Research stay, master's degree (2024)
Supervisor: Djalil Chafaï

Cutoff phenomenon for ergodic diffusion processes

By Samuel Chan-ashing

Abstract

The cutoff phenomenon for ergodic Markov processes was discovered towards the end of the last century, notably by David Aldous and Persi Diaconis, in the case of random walks on finite spaces. Since then, it has been the subject of intense exploration by the probabilistic community worldwide. Despite the accumulation of numerous results, there is still no general theory. The case of diffusion processes has been explored in detail, particularly by Laurent Saloff-Coste, in connection with notions of curvature, dimension, and diameter, as well as Sobolev-type functional inequalities. The present research stay memoir explores the cutoff phenomenon in various Markov processes, highlighting the challenges posed by existing methods and proposing directions for future research. This work builds on recent results and seeks to generalize them to different types of diffusions and spaces, pursuing two main objectives concerning this phenomenon: (1) to complete the study conducted by Chen and Saloff-Coste [10] by observing their cutoff condition for various distances, (2) to examine how the cutoff condition found by Salez in [23] for curved finite Markov chains can be more generally applied to diffusion processes under curvature conditions, both in compact and non-compact spaces.

Contents

1 Introduction to the cutoff phenomenon **2**
1.1 Cutoff Phenomenon 2
1.2 Distance selection and comparative analysis 4
1.3 Toy model: Dyson-Orstein-Uhlenbeck process 6

2 Cutoff for relaxed processes **9**
2.1 Relaxation condition for cutoff 9
2.2 Spectral criterion 10
2.3 Conclusion and perpectives 12

3 Varentropic method **13**
3.1 Main result 13
3.2 Proof 14
3.3 Applications for finite Markov chains 17
3.4 The Ornstein-Uhlenbeck case is not concerned 18
3.5 Conclusion and perspectives 20

Keywords: Cutoff, Diffusion processes, Dyson-Orstein-Uhlenbeck, Mixing time
Research stay for the Master 2 Probabilités et Modèles Aléatoires at Sorbonne Université, Département de Mathématiques et Applications (DMA), École Normale Supérieure (Paris).

1. Introduction to the cutoff phenomenon

Recently, the *cutoff phenomenon* has been examined for Dyson diffusions arising from random matrix theory. Also, the case of finite Markov chains with total variation distance has recently been thoroughly studied by Justin Salez using the notion of varentropy, an approach that avoids upper and lower bounds.

1.1. Cutoff Phenomenon

Let us consider an ergodic Markov process $\mathrm{X} = (\mathrm{X}_t)_{t\geqslant 0}$ with a state space S and a unique invariant law π for which

$$\lim_{t\to\infty} \mathrm{dist}(\mathrm{Law}(\mathrm{X}_t) \mid \pi) = 0,$$

where $\mathrm{dist}(\cdot \mid \cdot)$ is a distance or divergence on probability measures on S.

Suppose now that $\mathrm{X} = \mathrm{X}^n$ depends on a parameter of dimension, size, or complexity parameter n, and let us set $\mathrm{S} = \mathrm{S}^n$, $\pi = \pi^n$, and $\mathrm{X}_0 \sim \mu_0^n \in \mathcal{P}(\mathrm{S}^n)$.

Examples 1.
- Sequence of transition matrices (P^n) on a fixed finite space.
- System of n particles: n independent one-dimensional Ornstein-Uhlenbeck processes.
- Brownian motion on the sphere $\mathbb{S}^n$.

Observations (Cutoff Phenomenon). In many of such examples, it has been proved that when n is sufficiently enough, the supremum over a certain set of initial conditions μ_0^n of the quantity $\mathrm{dist}(\mathrm{Law}(\mathrm{X}_t^n) \mid \pi^n)$, collapses abruptly to 0 when t crosses a critical value $c = c_n$ which may depend on n.

More precisely, if dist ranges from 0 to max, then, for some sets of initial conditions $\mathrm{S}_0^n \subset \mathcal{P}(\mathrm{S}^n)$, we define the cutoff phenomenon as follow :

Definition 2 (Cutoff)**.** We say that the sequence (X_t^n) exhibits a **cutoff** if there exists some critical value $c = c_n$ such that for all $\eta \in (0,1)$,

$$\lim_{n\to\infty} \sup_{\mu_0^n \in \mathrm{S}_0^n} \mathrm{dist}(\mathrm{Law}(\mathrm{X}_{t_n}^n) \mid \pi^n) = \begin{cases} \max & \text{if } t_n = (1-\eta)c_n, \\ 0 & \text{if } t_n = (1+\eta)c_n. \end{cases}$$

To quantify this critical value, we introduce the following quantity:

Definition 3 (Mixing Time)**.** For an arbitrarily small threshold $\varepsilon > 0$,

$$t_{\mathrm{mix}}^n(\varepsilon) := t_{\mathrm{mix}}^{n,\mathrm{dist}}(\varepsilon, \mathrm{S}_0^n) := \inf\{t \geqslant 0 : \sup_{\mu_0^n \in \mathrm{S}_0^n} \mathrm{dist}(\mathrm{Law}(\mathrm{X}_t^n) \mid \pi^n) \leqslant \varepsilon\}$$

Of course, such a definition fully makes sense as soon as the following monotonicity condition is satisfied:

Definition 4 (Monotonicity)**.** The process is **monotone** for dist if from a certain rank n:

$$\mathrm{dist}_n^* : t \longmapsto \sup_{\mu_0^n \in \mathrm{S}_0^n} \mathrm{dist}(\mathrm{Law}(\mathrm{X}_t^n), \pi^n) \text{ is decreasing.}$$

Remark 5. From the definition of the mixing time, we hence have that for all $t \geqslant 0$:

$$t > t_{\mathrm{mix}}^n(\varepsilon) \quad \Longrightarrow \quad \sup_{\mu_0^n \in \mathrm{S}_0^n} \mathrm{dist}(\mathrm{Law}(\mathrm{X}_t^n) \mid \pi^n) \leqslant \varepsilon,$$

$$t < t_{\mathrm{mix}}^n(\varepsilon) \quad \Longrightarrow \quad \sup_{\mu_0^n \in \mathrm{S}_0^n} \mathrm{dist}(\mathrm{Law}(\mathrm{X}_t^n) \mid \pi^n) > \varepsilon.$$

The following proposition provides us a second viewpoint (temporal) of cutoff phenomenon: the distance to equilibrium remains close to its maximal value for a long time, and then suddenly drops to zero on a much shorter time-scale. We will then explore these two different perspectives further.

Proposition 6 (Characterization). *Supposed that* max $=1$ *and that* (X_t^n) *is monotone for dist. Then the process* (X_t^n) *exhibits a cutoff if and only if:*

$$\forall\,\varepsilon\in(0,1)\ :\quad \frac{t^n_{mix}(1-\varepsilon)}{t^n_{mix}(\varepsilon)} \;=\; 1+\underset{n\to\infty}{o}(1).$$

Proof. "$\Longrightarrow$" : Let c_n be the critical value as in the definition.

(1) (pseudo-uniqueness): Let's first show that if there exists a second critical value d_n, then $d_n\sim c_n$. By contradiction, let us suppose that:

$$\exists\,\eta\in(0,1),\quad \forall\,n\in\mathbb{N},\quad \exists\,\varphi(n)\geqslant n\ :\quad \left|\frac{c_{\varphi(n)}}{d_{\varphi(n)}}-1\right| > \eta,$$

i.e. by extracting a subsequence, we can assume that:

$$\forall\,n\geqslant 1\ :\quad c_{\varphi(n)} > (1+\eta)d_{\varphi(n)}\qquad (\text{or } c_{\varphi(n)} < (1-\eta)d_{\varphi(n)}).$$

By the hypothesis on c_n, we have:

$$\lim_{n\to\infty}\mathrm{dist}^*_{\varphi(n)}((1+\eta)d_{\varphi(n)}) \underset{n\to\infty}{\longrightarrow} \max = 1,$$

which is absurd by the definition of d_n.

(2) (pseudo-uniqueness): From (1), it suffices to show that, for all $0<\varepsilon<1$, $t^n_{\mathrm{mix}}(\varepsilon)$ is a critical value. By monotonicity and by definition of the infimum:

$$\forall\,t\geqslant t^n_{\mathrm{mix}}(\varepsilon)\ :\quad \mathrm{dist}^*_n(t)\leqslant\varepsilon\qquad\text{and}\qquad \forall\,t< t^n_{\mathrm{mix}}(\varepsilon)\ :\quad \mathrm{dist}^*_n(t)>\varepsilon.$$

Hence, for all $\eta\in(0,1)$:

$$\lim_{n\to\infty}\mathrm{dist}^*_n((1+\eta)t^n_{\mathrm{mix}}(\varepsilon))\leqslant\varepsilon\qquad\text{and}\qquad \lim_{n\to\infty}\mathrm{dist}^*_n((1-\eta)t^n_{\mathrm{mix}}(\varepsilon))>\varepsilon.$$

"$\Longleftarrow$" : Let's find the critical value c_n from the definition.

(1) Let's show that there exists (c_n) such that for every $\varepsilon\in(0,1)$, $c_n\sim t^n_{\mathrm{mix}}(\varepsilon)$ (independent of ε). For $0<\varepsilon<\varepsilon'<\frac12$ (so $1-\varepsilon>1-\varepsilon'$), by the definition of mixing time :

$$t^n_{\mathrm{mix}}(\varepsilon')\leqslant t^n_{\mathrm{mix}}(\varepsilon)\qquad\text{and}\qquad t^n_{\mathrm{mix}}(1-\varepsilon)\leqslant t^n_{\mathrm{mix}}(1-\varepsilon')\leqslant t^n_{\mathrm{mix}}(\varepsilon')\leqslant t^n_{\mathrm{mix}}(\varepsilon),$$

and $t^n_{\mathrm{mix}}(\varepsilon)\sim t^n_{\mathrm{mix}}(1-\varepsilon)$. Denoting $c_n := t^n_{\mathrm{mix}}(\varepsilon)\sim t^n_{\mathrm{mix}}(\varepsilon')$, we obtain the result.

