n)} \ll \frac{1}{\theta_n} \ll t_{\text{mix}}^{(n)}.$$

On vérifie aisément que la distance à l'équilibre de cette nouvelle chaîne s'écrit

$$(8) \quad \tilde{\mathcal{D}}_n(t) = e^{-\theta_n t} \mathcal{D}_n(t).$$

En particulier, le temps de mélange est de l'ordre de $\frac{1}{\theta_n} \ll t_{\text{mix}}^{(n)}$ et il n'y a pas cutoff : pour tout $\alpha > 0$,

$$\tilde{\mathcal{D}}_n\left(\frac{\alpha}{\theta_n}\right) = e^{-\alpha} \mathcal{D}_n\left(\frac{\alpha}{\theta_n}\right) \xrightarrow{n \rightarrow +\infty} e^{-\alpha}.$$

La chaîne vérifie cependant encore la condition (7), puisqu'en élevant à la puissance $1/t$ dans (8), en faisant tendre $t \rightarrow +\infty$ et en utilisant (1), on a

$$\frac{1}{\widetilde{t_{\text{rel}}^{(n)}}} = \theta_n + \frac{1}{t_{\text{rel}}^{(n)}},$$

ce qui implique $\widetilde{t_{\text{rel}}^{(n)}} \sim t_{\text{rel}}^{(n)} \ll \frac{1}{\theta_n}$.

Dans l'article Basu, Hermon et Peres (2017), qui établit que la condition (7) est suffisante sur les arbres, les auteurs démontrent une caractérisation du cutoff en termes de concentration des temps d'atteinte des « pires » sous-ensembles. Si cette caractérisation permet de mieux comprendre le cutoff comme un phénomène de concentration, elle est difficilement utilisable en pratique. La vérifier demande de calculer précisément le pré-facteur devant les temps d'atteinte, un problème généralement aussi dur que celui de calculer le pré-facteur devant le temps de mélange. Ce que l'on souhaiterait, c'est un critère tel que (7), qui ne demande que de comparer des ordres de grandeur.

2.2. Graphes expandeurs

Une importante classe de graphes dont les remarquables propriétés plaident en faveur du cutoff est celle des expandeurs à degrés bornés. Une famille de graphes $(G_n)_{n \geq 1}$ avec $|V_n| \rightarrow \infty$ (où V_n est l'ensemble des sommets de G_n) est dite famille d'expandeurs, s'il existe $\alpha > 0$ tel que pour tout $n \geq 1$

$$\gamma^{(n)} \geq \alpha,$$

où $\gamma^{(n)}$ est le trou spectral de la marche aléatoire simple sur G_n . Par la borne (6), on voit que si les degrés sont uniformément bornés, alors le temps de mélange de la marche sur G_n satisfait $t_{\text{mix}} = O(\log |V_n|)$. En ordre de grandeur, il s'agit du plus petit temps de mélange possible sur les graphes à degrés bornés. En effet, soit G_n un graphe dont tous les degrés sont compris entre 1 et Δ et, pour $x \in V_n$, notons $\mathcal{B}_x(2t)$ l'ensemble des sommets de V_n qui sont à distance au plus $2t$ de x . Alors, on a d'une part

$$|\mathcal{B}_x(2t)| \leq \Delta^{2t},$$

et d'autre part, la probabilité que la chaîne ne soit pas dans $\mathcal{B}_x(2t)$ au temps t est majorée par la probabilité que le nombre de sauts N_t entre 0 et t soit plus grand que $2t$. Par la méthode de Chernoff pour la loi de Poisson, on obtient

$$\mathcal{P}_t(x, \mathcal{B}_x(2t)^c) \leq \mathbf{P}(N_t > 2t) \leq e^{-(2\log(2)-1)t}.$$

Ainsi

$$\|\mathcal{P}_t(x, \cdot) - \pi\|_{\text{TV}} \geq 1 - e^{-(2\log(2)-1)t} - \frac{\Delta^{2t}}{|V_n|}.$$

Pour $t = \frac{\log(|V_n|(1-\varepsilon))}{2\log(\Delta)}$ et n assez grand, le terme de droite est supérieur à $1 - \varepsilon/2$. Sur un graphe à degrés bornés, le temps de mélange est donc d'ordre au moins logarithmique en $|V_n|$, et cette vitesse est atteinte sur les graphes expandeurs.