(2) Let $\eta,\varepsilon\in(0,1)$. Since $c_n\sim t^n_{\mathrm{mix}}(\varepsilon)$, there exists $n_{\varepsilon,\eta}\in\mathbb{N}$, such that:

$$\forall\,n\geqslant n_{\varepsilon,\eta}\ :\quad \left|\frac{t^n_{\mathrm{mix}}(\varepsilon)}{c_n}-1\right|\leqslant\eta\qquad\Longrightarrow\qquad \forall\,n\geqslant n_{\varepsilon,\eta}\ :\quad |t^n_{\mathrm{mix}}(\varepsilon)-c_n|\leqslant\eta c_n.$$

Hence, for every $\varepsilon\in(0,1)$, there exists $n_{\varepsilon,\eta}\in\mathbb{N}$ such that for every $n\geqslant n_{\varepsilon,\eta}$:

$$(1+\eta)c_n > t^n_{\mathrm{mix}}(\varepsilon) > (1-\eta)c_n.$$

Thus, by Remark 5 :

$$\forall\,\varepsilon>0,\quad \exists\,n_{\varepsilon,\eta}\in\mathbb{N},\quad \forall\,n\geqslant n_{\varepsilon,\eta}\ :\quad \mathrm{dist}^*_n((1+\eta)c_n)\leqslant\varepsilon,$$

hence

$$\forall\,\varepsilon>0\ :\quad \lim_{n\to\infty}\mathrm{dist}^*_n((1+\eta)c_n)\leqslant\varepsilon.$$

We do the same with $t^n_{\mathrm{mix}}(1-\varepsilon)$.

□

Remark 7. We do not need the process to be reversible or the state space to be compact. The analysis can be easily refined using a cutoff window concept.

Now that we have formalized this phenomenon, we specify the two main components used:

(*i*) distances or divergences,

(*ii*) the type of ergodic Markov process: diffusions.

1.2. **Distance selection and comparative analysis**

The cutoff phenomenon is intrinsically linked to the chosen distance. Therefore, we quantify the tendency to equilibrium of ergodic Markov processes for the following standard distances or divergences: Wasserstein, total variation (TV), Hellinger, Kullback Entropy, χ^2, $\mathrm{L}^p(\pi)$ and Fisher. It is important to note that for comparable distances or divergences (in a certain sense of comparison and for a particular process), it is sufficient to address the cutoff for the simplest distance to use.

Distances

- **Transport Distance: Wasserstein-Kantorovich-Monge**

Definition 8 (Transport Distance: Wasserstein-Kantorovich-Monge)**.** The Wasserstein distance of order 2 and with respect to the underlying Euclidean distance is defined for all probability measures μ and ν on $\mathbb{R}^n$ by

$$\mathrm{Wasserstein}(\mu,\nu) = \left(\inf_{\substack{X\sim\mu \\ Y\sim\nu}} \mathbb{E}\left[|X-Y|^2\right] \right)^{1/2} \in [0,+\infty]$$

where $|x| = \sqrt{x_1^2 + \cdots + x_n^2}$.

- **Total Variation Distance**

Definition 9 (Total Variation Distance)**.** The total variation distance between probability measures μ and ν on the same space is given by:

$$\|\mu-\nu\|_{\mathrm{TV}} = \sup_{A\in\mathcal{B}_{\mathrm{X}}} |\mu(A)-\nu(A)| \in [0,1].$$

Proposition 10 (Total Variation Distance Densities)**.** *If μ and ν are absolutely continuous with respect to a reference measure λ with densities f_μ and f_ν, then:*

$$\|\mu-\nu\|_{\mathrm{TV}} = \frac{1}{2}\int |f_\mu - f_\nu|\,\mathrm{d}\lambda = \frac{1}{2}\|f_\mu - f_\nu\|_{\mathrm{L}^1(\lambda)}.$$

- **Hellinger Distance**

Definition 11 (Hellinger Distance)**.** The Hellinger distance between probability measures μ and ν with densities f_μ and f_ν with respect to the same reference measure λ is given by:

$$\mathrm{Hellinger}(\mu,\nu) = \frac{1}{\sqrt{2}}\left\|\sqrt{f_\mu}-\sqrt{f_\nu}\right\|_{\mathrm{L}^2(\lambda)} = \left(1-\int\sqrt{f_\mu f_\nu}\mathrm{d}\lambda\right)^{1/2} \in [0,1].$$

Remark 12. This quantity does not depend on the choice of λ. Note that an alternative normalization is sometimes considered in the literature, making the maximum value of the Hellinger distance equal to $\sqrt{2}$.

- **Relative Entropy: Kullback-Leibler Divergence**

Definition 13 (Relative Entropy: Kullback-Leibler Divergence)**.** The Kullback-Leibler divergence or relative entropy is defined by

$$\mathrm{Kullback}(\nu\mid\mu) = \begin{cases} \displaystyle\int \log\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\mathrm{d}\nu = \int \frac{\mathrm{d}\nu}{\mathrm{d}\mu}\log\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\mathrm{d}\mu \in [0,+\infty] & \text{si } \nu \ll \mu \\ \infty & \text{sinon.} \end{cases}$$

- **Chi-square Divergence**

Definition 14 (Chi-square Divergence)**.** The chi-square divergence is given by

$$\chi^2(\nu \mid \mu) = \begin{cases} \left\| \frac{\mathrm{d}\nu}{\mathrm{d}\mu} - 1 \right\|^2_{\mathrm{L}^2(\mu)} = \left\| \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \right\|^2_{\mathrm{L}^2(\mu)} - 1 \in [0, +\infty] & \text{si } \nu \ll \mu \\ \infty & \text{sinon.} \end{cases}$$

Proposition 15. *If μ and ν have densities f_μ and f_ν with respect to a reference measure λ then:*

$$\chi^2(\nu \mid \mu) = \int (f_\nu^2 / f_\mu)\, \mathrm{d}\lambda - 1.$$

- $\mathrm{L}^p(\pi)$ **Distance**

Definition 16 (Distance $\mathrm{L}^p(\pi)$ to equilibrium)**.** For any $p \in [1, \infty]$ and $t \geqslant 0$, it is defined as follows:

$$\|\mu_t - \pi\|_p := \begin{cases} \|h_t - 1\|_{\mathrm{L}^p(\pi)} & \text{if } \mu_t \ll \pi \text{ with } \frac{d\mu_t}{d\pi} = h_t, \\ \frac{1}{2} & \text{if } \mu_t \not\ll \pi \text{ and } p = 1, \\ \infty & \text{if } \mu_t \not\ll \pi \text{ and } p > 1. \end{cases}$$

Proposition 17 (Duality)**.** *For any (p, q) conjugate pairs, and if $\mu_t \ll \pi$ with $\frac{d\mu_t}{d\pi} = h_t$, then:*

$$\|\mu_t - \pi\|_p = \sup\{(\mu_t - \pi)(g) : g \in \mathrm{L}^q(\pi), \|g\|_{\mathrm{L}^q} = 1\}.$$

In particular, the cases $p \in \{1, 2\}$ correspond respectively to the χ^2 divergence and the total variation distance:

$$\|\mu_t - \pi\|_{\mathrm{L}^1} = 2\, \|\mu_t - \pi\|_{\mathrm{TV}} \quad \text{and} \quad \|\mu_t - \pi\|^2_{\mathrm{L}^2} = \chi^2(\mu_t \| \pi).$$

Proof. The $\mathrm{L}^p - \mathrm{L}^q$ duality yields:

$$\|h_t - 1\|_{\mathrm{L}^p} = \sup \left\{ \int g(h_t - 1) d\pi : g \in L^q(\pi), \|g\|_{\mathrm{L}^q} = 1 \right\},$$

So, if $\mu_t \ll \pi$, then

$$\|\mu_t - \pi\|_p = \sup\{(\mu_t - \pi)(g) : g \in \mathrm{L}^q(\pi), \|g\|_{\mathrm{L}^q} = 1\}.$$

In the cases $p \in \{1, 2\}$:

$$\|h_t - 1\|_2^2 = \int (h_t^2 - 2h_t + 1) d\pi = \|h_t\|_2^2 - 1 = \mathrm{Var}_\pi(h_t) = \chi^2(\mu_t \| \pi),$$

$$\|h_t - 1\|_1 = \sup_{f \in \mathrm{L}^\infty(\pi)} \int (h_t - 1) f d\pi = 2 \sup_A |\mu_t(A) - \pi(A)| = 2\, \|\mu_t - \pi\|_{\mathrm{TV}}.$$

□

- **Divergence: Fisher Information**

Definition 18 (Divergence: Fisher Information)**.** The (logarithmic) Fisher information or divergence is defined by

$$\mathrm{Fisher}(\nu \mid \mu) = \int \left| \nabla \log \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \right|^2 \mathrm{d}\nu = \int \frac{\left| \nabla \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \right|^2}{\frac{\mathrm{d}\nu}{\mathrm{d}\mu}} \mathrm{d}\mu = 4 \int \left| \nabla \sqrt{\frac{\mathrm{d}\nu}{\mathrm{d}\mu}} \right|^2 \mathrm{d}\mu \in [0, +\infty]$$

if ν is absolutely continuous with respect to μ, and Fisher $(\nu \mid \mu) = +\infty$ otherwise.

Comparative analysis

Let dist_1, dist_2 be two distances/divergences.

Definition 19 (Domination)**.** Let $(f_n)_{n\geqslant 1}$ be a sequence of functions such that each f_n is a continuous, increasing function with $f_n(0) = 0$. We say that dist_2 *dominates* dist_1 under (f_n) for (X_t) if for all $n \geqslant 1$:

$$\forall t \geqslant 0 : \quad \mathrm{dist}_1(\mu_t \mid \pi) \leqslant f_n(\mathrm{dist}_2(\mu_t \mid \pi)).$$

Proposition 20 (Comparison of Mixing Times)**.** *If* dist_2 *dominates* dist_1 *under* (f_n) *for* (X_t)*, then for* $\varepsilon \in (0,1)$*:*

$$t_{mix}^{n,\mathrm{dist}_1}(f_n(\varepsilon)) \leqslant t_{mix}^{n,\mathrm{dist}_2}(\varepsilon).$$

Proof. Since f_n is increasing, for all $\varepsilon \in (0,1)$ and $t \geqslant 0$:

$$\mathrm{dist}_2(\mu_t \mid \pi) \leqslant \varepsilon \quad \Longrightarrow \quad \mathrm{dist}_1(\mu_t \mid \pi) \leqslant f_n(\varepsilon).$$

□

Proposition 21 (Comparison Product Condition)**.** *Supposed that* $\max = 1$ *and that* (X_t^n) *is monotone for* dist_1 *and* dist_2*. Let* dist_1*,* dist_2 *be such that* dist_2 *(resp .* dist_1*) dominates* dist_1 *(resp .* dist_2*) under* (f_n) *(resp .* (g_n)*) for* (X_t)*. If the process* (X_t^n) *exhibits a cutoff for* dist_1*, then it exhibits a cutoff for* dist_2*.*

Proof. Straightforward since we have:

$$t_{\mathrm{mix}}^{n,\mathrm{dist}_2}(g_n \circ f_n(\varepsilon)) \leqslant t_{\mathrm{mix}}^{n,\mathrm{dist}_1}(f_n(\varepsilon)) \leqslant t_{\mathrm{mix}}^{n,\mathrm{dist}_2}(\varepsilon).$$

□

1.3. **Toy model: Dyson-Orstein-Uhlenbeck process**

There is no general theory on cutoff, so our goal is to contribute insights and ideas to this theory. We focus mainly on diffusions and aim to establish a general criterion. To do this, we will use a simple toy model of diffusion: the Ornstein-Uhlenbeck process. A slightly more complex toy model will then be the Dyson-Ornstein-Uhlenbeck process. There are other relevant examples that we will describe later, namely:
− Brownian motion on the torus and on the sphere,
− The $(\mathrm{M/M/}\infty)^n$ queueing process,
− Dyson type dynamics associated to the Laguerre and Jacobi beta ensembles.

Ornstein-Uhlenbeck process

Among the few models for which the SDE has an explicit solution in terms of stochastic component, we find equations whose drift $(b(t,x))$ and diffusion coefficient $(\sigma(t,x))$ are linear in the state variable x. The Ornstein-Uhlenbeck process is one such example with the additional property that its law is completely computable.

Definition 22 (Ornstein-Uhlenbeck Process)**.** Any process defined by the stochastic differential equation:

$$\mathrm{dX}_t = -\theta \mathrm{X}_t \,\mathrm{d}t + \sigma \,\mathrm{dB}_t, \qquad \mathrm{X}_0 \sim \mu_0(\theta,\sigma),$$

where $\theta \in \mathbb{R}$, $\sigma > 0$ are constants, and $\mu_0(\theta,\sigma)$ is independent of B.

Remark 23. This model is particularly rich and interesting from a statistical perspective and can be studied from various angles since all computations are explicit (which is exceptional for a diffusion model).

The infinitesimal generator of an OU process is to be interpreted as follows: the differential of f at x is a linear form on E, which is applied to the vector x itself.

Proposition 24 (Infinitesimal Generator). *It is the second-order elliptic operator defined for all $f \in \mathrm{C}^2(\mathbb{R}^n)$ by:*

$$\mathrm{G}f \;=\; \frac{\sigma^2}{2}\sum_{i=1}^{n}\partial^2_{ii}f - \theta\sum_{i=1}^{n}x_i\partial_i f \;=\; \frac{\sigma^2}{2}\Delta f - \theta x\cdot\nabla f.$$

Theorem 25 (Mehler Formula). *Any OU process in $\mathbb{R}^n$ can be written as:*

$$\mathrm{X}_t \;=\; xe^{-\theta t} + \sigma\int_0^t e^{\theta(s-t)}\mathrm{dB}_s.$$

In particular, the OU process is a (non-centered) Gaussian process and for any $t \geqslant 0$:

$$\mathrm{X}_t \;\sim\; \mu_t^x \;:=\; \mathcal{N}\left(xe^{-\theta t}, \frac{\sigma^2}{2}\frac{1-e^{-2\theta t}}{\theta}\mathrm{I}_n\right).$$

Moreover, its coordinates are independent one-dimensional OU processes with initial condition x^i and a unique invariant law $\mathcal{N}\left(0, \frac{\sigma^2}{2\theta}\right)$.

Dyson-Ornstein-Uhlenbeck process

A Dyson process is a diffusion that describes a system of n one-dimensional Brownian particles interacting, subject to a confinement potential V and pairwise singular repulsion $\beta \geqslant 0$. This kind of process was essentially discovered by Dyson (in the cases $\beta \in \{1,2,4\}$ and for Coulomb interaction), as it describes the dynamics of the eigenvalues of symmetric/hermitian/symplectic random $n \times n$ matrices with independent Ornstein-Uhlenbeck entries.
The Dyson-Ornstein-Uhlenbeck process is the case where the confinement potentiel comes from Coulomb interaction :

$$\mathrm{V}(x) \;=\; \frac{x^2}{2}.$$

Definition 26 (DOU process). Any solution $\mathrm{X} := \mathrm{X}^n$ on $\mathbb{R}^n$ of the SDE defined for all $1 \leqslant i \leqslant n$ by:

$$\mathrm{dX}_t^i \;=\; \left(-\mathrm{X}_t^i \;+\; \frac{1}{n}\sum_{j\neq i}^{n}\frac{\beta}{\mathrm{X}_t^i - \mathrm{X}_t^j}\right)\mathrm{d}t \;+\; \sqrt{\frac{2}{n}}\mathrm{dB}_t, \qquad \mathrm{X}_0 \;=\; x_0 \in \mathbb{R}^n,$$

where $\beta \geqslant 0$ is the singular repulsion.

Remark 27 (Non-interacting case). Note that in the case without interaction ($\beta = 0$), the n particles are therefore independent one-dimensional Ornstein-Uhlenbeck processes. In particular, they still collide but since they do not interact, this poses no problem.

Remark 28 (Mean-field type parametrization).
- (Scale): We take an inverse temperature of order n in order to obtain a mean field limit without modifying the process over time.
- (Empirical measure): The process is actually a solution of the SDE:

$$\mathrm{dX}_t^i \;=\; b(\mathrm{X}_t^i, \hat{\mu}_t^N)\mathrm{d}t \;+\; \sigma(\mathrm{X}_t^i, \hat{\mu}_t^N)\mathrm{dB}_t, \qquad \mathrm{X}_0 \;=\; x_0 \in \mathbb{R}^n,$$

  where the drift and diffusion coefficient are given respectively by:

$$b(x,\mu) \;=\; -x + \int \frac{\beta}{|x-y|}\,\mu(\mathrm{d}y) \quad \text{and} \quad \sigma(x,\mu) \;=\; \sigma \;:=\; \text{const} > 0.$$

The drift in the above SDE is the gradient of the following function which can be interpreted as the energy of the configuration of particles $x_1,\ldots,x_n$.

Definition 29 (Energy)**.** The **energy** related to the confinement potential $\mathrm{V} \in \mathrm{C}^2(\mathbb{R},\mathbb{R})$ is:

$$\mathrm{E} \,:\, (x_1,\dots,x_n) \longmapsto n\sum_{i=1}^{n}\frac{x_i^2}{2} + \beta\sum_{i>j}\log\frac{1}{|x_i - x_j|}.$$

Remark 30 (DOU SDE in term of energy). The SDE can be rewritten as:

$$\mathrm{dX}_t^i \;=\; -\frac{1}{n}\nabla\mathrm{E}(\mathrm{X}_t)\mathrm{d}t \;+\; \sqrt{\frac{2}{n}}\mathrm{dB}_t, \qquad \mathrm{X}_0 \;=\; x_0 \in \mathrm{D}^n.$$

Proposition 31 (Infinitesimal Generator)**.** *It is the second-order elliptic operator defined for all $f \in \mathrm{C}^2(\mathbb{R}^d)$ by:*

$$\mathrm{G} \;:=\; \frac{1}{n}\left(\Delta - \nabla\mathrm{E}\cdot\nabla\right).$$

In particular, for all $f \in \mathrm{C}^2(\mathbb{R}^d)$ we have:

$$\begin{aligned}\mathrm{G}(f) \;&=\; \frac{1}{n}\sum_{i=1}^{n}\partial_{ii}^2 f - \sum_{i=1}^{n}x_i\partial_i f + \frac{\beta}{n}\sum_{\substack{i,j=1\\ j\neq i}}^{n}\frac{\partial_i f}{x_i - x_j}\\ &=\; \frac{1}{n}\sum_{i=1}^{n}\partial_{ii}^2 f - \sum_{i=1}^{n}x_i\partial_i f + \frac{\beta}{2n}\sum_{\substack{i,j=1\\ j\neq i}}^{n}\frac{\partial_i f - \partial_j f}{x_i - x_j}\end{aligned}$$

In the case that $0 < \beta < 1$, the particles collide leading to an explosion of the drift, with positive probability. However, one can define the process for all times, for example by adding a local time term to the stochastic differential equation. It is natural to expect that the universality of the cutoff phenomenon works as for $\beta \notin (0,1)$, but for simplicity, we do not consider this case here.

Suppose now that $\beta \geqslant 1$. The Markov process X^n is not irreducible, but fortunately there is a domain such that the particles never collide, and the order of the initial particles is preserved at all times:

Proposition 32. *Let us order the coordinates by defining the following convex domain*

$$\mathrm{D}_n \;:=\; \{x \in \mathbb{R}^n : \, x_1 \,<\, \cdots \,<\, x_n\}.$$

Then, if $x_0^n \in \overline{\mathrm{D}}_n$ we have that the Dyson SDE admits a unique strong solution $(\mathrm{X}_t^n)_{t\geqslant 0}$ that never leaves D_n in the sense that:

$$\forall\, t \,>\, 0, \quad \mathrm{X}_t^n \in \mathrm{D}_n \quad a.s.$$

In particular, this is true for x_0^n such that $x_0^{n,1} \,=\, \cdots \,=\, x_0^{n,n}$.