De plus, comme ils vérifient la condition produit (7), on a pensé pendant longtemps que de tels graphes devaient présenter le cutoff. Cependant, jusqu'à 2010, aucun exemple d'expandeurs avec cutoff n'était connu. Puis Lubetzky et Sly (2011) ont construit des expandeurs 3-réguliers à la fois *avec* et *sans* cutoff. Une conjecture, toujours ouverte, de Peres (2004) est qu'il y a cutoff sur tous les expandeurs *transitifs* de degré borné. Jusqu'ici, le résultat le plus général concernant les graphes expandeurs est Lubetzky et Peres (2016), établissant le cutoff pour les expandeurs optimaux que sont les graphes de Ramanujan. Ces graphes atteignent le plus grand trou spectral possible pour les graphes d -réguliers (donné par $1 - \frac{2\sqrt{d-1}}{d}$, d'après le Théorème d'Alon–Boppana (Nilli, 1991)).

3. Le critère varentropique

Le premier résultat de Salez (2023a) établit un critère très général pour le cutoff et permet de comprendre ce dernier comme résultant d'un phénomène de concentration entropique.

L'entropie relative (ou divergence de Kullback–Leibler) entre deux mesures de probabilité μ et π sur \mathcal{X} est définie par

$$\mathbf{D}(\mu\|\pi) = \sum_{x \in \mathcal{X}} \mu(x) \log \frac{\mu(x)}{\pi(x)}.$$

L'utilisation de l'entropie relative pour contrôler le temps de mélange des chaînes de Markov n'est pas nouvelle et se fait généralement via l'inégalité de Pinsker :

$$\|\mu - \pi\|_{\text{TV}}^2 \leq 2\mathbf{D}(\mu\|\pi).$$

Cette inégalité constitue le point de départ d'une approche permettant d'obtenir des bornes supérieures puissantes sur le temps de mélange, à l'aide d'inégalités fonctionnelles appelées inégalités de log-Sobolev modifiées. Pour voir cela, notons comme précédemment $\mu_t = \mu_0 \mathcal{P}_t$ et $f_t = \frac{\mu_t}{\pi}$ la densité de μ_t par rapport à π . Alors que la variance de f_t est reliée à la distance ℓ^2 entre μ_t et π , son entropie⁽¹⁾ est reliée à l'entropie relative entre les deux distributions :

$$\mathbf{Ent}(f_t) = \mathbf{E}[f_t \log f_t] = \mathbf{D}(\mu_t\|\pi).$$

Et de même que l'égalité $\partial_t \mathbf{Var}(f_t) = -2\mathcal{E}(f_t, f_t)$ avait motivé la définition du trou spectral, l'égalité

$$\partial_t \mathbf{D}(\mu_t\|\pi) = -\mathcal{E}(f_t, \log f_t)$$

motive la définition de la constante de log-Sobolev modifiée comme la plus grande constante ρ telle que l'inégalité

$$\rho \mathbf{Ent}(f) \leq \frac{1}{2} \mathcal{E}(f, \log f)$$

est valable pour toute fonction $f: \mathcal{X} \rightarrow \mathbb{R}_+$. On obtient alors $\partial_t \mathbf{D}(\mu_t\|\pi) \leq -2\rho \mathbf{D}(\mu_t\|\pi)$, ce qui donne

$$\mathbf{D}(\mu_t\|\pi) \leq \mathbf{D}(\mu_0\|\pi) e^{-2\rho t}.$$

En prenant μ_0 concentrée sur un point de masse stationnaire minimale et en utilisant l'inégalité de Pinsker, on obtient

$$(9) \quad \mathcal{D}(t) \leq \sqrt{2 \log \left(\frac{1}{\pi_\star} \right)} e^{-\rho t}.$$