Proposition 33 (Invariant law)**.** D_n *is a recurrent class carrying a unique invariant law which is reversible and given by:*

$$\pi_\beta^n \;:=\; \frac{e^{-\mathrm{E}(x_1,\dots,x_n)}}{C_n^\beta}\mathbb{1}_{(x_1,\dots,x_n)\in\overline{\mathrm{D}}_n}\mathrm{d}x_1\cdots\mathrm{d}x_n,$$

where C_n^β is the normalization factor given by:

$$C_n^\beta \;:=\; \int_{\overline{\mathrm{D}}_n} e^{-\mathrm{E}(x_1,\dots,x_n)}\mathrm{d}x_1\cdots\mathrm{d}x_n.$$

Remark 34 (Explicite formulation).

$$\pi_\beta^n \;:=\; \frac{\mathbb{1}_{(x_1,\dots,x_n)\in\overline{\mathrm{D}}_n}}{C_n^\beta}\prod_{i>j}(x_i - x_j)^\beta\prod_{i=1}^{n}e^{-nx_i^2/2}\mathrm{d}x_i.$$

Proposition 35. *For any $\beta \geqslant 0$, the invariant law π_β^n is log-concave with respect to the Lebesgue measure as well as to $\mathcal{N}\left(0, \frac{1}{n}\mathrm{I}_n\right)$.*

Proof. Since $-\log$ is convex on $(0,+\infty)$, the application $(x_1,\dots,x_n) \in \mathrm{D}_n \longmapsto \beta \sum_{i>j} \log \frac{1}{x_i - x_j}$, is convex. Hence, since V is convex on $\mathbb{R}$, it follows that E is convex on D_n.

□

Proposition 36 (Spectral gap)**.** *The spectral gap is 1 for all $n \geqslant 1$.*

Theorem 37 (Exponential convergence to equilibrium)**.** *If there exists some constant $\rho \geqslant 0$ such that $\mathrm{V} - \frac{\rho}{2}|\cdot|^2$ is convex, then the Dyson process has exponential convergence to equilibrium.*

2. Cutoff for relaxed processes

The following theorem comes from Saloff-Coste's article [10, 24]. This theorem states that under a so-called *product condition* and a relaxation condition, the process admits a cutoff. One might wonder if the product condition is necessary and sufficient for diffusions. However, in the general case, it is not, as Saloff-Coste demonstrates in [10]. We will then see that this theorem provides a very interesting spectral criterion that can be used, for example, on the D.O.U. process.

2.1. Relaxation condition for cutoff

We consider the following framework:
– $(\mathrm{P}_t)_{t\geqslant 0}$: an ergodic Markov semigroup with a state space S^n,
– π : unique invariant probability measure,
– S_0^n: set of initial conditions.
Consider a distance or divergence dist and denote, when it exists :

$$\mathrm{dist}_n^* : t \longmapsto \sup_{\mu_0 \in \mathrm{S}_0^n} \mathrm{dist}(\mu_0 \mathrm{P}_t^n, \pi^n) \qquad \text{and} \qquad t_{\mathrm{mix}}^n(\varepsilon) := t_{\mathrm{mix}}^{n,\mathrm{d}}(\varepsilon, \mathrm{S}_0^n).$$

In the following, we denote : $a \ll b$ for $a/b = o(1)$ and $a \lesssim b$ for $a/b = O(1)$.

Let us associate a relaxation time to the ergodic Markov semigroup $(\mathrm{P}_t)_{t\geqslant 0}$ (when it exists). We will see later how it is related to the spectral gap or to the curvature of this process.

Definition 38 (Relaxation Time)**.** We say that the semigroup $(\mathrm{P}_t)_{t\geqslant 0}$ has a *relaxation time* t_{rel}^n for dist over S_0^n if there exists a constant $C > 0$ such that:

$$\forall s,t \geqslant 0 : \quad \mathrm{dist}_n^*(t+s) \leqslant Ce^{-t/t_{\mathrm{rel}}^n} \mathrm{dist}_n^*(s).$$

When $C = 1$, we then say that the process is *curved.*

Remark 39 (Relaxation time for a pointwise initial condition).

Theorem 40 (Cutoff for relaxed process)**.** *Let $(\mathrm{P}_t)_{t\geqslant 0}$ be an ergodic Markov semigroup and a unique invariant law π. Suppose that :*

(i) (monotonicity) : $(\mathrm{P}_t)_{t\geqslant 0}$ is monotone for dist *over S_0^n,*
(ii) (relaxation condition) : $(\mathrm{P}_t)_{t\geqslant 0}$ has a relaxation time t_{rel}^n for dist *over S_0^n,*
(iii) (product-like condition) :

$$t_{mix}^n(\varepsilon) \gg t_{rel}^n.$$

Then, the sequence exhibits cutoff.

Remark 41. In the general case, the product condition is sufficient but not necessary for the cutoff. One may wonder if for diffusions, this condition is necessary.

Proof. Let's denote $\mathrm{d}_n^* := \mathrm{dist}_n^*$ and $t_{\mathrm{mix}}^n := t_{\mathrm{mix}}^n(\varepsilon)$ for $\varepsilon \in (0,1)$ fixed. By relaxation condition and monotonicity, we have for :

– $t = t^n_{\text{mix}}(\varepsilon)$, $s = \varepsilon t^n_{\text{mix}}(\varepsilon)$:

$$\sup_{t>(1+\varepsilon)t^n_{\text{mix}}} \mathrm{d}^*_n(t) \leqslant \mathrm{d}^*_n((1+\varepsilon)t^n_{\text{mix}}) \leqslant \mathrm{d}^*_n(t^n_{\text{mix}})Ce^{-\varepsilon t^n_{\text{mix}}/t^n_{\text{rel}}} \leqslant \varepsilon Ce^{-\varepsilon t^n_{\text{mix}}/t^n_{\text{rel}}} \xrightarrow[n+\infty]{(iii)} 0,$$

– $t = (1-\varepsilon)t^n_{\text{mix}}(\varepsilon)$, $s = \frac{\varepsilon}{2}t^n_{\text{mix}}(\varepsilon)$:

$$\inf_{t<(1-\varepsilon)t^n_{\text{mix}}} \mathrm{d}^*_n(t) \geqslant \mathrm{d}^*_n((1-\varepsilon)t^n_{\text{mix}}) \geqslant \mathrm{d}^*_n((1-\frac{\varepsilon}{2})t^n_{\text{mix}})C^{-1}e^{\frac{\varepsilon}{2}t^n_{\text{mix}}/t^n_{\text{rel}}} \geqslant \varepsilon C^{-1}e^{\frac{\varepsilon}{2}t^n_{\text{mix}}/t^n_{\text{rel}}} \xrightarrow[n+\infty]{(iii)} +\infty,$$

□

Remark 42. If for all $0 \leqslant s \leqslant t$,

The following application has already been covered in [6], but we simply provide an example of how the previous introduction can, in some cases, easily allow one to conclude the existence of a cutoff through comparison.

Application 43 (Cutoff for DOU)**.** Let $(\mathrm{X}^n_t)_{t\geqslant 0}$ be the DOU process with $\beta = 0$ or $\beta \geqslant 1$ and a unique invariant law π^n_β. Let's prove that for all $\mathrm{d} \in \{\mathrm{TV}, \mathrm{H}, \text{Hellinger}, \mathrm{I}, \mathrm{W}, \chi^2, \mathrm{L}^{p>1}(\pi)\}$, and $\varepsilon \in (0,1)$, the sequence exhibits cutoff.
It suffices to verify that the three previous conditions are satisfied for the different measures, and in some cases use the comparison criterion. Supposed that we already have the cutoff for TV, Hellinger, W (proved in [6], with a mixing time $t^n_{\text{mix}} \longrightarrow \infty$), the we already have the product-like condition for $\mathrm{H}, \chi^2, \mathrm{L}^p(\pi), \mathrm{I}$ with comparison to TV. It then suffices to show the monotonicity, the relaxation condition and the comparison of distances and divergences.

– (Monotonicity): for all $\mathrm{d} \in \left\{\mathrm{TV}, \mathrm{H}, \text{Hellinger}, \mathrm{I}, \mathrm{W}, \chi^2, \|.\|_{p>1}\right\}$, we have :

$$\mathrm{d}^*_n : t \geqslant 0 \longmapsto \sup_{\mu_0\in \mathrm{S}^n_0} \mathrm{d}(\mu^n_t \mid \pi^n_\beta) \text{ is non-increasing.}$$

The proof is given in [6] for $\mathrm{d} \in \left\{\mathrm{TV}, \mathrm{H}, \text{Hellinger}, \mathrm{I}, \mathrm{W}, \chi^2, \|.\|_{p>1}\right\}$ and [10] for $\mathrm{d} = \|.\|_{p>1}$.

– (Relaxation Condition): for all $t \geqslant 0$, we have the sub-exponential convergences :

$$\forall\, \mathrm{d} \in \{\mathrm{H}, \mathrm{I}, \chi^2\} \ : \quad \mathrm{d}\left(\mu^n_t \mid \pi^n_\beta\right) \leqslant e^{-2t}\mathrm{d}\left(\mu^n_0 \mid \pi^n_\beta\right).$$

and

$$\|\mu_t - \pi\|_p \leqslant C_p e^{-c_p t} \|\mu_0 - \pi\|_p .$$

The proof is given in [6] and [10].

– (Comparison of distances/divergences): for all probability measures μ and ν on the same space,

$$\begin{aligned} \|\mu-\nu\|^2_{\mathrm{TV}} &\leqslant 2\,\mathrm{H}(\nu \mid \mu) \\ \mathrm{H}(\nu \mid \mu) &\leqslant 2\chi(\nu \mid \mu) + \chi^2(\nu \mid \mu) \\ \mathrm{H}(\nu \mid P^\beta_n) &\leqslant \frac{1}{2n}\mathrm{I}(\nu \mid P^\beta_n) \\ 2\,\|\mu_t - \pi\|_{\mathrm{TV}} &\leqslant \|\mu_t - \pi\|_p \end{aligned}$$

The proof is given in [6] and Hölder.

2.2. **Spectral criterion**

The following corollary provides a sufficient condition for having a cutoff. This corollary is particularly interesting when we know an eigenfunction associated with the spectral gap.