Cette approche, développée par Bobkov et Tetali (2006), donne souvent des bornes supérieures bien plus fines que celles découlant du temps de relaxation. En effet, même si l'on a toujours $\rho \leq \gamma$, il existe de nombreuses situations où ces deux quantités sont du même ordre. On gagne alors à utiliser (9) plutôt que (5), du fait du log supplémentaire devant π_\star^{-1} . Cependant, ce genre de raffinement n'est pas d'une grande aide pour établir le cutoff, qui nécessite d'avoir des bornes dans les deux directions. Un des apports de Salez (2023a) est de fournir une forme de réciproque à l'inégalité de Pinsker, en établissant un lien bidirectionnel entre l'entropie et le mélange, via la notion de varentropie. Pour la borne inférieure, on souhaiterait pouvoir dire que la distance en variation totale est grande quand l'entropie relative est grande. Cela a peu de chance d'être vrai sans un contrôle supplémentaire. En effet, s'il existe de petites régions de l'espace où le rapport $\frac{\mu}{\pi}$ est grand, cela peut fortement augmenter l'entropie relative,

1. On rappelle que pour $f: \mathcal{X} \rightarrow \mathbb{R}_+$, on a $\mathbf{Ent}(f) = \mathbf{E}[f \log f] - \mathbf{E}[f] \log \mathbf{E}[f]$, où l'espérance est prise sous la loi π .

tout en n'ayant qu'un faible effet en variation totale. Pour éviter ce genre de pathologies, on peut chercher à contrôler les fluctuations typiques de $\frac{\mu}{\pi}$ en considérant la varentropie :

$$\mathbf{V}(\mu\|\pi) = \sum_{x \in \mathcal{X}} \mu(x) \left(\log \frac{\mu(x)}{\pi(x)} - \mathbf{D}(\mu\|\pi) \right)^2 .$$

Le lemme suivant peut être vu comme une réciproque de l'inégalité de Pinsker.

LEMME 3.1 (Borne inférieure entropique). — *Pour toute distribution μ sur \mathcal{X} et pour tout $\varepsilon \in]0, 1[$, on a*

$$\|\mu - \pi\|_{\text{TV}} \leq 1 - \varepsilon \quad \implies \quad \mathbf{D}(\mu\|\pi) \leq \frac{1 + \sqrt{\mathbf{V}(\mu\|\pi)}}{\varepsilon} .$$

Preuve du Lemme 3.1. — Soit $A \subset \mathcal{X}$ le sous-ensemble défini par

$$A = \left\{ x \in \mathcal{X} , \mu(x) \geq \pi(x)e^\theta \right\} ,$$

avec

$$\theta = \mathbf{D}(\mu\|\pi) - \frac{\sqrt{\mathbf{V}(\mu\|\pi)}}{\varepsilon} .$$

L'inégalité de Chebyshev implique $\mu(A) \geq 1 - \varepsilon^2$. D'autre part, par définition de A , on a $\pi(A) \leq e^{-\theta} \mu(A)$. En combinant ces deux inégalités, on obtient

$$\|\mu - \pi\|_{\text{TV}} \geq \mu(A) - \pi(A) \geq (1 - \varepsilon^2)(1 - e^{-\theta}) .$$

Ainsi, si $\|\mu - \pi\|_{\text{TV}} \leq 1 - \varepsilon$, alors

$$(1 + \varepsilon)(1 - e^{-\theta}) \leq 1 ,$$

soit $\theta \leq \log\left(\frac{1}{\varepsilon} + 1\right) \leq \frac{1}{\varepsilon}$, comme souhaité. □

Revenons maintenant aux chaînes de Markov. Notons

$$\mathbf{D}^*(t) = \max_{x \in \mathcal{X}} \mathbf{D}(\mathcal{P}_t(x, \cdot) \| \pi) ,$$

et

$$\mathbf{V}^*(t) = \max_{x \in \mathcal{X}} \mathbf{V}(\mathcal{P}_t(x, \cdot), \pi) .$$

Le théorème suivant montre que la quantité $\mathbf{V}^*(t)$ est déterminante dans l'apparition du cutoff, et fournit un critère simple pour l'occurrence de ce phénomène : le critère varentropique.