Corollary 44 (Spectral criterion for the $\mathrm{L}^2(\pi)$ distance)**.** *Let $x_0^n \in \mathrm{S}_0^n$. If there exists $\varphi_n \in \mathrm{L}^2(\pi^n)$ such that: for a constant $c > 0$ and all $t \geqslant 0$ and $n \geqslant 1$,*

$$\lim_{n\to\infty} \frac{|\varphi_n(x_0^n)|}{\|\varphi_n\|_2} = +\infty \quad \text{and} \quad |(\mathrm{P}_t^n - \pi^n)(\varphi_n)(x_0^n)| \geqslant e^{-c\lambda_n t}|\varphi_n(x_0^n)|,$$

then the product condition is satisfied for $\mathrm{S}_0^n = \{\delta_{x_0^n}\}$.

Proof. Indeed, at $t = t_{\mathrm{mix}}$, by duality,

$$\eta \geqslant \|\mu_t^n - \pi^n\|_2 = \sup_{\|g\|_2=1} |(\mathrm{P}_t^n - \pi^n)(g)| \geqslant e^{-c\lambda_n t_{\mathrm{mix}}^n} \frac{|\varphi_n(x_n)|}{\|\varphi_n\|_2},$$

which imposes the product condition $\lim_{n\to\infty} \lambda_n t_{\mathrm{mix}}^n = +\infty$. □

Application of the spectral criterion for DOU

We will now use this theorem in the case of the DOU process. Let us begin with a spectral analysis of its generator, which, as a reminder, is given by:

$$\mathrm{G} := \frac{1}{n}(\Delta - \nabla \mathrm{E}\cdot\nabla).$$

We denote $\pi_0^n := \mathcal{N}\left(0, \frac{\sigma^2}{2\theta}\mathrm{I}_n\right)$.

Proposition 45 (Spectral analysis for the non-interactive case OU)**.** *The operator:*

$$\mathrm{G} : \mathrm{L}^2(\mathbb{R}^n, \pi_0^n) \longrightarrow \mathrm{L}^2(\mathbb{R}^n, \pi_0^n),$$

is well-defined and self-adjoint. Moreover:

- G *globally preserves the set of polynomials,*
- $\mathrm{Sp}(\mathrm{G}) = \mathbb{Z}_-$,
- *The eigenspaces $(\mathrm{F}_{-m})_{m\in\mathbb{N}}$ are finite-dimensional and each F_{-m} is generated by multivariate Hermite polynomials of degree m (that is, by tensor products of univariate Hermite polynomials).*

In particular, F_0 is the vector space of constant functions, while F_1 is the n-dimensional vector space of all linear functions.

Corollary 46 (Invariant law)**.** *It is the reversible invariant distribution given by:*

$$\pi_0^n = \mathcal{N}\left(0, \frac{\sigma^2}{2\theta}\mathrm{I}_n\right) = \mathcal{N}\left(0, \frac{\sigma^2}{2\theta}\right)^{\otimes n}$$

Definition 47 (Heat kernel)**.** It is the the density of the law $\mathcal{L}(\mathrm{X}_t^n \mid \mathrm{X}_0^n = x)$ with respect to the invariant measure π_0^n:

$$p_t(x, \cdot) := \frac{\mathrm{d}\mu_t^x}{\mathrm{d}\pi_0^n}.$$

Proposition 48 (Properties of the heat kernel)**.** *For all $x \in \mathbb{R}^n$, we have:*

- *(long-time behavior):*
$$\lim_{t\to\infty} p_t(x,\cdot) = 1.$$
- *(norm $\mathrm{L}^p(\pi_0^n)$):*
$$\|\mu_t^x - \pi_0^n\|_{\mathrm{TV}} = \|p_t(x,\cdot) - 1\|_1 \leqslant \|p_t(x,\cdot) - 1\|_p \leqslant \|p_t(x,\cdot) - 1\|_q, \quad 1 \leqslant p \leqslant q.$$

- *(case $p=2$): let* A_m *be an orthonormal basis of* $\mathrm{F}_m \subset \mathrm{L}^2(\pi_0^n)$*,*

$$\|p_t(x,\cdot)-1\|_2 = \sum_{m=1}^{\infty}\sum_{\psi\in\mathrm{A}_m} e^{-2mt}\|\psi(x)\|_2^2 \qquad \text{and} \qquad \|p_t(x,\cdot)-1\|_2 \geqslant e^{-2t}\sum_{\psi\in\mathrm{A}_1}|\psi(x)|^2.$$

Remark 49 (Case $p=2$ and cutoff)*.* Since we can estimate $\sum_{\psi\in\mathrm{A}_1}|\psi(x)|^2$ (this is the square of the norm of the projection of δ_x onto A_1), the above inequality leads to a lower bound on the χ^2 cutoff (*i.e.* L^2).

Let $\mathrm{L}^2_{\mathrm{sym}}(\pi_\beta^n)$ denote the Hilbert space of symmetric functions of n variables $(x_1,\ldots,x_n)$, defined on $\mathbb{R}^n$, and square-integrable with respect to the measure π_β^n.

Proposition 50 (Spectral analysis for the interactive case DOU with $\beta \geqslant 1$)**.** *The operator:*

$$\mathrm{G}_{\mathrm{DOU}} : \mathrm{L}^2_{sym}(\pi_\beta^n) \longrightarrow \mathrm{L}^2_{sym}(\pi_\beta^n)$$

is well-defined and self-adjoint. Moreover:

- G *globally preserves the set of symmetric polynomials,*
- *(spectrum):* $\mathrm{Sp}(\mathrm{G}) = \mathbb{Z}_-$*,*
- *(eigenspaces):* $(\mathrm{F}_{-m})_{m\in\mathbb{N}}$ *are finite-dimensional and each* F_{-m} *is generated by multivariate Hermite polynomials of degree* m*.*

In particular, F_0 *is the vector space of constant functions, while* F_1 *is the* n*-dimensional vector space of all linear functions.*

Proof. See [9]. □

Remark 51. Note that, in the present interactive case, the integrability properties of the heat kernel are not known: in particular, we do not know if $p_t(x,\cdot)$ belongs to $\mathrm{L}^p(\pi_\beta^n)$ for $t>0$, $x\in D^n$, and $p>1$.

We use the spectral criterion to demonstrate that the DOU process has a cutoff in $\mathrm{L}^2(\pi_\beta^n)$ norm and for $\mathrm{S}_0^n = \{\delta_{x_0^n}\}$. To do this, let us observe how to choose $x_0^n \in \mathrm{S}_0^n$.

Application 52 (Spectral criterion for DOU)**.** Let φ_n be the eigenvector associated with the eigenvalue -1 of G, defined by:

$$\varphi_n : x \longmapsto x_1 + \cdots + x_n.$$

Then we have:

$$\|\varphi_n\|^2_{\mathrm{L}^2(\pi_\beta^n)} := \langle\varphi_n,\varphi_n\rangle_{\mathrm{L}^2(\pi_\beta^n)} = -\langle \mathrm{G}\varphi_n,\varphi_n\rangle_{\mathrm{L}^2(\pi_\beta^n)} \overset{\text{def.}}{=} \int \|\nabla\varphi_n\|^2\,\mathrm{d}\pi_\beta^n = \int n\,\mathrm{d}\pi_\beta^n = n,$$

since $\nabla\varphi_n = (1,\ldots,1)$. Therefore, it is sufficient to choose x_0^n such that $x_0^n \to +\infty$ as $n\to\infty$.

Thus, since we have monotonicity, the spectral gap relaxation condition, and the product-like condition by the spectral criterion, the DOU process indeed has a cutoff in $\mathrm{L}^2(\pi_\beta^n)$ norm and for $\mathrm{S}_0^n = \{\delta_{x_0^n}\}$, according to Theorem 40.

2.3. **Conclusion and perpectives**

Another promising research direction is to extend the results obtained in [6] to other integrable Dyson-type diffusions, particularly the Laguerre and Jacobi beta ensembles. These studies would allow us to revisit the constraints on initial conditions using the microscopic analysis developed in random matrix theory.

Moreover, the cutoff phenomenon in McKean-Vlasov processes, with potential links to metastability, is an essential topic for future research. Discussions with my advisor and during my research visit have led to the following perspectives:

1. Explore whether the product condition is necessary for diffusions.
2. Explore the cutoff phenomenon in the remaining integrable Dyson-type dynamics associated with the Laguerre and Jacobi beta ensembles. Revisit the constraints on initial conditions using the microscopic analysis developed in beta ensembles within random matrix theory.
3. Study the cutoff phenomenon for general ergodic diffusions in non-compact spaces by connecting the works of [6] and [10, 11, 24, 26].
4. Investigate the cutoff for the dynamics of systems composed of n i.i.d. ergodic particles.
5. Explore the cutoff for processes with polynomial convergence towards equilibrium.
6. Study the cutoff for hypoelliptic dynamics, particularly Gaussian chains of oscillators.
7. Analyze the link between the cutoff of McKean-Vlasov processes and metastability.

3. **Varentropic method**

In this section, we reproduce the proof from [23] in the case of a curved ergodic Markov process with values in a general state space. We highlight that the obtained condition is a control of the semi-group (P_t^n). In particular, we will see applications of this theorem where such control is possible: the case of finite Markov chains (the case from [23] which is the basis of our study), the case of the O.U. process, and finally the case of a compact state space with Brownian motion on the sphere. The objective is to highlight the discrete and continuous cases, as well as the types of conditions on the initial conditions.

3.1. **Main result**

Let S_0^n be the set of pointwise initial conditions, *i.e.* a subset of $\mathbb{R}^n$. We are interested in the cutoff phenomenon for the total variation distance:

$$t_{\mathrm{mix}}^n(\varepsilon) := t_{\mathrm{mix}}^{n,\mathrm{TV}}(\varepsilon, \mathrm{S}_0^n) := \inf\{t \geqslant 0 : \sup_{\mathrm{S}_0^n} \|\mathrm{Loi}(\mathrm{X}_t^n) - \pi^n\|_{\mathrm{TV}} \leqslant \varepsilon\}.$$

Theorem 53 (Cutoff in the general case)**.** *Let $(\mathrm{P}_t^n)_{t\geqslant 0}$ be an ergodic Markov semigroup on S_0^n, a unique invariant law π^n and its spectral gap λ_n. Suppose that for all $\varepsilon \in (0,1)$:*
(i) (curvature condition) : $(\mathrm{P}_t^n)_{t\geqslant 0}$ has a non-negative curvature : $\mathrm{CD}(0,\infty)$,
(ii) (product condition) : $1 \ll \lambda_n t_{mix}^n(\varepsilon)$,
(iii) (Fisher condition) :

$$\mathrm{I}_n^*(t_{mix}^n(\varepsilon)) \ll \lambda_n^2 t_{mix}^n(\varepsilon) \qquad \text{where} \qquad \mathrm{I}_n^*(t) = \sup_{x\in\mathrm{S}_0^n} \mathrm{Fisher}(\delta_x \mathrm{P}_t^n \mid \pi^n).$$

Then, for all $\varepsilon \in (0,1)$, the sequence exhibits cutoff for TV.

Remark 54. This slightly more general formulation is very enlightening because it specifies what happens in a finite space, and it gives us clues for the infinite case on what needs to be done in the unbounded case:
– For finite Markov chains, the logarithm of the kernel is a Lipschitz function, so it is bounded on a finite set. The interpretation of this bound is discussed in [23].
– For O.U., we have an explicit formula for the Fisher information in the case of point initial conditions. However, this is not bounded, but by taking controlled initial conditions, it becomes bounded.

Remark 55 (Type of curvature). It should be noted that in the proof below, we use tools related to Bakry-Émery curvature and not Ollivier-Ricci curvature.

3.2. Proof

We denote μ_t^x for $\delta_x \mathrm{P}_t^n$.

Setup

In this section, we consider :

$$\mathrm{d}(\mu \mid \nu) \;:=\; \|\mu - \nu\|_{\mathrm{TV}} \qquad \text{and} \qquad \mathrm{d}_n^* \;:=\; \sup_{x \in \mathrm{S}_0^n} \mathrm{d}(\mu_t^x \mid \pi).$$

Recall that the relative entropy (Kullback-Leibler Divergence) is defined by

$$\mathrm{H}(\mu \mid \nu) \;:=\; \mathrm{Kullback}(\mu \mid \nu) \;=\; \begin{cases} \displaystyle\int \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right) \mathrm{d}\mu \in [0, +\infty] & \text{si } \mu \ll \nu, \\ \infty & \text{sinon,} \end{cases}$$

and the Fisher Information by :

$$\mathrm{I}(\mu \mid \nu) \;:=\; \mathrm{Fisher}(\mu \mid \nu) \;=\; \int \left|\nabla \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right)\right|^2 \mathrm{d}\mu.$$

In [23], to exhibit a cutoff, the author uses the varentropy defined by :

Definition 56 (Varentropy)**.**

$$\mathrm{V}(\mu \mid \nu) \;:=\; \mathrm{Varentropy}(\mu \mid \nu) \;=\; \begin{cases} \displaystyle\int \left(\log \frac{\mathrm{d}\mu}{\mathrm{d}\nu} - \mathrm{H}(\mu \mid \nu)\right)^2 \mathrm{d}\mu \in [0, +\infty] & \text{si } \mu \ll \nu, \\ \infty & \text{sinon.} \end{cases}$$

We then define the following:

$$\mathrm{H}_n^*(t) \;:=\; \sup_{x \in \mathrm{S}_0^n} \mathrm{H}(\mu_t^x \mid \pi), \qquad \mathrm{V}_n^*(t) \;:=\; \sup_{x \in \mathrm{S}_0^n} \mathrm{V}(\mu_t^x \mid \pi) \qquad \text{and} \qquad \mathrm{I}_n^*(t) \;:=\; \sup_{x \in \mathrm{S}_0^n} \mathrm{I}(\mu_t^x \mid \pi).$$

Outline of the proof

To show that there exists a cutoff, it is equivalent to show that:

$$t_{\mathrm{mix}}^n(\varepsilon) - t_{\mathrm{mix}}^n(1-\varepsilon) \;=\; \underset{n \to \infty}{o}\left(t_{\mathrm{mix}}^n(1-\varepsilon)\right)$$

To do so, we proceed as follows. On the one hand, we start by establishing an entropic upper-bound by linking the mixing time to the entropy (Kullback), using the spectral gap and considering a sufficiently large time:

$$t_{\mathrm{mix}}^n(\varepsilon) \;\leqslant\; t + \frac{1}{\lambda_n \varepsilon}\left(1 + \varepsilon + 2\mathrm{H}_n^*(t)\right).$$

Next, we exploit an entropic lower-bound and relate the entropy to the varentropy by viewing entropy as the expectation of a log and using the Bienaymé-Tchebychev inequality:

$$\|\mu - \pi\|_{\mathrm{TV}} \;\leqslant\; 1 - \varepsilon \quad \Longrightarrow \quad \mathrm{H}_n^*(t) \;\leqslant\; \frac{1 + \sqrt{\mathrm{V}_n^*(t)}}{\varepsilon}.$$

From this, we obtain the first following inequality:

$$t_{\mathrm{mix}}^n(\varepsilon) - t_{\mathrm{mix}}^n(1-\varepsilon) \;\leqslant\; \frac{2}{\varepsilon^2 \lambda_n}\left(2 + \sqrt{\mathrm{V}_n^*\left(t_{\mathrm{mix}}^n(1-\varepsilon)\right)}\right).$$

On the other hand, we exploit the positive curvature through the local Poincaré inequality:

$$\mathrm{P}_t(f^2) - (\mathrm{P}_t f)^2 \;\leqslant\; 2t\mathrm{P}_t\Gamma(f).$$

By applying this inequality to the logarithm of the kernel ($f = \log \frac{\mathrm{d}\mu_t^x}{\mathrm{d}\pi}$), we obtain the link between varentropy and Fisher information as follows:

$$\mathrm{V}_n^*(t) \;\leqslant\; 2t \times \mathrm{I}_n^*(t).$$

In proofs and lemmas, we drop the n sub/superscripts for simplicity.

The entropic concentration phenomenon

Proposition 57 (Entropic concentration implies cutoff)**.** *For all* $\varepsilon \in (0,1)$,

$$t^n_{mix}(\varepsilon) - t^n_{mix}(1-\varepsilon) \leqslant \frac{2}{\varepsilon^2 \lambda_n}\left(2 + \sqrt{\mathrm{V}^*_n\left(t^n_{mix}(1-\varepsilon)\right)}\right).$$

Lemma 58. *Let* ν *be a probability measure such that* $\nu \ll \pi$*. Then :*

$$\|\nu \mathrm{P}_t - \pi\|_{\mathrm{TV}} \leqslant \frac{1}{2} e^{-\lambda t} \left\|\frac{\mathrm{d}\nu}{\mathrm{d}\pi}\right\|_\infty,$$

where λ *is the spectral gap.*

Proof. By the definition of the spectral gap λ, we have, for all $s \geqslant 0$ and $f \in \mathrm{L}^2(\pi)$:

$$\|\mathrm{P}_s(f) - \pi(f)\|^2_{\mathrm{L}^2(\pi)} \leqslant e^{-2\lambda t} \|f\|^2_{\mathrm{L}^2(\pi)},$$

In particular, since $f = \frac{\mathrm{d}\nu}{\mathrm{d}\pi}$ exists, then, denoting $\mathrm{d}(\nu\mathrm{P}_s) =: f_s \mathrm{d}\pi$:

$$\|f_s - 1\|^2_{\mathrm{L}^2(\pi)} \leqslant e^{-2\lambda t}\|f-1\|^2_{\mathrm{L}^2(\pi)}$$

and

$$\|\nu\mathrm{P}_s - \pi\|_{\mathrm{TV}} = \frac{1}{2}\|f_s - 1\|_{\mathrm{L}^1(\pi)} \leqslant \frac{1}{2}\|f_s - 1\|_{\mathrm{L}^2(\pi)} \leqslant \frac{1}{2}e^{-\lambda t}\|f-1\|_{\mathrm{L}^2(\nu)} \leqslant \frac{1}{2}e^{-\lambda t}\|f\|_\infty.$$

□

Lemma 59 (Entropic upper-bound)**.** *For all* $t \geqslant 0$ *and all* $\varepsilon \in (0,1)$*:*

$$t^n_{mix}(\varepsilon) \leqslant t + \frac{1}{\lambda\varepsilon}\left(1 + \varepsilon + 2\mathrm{H}^*_n(t)\right).$$

Outline of the proof. We recall that by definition of the mixing time: $t^n_{\mathrm{mix}}(\varepsilon) \leqslant t$ if and only if $\mathrm{d}^*_n(t) \leqslant \varepsilon$. In particular:

$$\forall s,t \geqslant 0 : \quad \mathrm{d}^*_n(t+s) \leqslant \varepsilon \qquad \Longleftrightarrow \qquad t^n_{\mathrm{mix}}(\varepsilon) \leqslant t+s.$$

(1) To show that $\mathrm{d}^*_n(t+s) \leqslant \varepsilon$, it suffices, by definition of the supremum, to prove that:

$$\forall \mu \in \mathcal{P}(\mathrm{X}) : \quad \|\mu\mathrm{P}_s - \pi\|_{\mathrm{TV}} \leqslant \varepsilon,$$

since this holds for $\mu := \delta_{x_0}\mathrm{P}_t$ and

$$\delta_{x_0}\mathrm{P}_t\mathrm{P}_s = \delta_{x_0}\mathrm{P}_{t+s}.$$

(2) We conclude by taking: $s := s(t,x) := \frac{1}{\lambda\varepsilon}\left(1+\varepsilon+2\mathrm{H}(\mu^x_t \mid \pi)\right)$, because:

$$\|\delta_{x_0}\mathrm{P}_{t+s} - \pi\|_{\mathrm{TV}} \leqslant \varepsilon \qquad \Longleftrightarrow \qquad t^n_{\mathrm{mix}}(\varepsilon) \leqslant t + s(t,x_0) \leqslant t + s^*(t),$$

where $s_*(t) := \sup_{x_0} s(t,x_0) = \frac{1}{\lambda\varepsilon}\left(1+\varepsilon+2\mathrm{H}^*_n(t)\right)$.