THÉORÈME 3.2 (Le critère varentropique). — *Pour toute matrice de transition P et pour tout $\varepsilon \in]0, 1/2[$,*

$$t_{\text{mix}}(\varepsilon) - t_{\text{mix}}(1 - \varepsilon) \leq \frac{2t_{\text{rel}}}{\varepsilon^2} \left(1 + \sqrt{\mathbf{V}^*(t_{\text{mix}}(1 - \varepsilon))} \right) .$$

En particulier, pour qu'une suite $(P_n)_{n \geq 1}$ de matrices de transition présente le cutoff, il suffit que, pour tout $\varepsilon \in]1/2, 1[$,

$$1 + \sqrt{\mathbf{V}_n^*(t_{\text{mix}}^{(n)}(\varepsilon))} \ll \frac{t_{\text{mix}}^{(n)}(\varepsilon)}{t_{\text{rel}}^{(n)}} .$$

Dans ce cas, pour tout $\varepsilon \in]0, 1[$, on a $t_{\text{mix}}^{(n)}(\varepsilon) \sim t_n$ où t_n est solution de

$$\mathbf{D}_n^*(t_n) \asymp 1 + \sqrt{\mathbf{V}_n^*(t_n)},$$

où \asymp correspond à l'égalité à constante multiplicative près.

Preuve du Théorème 3.2. — Commençons par montrer que pour toute loi μ , et pour

$$(10) \quad s = \frac{t_{\text{rel}}}{\varepsilon} (1 + \mathbf{D}(\mu\|\pi)),$$

on a

$$(11) \quad \|\mu_{\mathcal{P}_s} - \pi\|_{\text{TV}} \leq \varepsilon.$$

Soit $A \subset \mathcal{X}$ le sous-ensemble défini par

$$A = \left\{ x \in \mathcal{X}, \log \frac{\mu(x)}{\pi(x)} < 1 + \frac{2\mathbf{D}(\mu\|\pi)}{\varepsilon} \right\},$$

et soit $\mu_A = \mu(\cdot|A)$, la distribution μ conditionnée à être dans A . Par l'inégalité triangulaire, on a

$$\|\mu_{\mathcal{P}_s} - \pi\|_{\text{TV}} \leq \|\mu_{\mathcal{P}_s} - \mu_A_{\mathcal{P}_s}\|_{\text{TV}} + \|\mu_A_{\mathcal{P}_s} - \pi\|_{\text{TV}}.$$

On va montrer que chacun des deux termes du membre de droite ci-dessus est inférieur à $\varepsilon/2$. D'une part, on a

$$\|\mu_{\mathcal{P}_s} - \mu_A_{\mathcal{P}_s}\|_{\text{TV}} \leq \|\mu - \mu_A\|_{\text{TV}} = \mu(A^c).$$

Or, par définition de A^c , on a

$$\begin{aligned} \left(1 + \frac{2\mathbf{D}(\mu\|\pi)}{\varepsilon}\right) \mu(A^c) &\leq \sum_{x \in A^c} \mu(x) \log \frac{\mu(x)}{\pi(x)} \\ &= \mathbf{D}(\mu\|\pi) + \sum_{x \in A} \mu(x) \log \frac{\pi(x)}{\mu(x)}. \end{aligned}$$

En utilisant l'inégalité $\log u \leq u - 1$, on a

$$\sum_{x \in A} \mu(x) \log \frac{\pi(x)}{\mu(x)} \leq \pi(A) - \mu(A) \leq 1 - \mu(A) = \mu(A^c).$$

Il suffit alors de ré-arranger les termes pour obtenir $\mu(A^c) \leq \varepsilon/2$. D'autre part, on sait par (4) que

$$\|\mu_A_{\mathcal{P}_s} - \pi\|_{\text{TV}} \leq \frac{e^{-s/t_{\text{rel}}}}{2} \sqrt{\left\| \frac{\mu_A}{\pi} \right\|_{\infty}}.$$

Or, par définition de A , on a

$$\left\| \frac{\mu_A}{\pi} \right\|_{\infty} = \frac{1}{\mu(A)} \max_{x \in A} \frac{\mu(x)}{\pi(x)} \leq \frac{1}{\mu(A)} \exp \left(1 + \frac{2\mathbf{D}(\mu\|\pi)}{\varepsilon} \right),$$

Comme $\mu(A) \geq 1/2 \geq 1/e$, on obtient

$$\|\mu_A_{\mathcal{P}_s} - \pi\|_{\text{TV}} \leq \frac{1}{2} \exp \left(-\frac{s}{t_{\text{rel}}} + 1 + \frac{\mathbf{D}(\mu\|\pi)}{\varepsilon} \right).$$

En remplaçant s par sa valeur donnée en (10), on a alors

$$\|\mu_A \mathcal{P}_s - \pi\|_{\text{TV}} \leq \frac{1}{2} e^{-(\frac{1}{\varepsilon}-1)} \leq \frac{\varepsilon}{2},$$

et l'on a bien établi (11).