□

Proof. Let μ be such that $\mathrm{H}(\mu \mid \pi) > 0$. For all $\varepsilon \in (0,1)$, define the event

$$\mathrm{A}_\varepsilon := \left\{x : \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\pi}(x)\right) < 1 + \frac{2\mathrm{H}(\mu \mid \pi)}{\varepsilon}\right\}.$$

(1) By definition of A_ε^c and since $\mu(\mathrm{A}_\varepsilon^c) = \int_{\mathrm{A}_\varepsilon^c} \mathrm{d}\mu$, we have :

$$\left(1 + \frac{2\mathrm{H}(\mu \mid \pi)}{\varepsilon}\right) \mu(\mathrm{A}_\varepsilon^c) \leqslant \int_{\mathrm{A}_\varepsilon^c} \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\pi}\right) \mathrm{d}\mu = \mathrm{H}(\mu \mid \pi) - \int_{\mathrm{A}_\varepsilon} \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\pi}\right) \mathrm{d}\mu.$$

Since $-\log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\pi}\right) = \log\left(\frac{\mathrm{d}\pi}{\mathrm{d}\mu}\right)$ and $\log(u) \leqslant u - 1$, we have :

$$\left(1 + \frac{2\mathrm{H}(\mu \mid \pi)}{\varepsilon}\right) \mu(\mathrm{A}_\varepsilon^c) \leqslant \mathrm{H}(\mu \mid \pi) + \pi(\mathrm{A}_\varepsilon) - \mu(\mathrm{A}_\varepsilon) \leqslant \mathrm{H}(\mu \mid \pi) + \mu(\mathrm{A}_\varepsilon^c).$$

After simplification, this gives

$$\mu(\mathrm{A}_\varepsilon^c) \leqslant \frac{\varepsilon}{2}.$$

In particular, $\mu(\mathrm{A}_\varepsilon) \geqslant \frac{1}{2}$.

(2) The conditional probability measure $\mu^\varepsilon := \mu(\cdot \mid \mathrm{A}_\varepsilon) := \frac{\mu(\cdot \cap \mathrm{A}_\varepsilon)}{\mu(\mathrm{A}_\varepsilon)}$ satisfies

$$\frac{\mathrm{d}\mu^\varepsilon}{\mathrm{d}\pi} = \frac{\mathbb{1}_{\mathrm{A}_\varepsilon}}{\mu(\mathrm{A}_\varepsilon)} \frac{\mathrm{d}\mu}{\mathrm{d}\pi} \implies \left\|\frac{\mathrm{d}\mu^\varepsilon}{\mathrm{d}\pi}\right\|_\infty = \frac{1}{\mu(\mathrm{A}_\varepsilon)} \left\|\mathbb{1}_{\mathrm{A}_\varepsilon} \frac{\mathrm{d}\mu}{\mathrm{d}\pi}\right\|_\infty \leqslant e^{2+\frac{2\mathrm{H}(\mu \mid \pi)}{\varepsilon}},$$

where, for the last inequality, we used that $\mu(\mathrm{A}_\varepsilon) \geqslant \frac{1}{2} \geqslant \frac{1}{e}$ and :

$$\left\|\mathbb{1}_{\mathrm{A}_\varepsilon} \frac{\mathrm{d}\mu}{\mathrm{d}\pi}\right\|_\infty = \sup_{x \in \mathrm{A}_\varepsilon} \left|\frac{\mathrm{d}\mu}{\mathrm{d}\pi}(x)\right| \overset{\text{déf.}}{\leqslant} \exp\left(1 + \frac{2\mathrm{H}(\mu \mid \pi)}{\varepsilon}\right).$$

(3) We denote $\mu_s^\varepsilon := \mu^\varepsilon \mathrm{P}_s \ll \pi$. Consequently, for all $s \geqslant 0$, by using step (2) and Lemma 58:

$$\|\mu_s^\varepsilon - \pi\|_{\mathrm{TV}} \leqslant \frac{1}{2} e^{-\lambda s + 2 + \frac{2\mathrm{H}(\mu \mid \pi)}{\varepsilon}}.$$

Hence, for all $s \geqslant s_* := \frac{1}{\lambda}\left(1 + \frac{2\mathrm{H}(\mu \mid \pi) + 1}{\varepsilon}\right)$, we get

$$\|\mu_s^\varepsilon - \pi\|_{\mathrm{TV}} \leqslant \frac{1}{2} e^{1 - \frac{1}{\varepsilon}} \leqslant \frac{\varepsilon}{2}.$$

(4) On the other hand, we simply have by contractivity of Markov kernels

$$\|\mu \mathrm{P}_s - \mu_s^\varepsilon\|_{\mathrm{TV}} = \|\mu \mathrm{P}_s - \mu^\varepsilon \mathrm{P}_s\|_{\mathrm{TV}} \leqslant \|\mu - \mu^\varepsilon\|_{\mathrm{TV}} = \mu(\mathrm{A}_\varepsilon^c) \leqslant \frac{\varepsilon}{2}.$$

(5) Thus, by the triangle inequality, we obtain, for all $s \geqslant s_*$,

$$\|\mu \mathrm{P}_s - \pi\|_{\mathrm{TV}} \leqslant \|\mu \mathrm{P}_s - \mu_s^\varepsilon\|_{\mathrm{TV}} + \|\mu_s^\varepsilon - \pi\|_{\mathrm{TV}} \leqslant \varepsilon.$$

□

Lemma 60 (Entropic lower-bound)**.** *For any* $\mu \in \mathcal{P}(\mathrm{X})$ *and any* $\varepsilon \in (0,1)$ *:*

$$\mathrm{d}_{\mathrm{TV}}(\mu, \pi) \leqslant 1 - \varepsilon \qquad \implies \qquad \mathrm{H}_n^*(t) \leqslant \frac{1 + \sqrt{\mathrm{V}_n^*(t)}}{\varepsilon}.$$

Proof. Proved in [23]. □

Proof of proposition 57. Let $\varepsilon \in (0,1)$ and $t := t_{\mathrm{mix}}^n(1-\varepsilon)$. The first lemma gives

$$t_{\mathrm{mix}}^n(\varepsilon) \leqslant t + \frac{1}{\lambda\varepsilon}\left(1 + \varepsilon + 2\mathrm{H}_n^*(t)\right).$$

But the second lemma with $\mu = \mu_t^x$ gives

$$\mathrm{H}_n^*(t) \leqslant \frac{1 + \sqrt{\mathrm{V}_n^*(t)}}{\varepsilon}.$$

Combining the two inequalities and using $\varepsilon \leqslant 1$ provides the desired result. □

Non-negative curvature implies entropic concentration

Lemma 61 (Local concentration inequality)**.** *For all $t \geqslant 0$:*

$$\mathrm{P}_t^n(f^2) - (\mathrm{P}_t^n f)^2 \;\leqslant\; \frac{1-e^{-2\rho t}}{\rho}\mathrm{P}_t^n\Gamma(f) \;\leqslant\; 2t\mathrm{P}_t^n\Gamma(f).$$

Proof. Since $\mathrm{CD}(\rho,\infty)$ is satisfied by $(\mathrm{P}_t)_{t\geqslant 0}$, then :

$$\mathrm{P}_t(f^2) - (\mathrm{P}_t f)^2 \;=\; 2\int_0^t \mathrm{P}_{t-s}\Gamma(\mathrm{P}_s f)\,\mathrm{d}s \qquad \text{and} \qquad \Gamma(\mathrm{P}_s f) \;\leqslant\; e^{-2\rho s}\mathrm{P}_s\Gamma(f).$$

For the second inequality we use that $1-e^{-x} \leqslant x$. □

Proof of theorem 53. Applying this lemma to $f = \log\frac{\mathrm{d}\mu_t}{\mathrm{d}\pi}$, we obtain:

$$\mathrm{V}(\mu_t \mid \pi) \;\leqslant\; 2t \times \int \left|\nabla \log\left(\frac{\mathrm{d}\mu_t}{\mathrm{d}\pi}\right)\right|^2 \mathrm{d}\mu_t \;=\; 2t \times \mathrm{I}(\mu_t \mid \pi),$$

since:

$$\mathrm{P}_t\left(\log\frac{\mathrm{d}\mu_t}{\mathrm{d}\pi}\right) \;=\; \mathrm{H}(\mu_t|\pi) \qquad \text{and} \qquad \Gamma\left(\log\frac{\mathrm{d}\mu_t}{\mathrm{d}\pi}\right) \;=\; \left|\nabla\log\left(\frac{\mathrm{d}\mu_t}{\mathrm{d}\pi}\right)\right|^2.$$

Now taking the maximum over S_0^n :

$$\mathrm{V}_n^*(t) \;\leqslant\; 2t\mathrm{I}_n^*(t).$$

Then :

$$\frac{\sqrt{\mathrm{V}_n^*\left(t_{\mathrm{mix}}^n(1-\varepsilon)\right)}}{\lambda_n t_{\mathrm{mix}}^n(1-\varepsilon)} \;\leqslant\; \frac{\sqrt{2\mathrm{I}_n^*(t_{\mathrm{mix}}^n(1-\varepsilon))}}{\lambda_n\sqrt{t_{\mathrm{mix}}^n(1-\varepsilon)}}.$$

However, by Proposition **??**, we have:

$$\begin{aligned}\frac{t_{\mathrm{mix}}^n(\varepsilon) - t_{\mathrm{mix}}^n(1-\varepsilon)}{t_{\mathrm{mix}}^n(1-\varepsilon)} \;&\leqslant\; \frac{2}{\varepsilon^2}\left(\frac{2}{\lambda_n t_{\mathrm{mix}}^n(1-\varepsilon)} + \frac{\sqrt{\mathrm{V}_n^*\left(t_{\mathrm{mix}}^n(1-\varepsilon)\right)}}{\lambda_n t_{\mathrm{mix}}^n(1-\varepsilon)}\right)\\ &\leqslant\; \frac{2}{\varepsilon^2}\left(\frac{2}{\lambda_n t_{\mathrm{mix}}^n(1-\varepsilon)} + \sqrt{2}\frac{\sqrt{\mathrm{I}_n^*(t_{\mathrm{mix}}^n(1-\varepsilon))}}{\lambda_n\sqrt{t_{\mathrm{mix}}^n(1-\varepsilon)}}\right) \;\xrightarrow[n\to\infty]{}\; 0.\end{aligned}$$

□

3.3. **Applications for finite Markov chains**

In all that follows, the objective is to control the Fisher information. In this subsection, we consider a finite state space S_0^n. The full proof of the following corollary is given in [23].