En appliquant cette inégalité à $\mu = \mathcal{P}_t(x, \cdot)$ et en maximisant sur $x \in \mathcal{X}$, on obtient

$$t_{\text{mix}}(\varepsilon) \leq t + \frac{t_{\text{rel}}}{\varepsilon} (1 + \mathbf{D}^*(t)).$$

De plus, par le Lemme 3.1, on a

$$\mathbf{D}^*(t_{\text{mix}}(1 - \varepsilon)) \leq \frac{1 + \sqrt{\mathbf{V}^*(t_{\text{mix}}(1 - \varepsilon))}}{\varepsilon}.$$

Ainsi, en utilisant le fait que $\varepsilon \leq 1$, on a bien

$$t_{\text{mix}}(\varepsilon) - t_{\text{mix}}(1 - \varepsilon) \leq \frac{2t_{\text{rel}}}{\varepsilon^2} \left(1 + \sqrt{\mathbf{V}^*(t_{\text{mix}}(1 - \varepsilon))}\right).$$

□

Récemment Salez (2023b) montre que ce critère est équivalent au cutoff pour les suites de matrices creuses expansives à support symétrique. Ces matrices peuvent être vues comme une généralisation du cas des graphes expandeurs à degrés bornés introduits en Section 2.2 et vérifient les trois hypothèses suivantes :

1. Symétrie du support :

$$\forall x, y \in \mathcal{X}_n, P_n(x, y) > 0 \Leftrightarrow P_n(y, x) > 0.$$

2. Matrices creuses :

$$\inf_{n \geq 1} \delta(P_n) > 0,$$

où $\delta(P_n)$ correspond à la plus petite entrée non nulle de P_n .

3. Matrices expansives :

$$\inf_{n \geq 1} \gamma(P_n) > 0,$$

où $\gamma(P_n)$ correspond au trou spectral de P_n .

Sous ces hypothèses, Salez (2023b) établit que la suite $(P_n)_{n \geq 1}$ présente le cutoff si et seulement si, pour tout $\varepsilon \in]0, 1[$, on a

$$\mathbf{V}_n^*(t_{\text{mix}}^{(n)}(\varepsilon)) \ll (\log |\mathcal{X}_n|)^2.$$

4. Chaînes à courbure positive

Venons-en maintenant à la preuve du Théorème 0.1, qui résulte d’une estimation de la varentropie pour les chaînes à courbure positive.

La notion de courbure de Ricci est un concept central en géométrie riemannienne. Il a fallu du temps pour qu’émergent des analogues satisfaisants de cette notion dans le cas discret des graphes et des chaînes de Markov. Une première approche, reposant sur le Γ -calcul, a été proposée par Bakry et Émery (2006). Plus récemment, Ollivier (2009) a proposé une autre définition de courbure discrète, issue de la théorie du transport optimal. Si le Théorème 0.1 s’applique indifféremment à ces deux notions de courbure, nous ne détaillerons ici que le cas de la courbure au sens d’Ollivier.

Si l’on suppose, en plus de l’irréductibilité, que le support de la matrice P est symétrique, alors on définit naturellement une distance sur \mathcal{X} en posant

$$\text{dist}(x, y) = \min\{k \geq 0, P^k(x, y) > 0\}.$$

On appelle diamètre de \mathcal{X} la plus grande distance entre deux points :

$$\text{diam}(\mathcal{X}) = \max_{x, y \in \mathcal{X}} \text{dist}(x, y).$$

La distance de Wasserstein entre deux lois μ et ν sur \mathcal{X} est définie à partir de dist par

$$\mathcal{W}(\mu, \nu) = \inf \mathbf{E} [\text{dist}(X, Y)],$$

où l’infimum porte sur tous les couplages (X, Y) avec $X \sim \mu$ et $Y \sim \nu$. On appelle alors courbure d’Ollivier-Ricci le plus grand réel κ tel que, pour tous $x, y \in \mathcal{X}$ et pour tout $t \geq 0$,

$$\mathcal{W}(\mathcal{P}_t(x, \cdot), \mathcal{P}_t(y, \cdot)) \leq e^{-\kappa t} \text{dist}(x, y).$$

Par dualité de Kantorovich, cela équivaut à

$$\|\mathcal{P}_t f\|_{\text{LIP}} \leq e^{-\kappa t} \|f\|_{\text{LIP}},$$

pour tout $t \geq 0$ et pour toute fonction $f : \mathcal{X} \rightarrow \mathbb{R}$, avec

$$\|f\|_{\text{LIP}} = \max_{\substack{x, y \in \mathcal{X} \\ x \neq y}} \frac{|f(y) - f(x)|}{\text{dist}(x, y)}.$$

Pour établir qu’une chaîne a une courbure positive, il suffit de montrer que le noyau P rapproche les trajectoires en un pas, et l’on peut même se restreindre à des points de départ voisins : si pour tous x et y avec $\text{dist}(x, y) = 1$, on a

$$\mathcal{W}(P(x, \cdot), P(y, \cdot)) \leq 1 - \kappa,$$

alors la courbure est au moins égale à κ .

Une conséquence simple mais cruciale d’une courbure positive est la suivante.

LEMME 4.1. — Si $\kappa \geq 0$, alors pour tout $t \geq 0$ et pour toute fonction $f: \mathcal{X} \rightarrow \mathbb{R}$, on a

$$\mathbf{Var}_{\mu_t}(f) \leq \min \left\{ t, \frac{1}{2\kappa} \right\} \|f\|_{LIP}^2,$$

où $\mu_t = \mu_0 \mathcal{P}_t$ et où \mathbf{Var}_{μ_t} correspond à la variance sous la loi μ_t .

Preuve du Lemme 4.1. — La décomposition suivante de la variance peut se vérifier en dérivant de chaque côté :

$$\mathbf{Var}_{\mu_t}(f) = \int_0^t \sum_{x \in \mathcal{X}} \mu_{t-s}(x) \sum_{y \in \mathcal{X}} P(x, y) (\mathcal{P}_s f(x) - \mathcal{P}_s f(y))^2 ds.$$

On a donc

$$(12) \quad \mathbf{Var}_{\mu_t}(f) \leq \int_0^t \|\mathcal{P}_s f\|_{LIP}^2 ds \leq \int_0^t e^{-2\kappa s} \|f\|_{LIP}^2 ds \leq t \|f\|_{LIP}^2,$$

où la dernière inégalité utilise le fait que $\kappa \geq 0$. □

En appliquant le Lemme 4.1 à la fonction $f: y \mapsto \log \frac{\mathcal{P}_t(x, y)}{\pi(y)}$ et en prenant le maximum sur $x \in \mathcal{X}$, on arrive à

$$\mathbf{V}^*(t) \leq t \max_{x \in \mathcal{X}} \left\| \log \frac{\mathcal{P}_t(x, \cdot)}{\pi(\cdot)} \right\|_{LIP}^2.$$

Le théorème 0.1 découle alors des deux lemmes suivants, dont les preuves peuvent être trouvées dans Salez (2023a).

LEMME 4.2. — Si le support de P est symétrique, alors pour tout $x \in \mathcal{X}$ et pour tout $t \geq \frac{\text{diam}(\mathcal{X})}{4}$, on a

$$\left\| \log \frac{\mathcal{P}_t(x, \cdot)}{\pi(\cdot)} \right\|_{LIP} \leq 3 \left(1 + \log \left(\frac{1}{\delta} \right) \right),$$

où δ est la plus petite entrée non nulle de P .

LEMME 4.3. — Pour tout $\varepsilon \in]0, 1[$, on a

$$\text{diam}(\mathcal{X}) \leq 2 \left(t_{\text{mix}}(\varepsilon) + \sqrt{\frac{2t_{\text{mix}}(\varepsilon)}{1-\varepsilon}} + \sqrt{\frac{2t_{\text{rel}}}{1-\varepsilon}} \right).$$

Notons qu'à l'étape (12), on aurait pu majorer l'intégrale par $\frac{1}{2\kappa}$ et obtenir

$$\mathbf{Var}_{\mu_t}(f) \leq \frac{1}{2\kappa} \|f\|_{LIP}^2.$$

Cela nous donne une version raffinée du Théorème 0.1, qui suppose d'avoir plus d'informations sur la courbure : si une chaîne à courbure positive est telle que pour tout $\varepsilon \in]0, 1[$,

$$(13) \quad t_{\text{mix}}(\varepsilon) \gg \frac{t_{\text{rel}} \log(1/\delta)}{\sqrt{\kappa}},$$

alors elle présente le cutoff.

Références

- David Aldous (1983). « Random walks on finite groups and rapidly mixing Markov chains », in : *Seminar on probability, XVII*. T. 986. Lecture Notes in Math. Springer, Berlin, p. 243-297.
- Dominique Bakry et Michel Émery (2006). « Diffusions hypercontractives », in : *Séminaire de Probabilités XIX 1983/84 : Proceedings*. Springer, p. 177-206.
- Riddhipratim Basu, Jonathan Hermon et Yuval Peres (2017). « Characterization of cutoff for reversible Markov chains », *The Annals of Probability* **45** (3), p. 1448-1487.
- Dave Bayer et Persi Diaconis (1992). « Trailing the dovetail shuffle to its lair », *The Annals of Applied Probability*, p. 294-313.
- Sergey G. Bobkov et Prasad Tetali (2006). « Modified logarithmic Sobolev inequalities in discrete settings », *Journal of Theoretical Probability* **19**, p. 289-336.
- Guan-Yu Chen et Laurent Saloff-Coste (2008). « The cutoff phenomenon for ergodic Markov processes », *Electronic Journal of Probability* **13** (3), p. 26-78.
- Persi Diaconis (1996). « The cutoff phenomenon in finite Markov chains », *Proc. Nat. Acad. Sci. U.S.A.* **93** (4), p. 1659-1664.
- Persi Diaconis et Mehrdad Shahshahani (1981). « Generating a random permutation with random transpositions », *Probability Theory and Related Fields* **57** (2), p. 159-179.
- Jian Ding, Eyal Lubetzky et Yuval Peres (2010). « Total variation cutoff in birth-and-death chains », *Probability theory and related fields* **146** (1-2), p. 61-85.
- Cyril Labbé et Hubert Lacoïn (2019). « Cutoff phenomenon for the asymmetric simple exclusion process and the biased card shuffling », *Annals of Probability* **47** (3), p. 1541-1586.
- Hubert Lacoïn (2016). « Mixing time and cutoff for the adjacent transposition shuffle and the simple exclusion », *Annals of Probability* **44** (2), p. 1426-1487.
- David A. Levin, Malwina J. Luczak et Yuval Peres (2010). « Glauber dynamics for the mean-field Ising model : cut-off, critical power law, and metastability », *Probability Theory and Related Fields* **146** (1-2), p. 223-265.
- Eyal Lubetzky et Yuval Peres (2016). « Cutoff on all Ramanujan graphs », *Geometric and Functional Analysis* **26** (4), p. 1190-1216.
- Eyal Lubetzky et Allan Sly (2011). « Explicit expanders with cutoff phenomena », *Electronic Journal of Probability* **16**, p. 419-435.
- (2013). « Cutoff for the Ising model on the lattice », *Inventiones mathematicae* **191** (3), p. 719-755.
- Alon Nilli (1991). « On the second eigenvalue of a graph », *Discrete Mathematics* **91** (2), p. 207-210.
- Yann Ollivier (2009). « Ricci curvature of Markov chains on metric spaces », *Journal of Functional Analysis* **256** (3), p. 810-864.

- Yuval Peres (2004). « American Institute of Mathematics (AIM) research workshop “Sharp Thresholds for Mixing Times” (Palo Alto, December 2004) », *Summary available at <http://www.aimath.org/WWN/mixingtimes>*.
- Justin Salez (2023a). « Cutoff for non-negatively curved Markov chains », *Journal of the European Mathematical Society*.
- (2023b). « The varentropy criterion is sharp on expanders », *arXiv preprint arXiv:2307.10066*.
- Laurent Saloff-Coste (2004). « Random walks on finite groups », in : *Probability on discrete structures*. Springer, p. 263-346.

Anna Ben-Hamou

Sorbonne Université, LPSM

4, place Jussieu 75005 Paris

E-mail : `anna.ben_hamou@sorbonne-universite.fr`