Corollary 62 (Cutoff for finite Markov chains, SALEZ)**.** *Consider a sequence of irreducible transitions matrices with symmetric support and non-negative Bakry-Émery curvature. Suppose that for every fixed $\varepsilon \in (0,1)$, we have*

$$t_{mix}^n(\varepsilon) \;\gg\; \left(t_{rel}^n \log\Delta^n\right)^2, \tag{$\star$}$$

where $\Delta^n \;=\; \max\left\{\mathrm{P}^n(x,y)^{-1},\; x\sim y\right\}$ is the sparsity parameter. Then, the sequence exhibits cutoff. More precisely, for every $\varepsilon \in (0,\frac{1}{2})$, we have

$$t_{mix}^n(\varepsilon) - t_{mix}^n(1-\varepsilon) \;\lesssim\; \sqrt{t_{mix}^n(1/4)}\,t_{rel}^n\log\Delta^n.$$

Note that the "product-like" condition $(\star)$ implies the product condition (ii). It remains to control the Fisher information.

Lemma 63 (Control of the Fisher Information for finite Markov chains)**.**

$$\mathrm{I}_n^*(t^n_{mix}(\varepsilon)) \;\lesssim\; \log^2 \Delta^n.$$

Outline of the proof.

(1) Since our Markov chain is finite, we have the trivial bound $\Gamma(f) \leqslant \|f\|^2_{\mathrm{Lip}}$, and by the stochasticity of the operator P_s :

$$\mathrm{P}_s\Gamma(f) \leqslant \|f\|^2_{\mathrm{Lip}} \quad\Longrightarrow\quad \mathrm{I}_n^*(t) \;=\; \max_{x\in \mathrm{S}_0^n} \mathrm{P}_s\Gamma\left(\log\frac{\mathrm{P}_t(x,\cdot)}{\pi}\right) \;\leqslant\; \max_{x\in \mathrm{S}_0^n}\left\|\log\frac{\mathrm{P}_t(x,\cdot)}{\pi}\right\|^2_{\mathrm{Lip}}.$$

(2) Hence, one shows that if P has symmetric support, then :

$$\forall\, t \;\geqslant\; \frac{\mathrm{diam}(\mathrm{S}_0^n)}{4} \;:\quad \max_{x\in \mathrm{S}_0^n}\left\|\log\frac{\mathrm{P}_t(x,\cdot)}{\pi}\right\|^2_{\mathrm{Lip}} \;\leqslant\; 3(1+\log\Delta).$$

(3) Thus, we show that for all $\varepsilon \in (0,1)$:

$$\mathrm{diam}(\mathrm{S}_0^n) \;\leqslant\; 2t^n_{\mathrm{mix}}(\varepsilon) + \sqrt{\frac{8t^n_{\mathrm{mix}}(\varepsilon)}{1-\varepsilon}} + \sqrt{\frac{8t_{\mathrm{rel}}}{1-\varepsilon}}.$$

(4) Since $\sqrt{t_{\mathrm{rel}}} \ll t^n_{\mathrm{mix}}(\varepsilon)$, by step (3) we have :

$$\mathrm{diam}(\mathrm{S}_0^n) \;\leqslant\; (2+o(1))t^n_{\mathrm{mix}}(\varepsilon) \qquad \text{and} \qquad t^n_{\mathrm{mix}}(\varepsilon) \;\geqslant\; \frac{\mathrm{diam}(\mathrm{S}_0^n)}{4}.$$

(5) Applying step (2) and to remark that since $\Delta \;\geqslant\; 2$, we have :

$$\mathrm{I}_n^*(t^n_{\mathrm{mix}}(\varepsilon)) \;\leqslant\; 9(1+\log\Delta)^2 \;\lesssim\; \log^2\Delta \;\ll\; \lambda_n^2 t^n_{\mathrm{mix}}(\varepsilon).$$

□

3.4. **The Ornstein-Uhlenbeck case is not concerned**

In this section, we will examine whether we can apply the previous results to OU processes. Recall the properties of OU process: for $\sigma_n^2 \,>\, 0$ and $\theta_n \in \mathbb{R}$,

$$\mathrm{dX}_t^n \;=\; \sigma_n\,\mathrm{dB}_t^n - \theta_n \mathrm{X}_t^n\,\mathrm{d}t, \qquad \mathrm{X}_0^n \sim \mu_0^n(\theta_n,\sigma_n).$$

– (Mehler formula) :

$$\mu_t^x \;:=\; \mathrm{Law}(\mathrm{X}_t^n \mid \mathrm{X}_0^n = x) \;=\; \mathcal{N}\left(xe^{-\theta_n t}, \frac{\sigma_n^2}{2}\frac{1-e^{-2\theta_n t}}{\theta_n}\mathrm{I}_n\right) \;=\; \bigotimes_{i=1}^n \mathcal{N}\left(x_i e^{-\theta_n t}, \frac{\sigma_n^2}{2}\frac{1-e^{-2\theta_n t}}{\theta_n}\right).$$

– (Invariant law) :

$$\pi^n \;:=\; \mathcal{N}\left(0,\, \frac{\sigma_n^2}{2\theta_n}\mathrm{I}_n\right) \;=\; \mathcal{N}\left(0,\, \frac{\sigma_n^2}{2\theta_n}\right)^{\otimes n}.$$

– (Spectral gap) : $\lambda_n \;=\; \theta_n$.
– (Curvature) : $\rho \;=\; 1$.

Lemma 64 (Entropy and Fisher Information for OU)**.**

- *(Entropy) :*

$$\mathrm{H}(\mu_t^x \mid \pi) \;=\; -\frac{n}{2}\log(1-e^{-2\theta_n t}) + e^{-2\theta_n t}\left(\frac{\theta_n |x|^2}{\sigma_n^2} - \frac{n}{2}\right).$$

- *(Fisher Information) :*

$$\mathrm{I}(\mu_t^x \mid \pi) \;=\; \frac{ne^{-4\theta_n t}}{\sigma_n^2(1-e^{-2\theta_n t})} + \frac{e^{-2\theta_n t}|x|^2}{\sigma_n^4}.$$

Proof. Hence, we have:

$$\forall\, y \in \mathbb{R}^n \;:\quad \log\left(\frac{\mathrm{d}\mu_t^x}{\mathrm{d}\pi}\right)(y) \;=\; -\frac{n}{2}\log(1-e^{-2\theta_n t}) + \left(-\frac{|y-xe^{-\theta_n t}|^2}{2(1-e^{-2\theta_n t})} + \frac{|y|^2}{2}\right)\frac{2\theta_n}{\sigma_n^2}.$$

Integrating with respect to $\mu_t^x(\mathrm{d}y)$, we have:

$$\mathrm{H}(\mu_t^x \mid \pi) \;=\; -\frac{n}{2}\log(1-e^{-2\theta_n t}) + \left(-\frac{\mathbb{V}(\mu_t^x)}{2(1-e^{-2\theta_n t})} + \frac{m_2(\mu_t^x)}{2}\right)\frac{2\theta_n}{\sigma_n^2},$$

where: $\mathbb{V}(\mu_t^x) \;=\; n(1-e^{-2\theta_n t})\frac{\sigma_n^2}{2\theta_n}$, $m_2(\mu_t^x) \;:=\; \mathbb{V}(\mu_t^x) + m_1(\mu_t^x)^2$ and $m_1(\mu_t^x) \;:=\; xe^{-\theta_n t}$. Then:

$$\frac{\sigma_n^2}{2\theta_n}\nabla\log\left(\frac{\mathrm{d}\mu_t^x}{\mathrm{d}\pi}\right)(y) \;=\; -\frac{y-xe^{-\theta_n t}}{1-e^{-2\theta_n t}} + y \;=\; -\frac{e^{-2\theta_n t}}{1-e^{-2\theta_n t}}y + \frac{e^{-\theta_n t}}{1-e^{-2\theta_n t}}x.$$

Thus, squaring and integrating with respect to $\mu_t^x(\mathrm{d}y)$, we have:

$$\left(\frac{\sigma_n^2}{2\theta_n}\right)^2\mathrm{I}(\mu_t^x \mid \pi) \;=\; \left(\frac{e^{-2\theta_n t}}{1-e^{-2\theta_n t}}\right)^2 m_2(\mu_t^x) - \frac{2e^{-3\theta_n t}}{(1-e^{-2\theta_n t})^2}x\cdot m_1(\mu_t^x) + \left(\frac{e^{-\theta_n t}}{1-e^{-2\theta_n t}}\right)^2|x|^2.$$

Hence we get the result. □

Lemma 65 (Mixing time for OU)**.** *Let (X_t^n) be the OU process and invariant law π. Suppose that (a_n) is a real sequence satisfying $\inf_n a_n > 0$. Then, for all $\varepsilon \in (0,1)$, we have that the sequence exhibits cutoff for TV and the mixing time is given by:*

$$t_{mix}^n(\{x\}) \;=\; \frac{1}{2\theta_n}\log\left(\frac{\theta_n}{2\sigma_n^2}|x|^2\right) \qquad\text{and}\qquad t_{mix}^n([-a_n,a_n]) \;=\; \frac{1}{2\theta_n}\log\left(\frac{n\theta_n}{2\sigma_n^2}a_n^2\right).$$

The case where $\theta_n = 1$ and $\sigma_n^2 = \frac{2}{n}$ is given by:

$$t_{mix}^n(\{x\}) \;=\; \log\left(\frac{\sqrt{n}\,|x|}{2}\right) \qquad\text{and}\qquad t_{mix}^n([-a_n,a_n]) \;=\; \log\left(\frac{na_n}{2}\right).$$

$$t_{mix}^n(\{x\}) \;=\; \log\left(\sqrt{n}\,|x|\right) \qquad\text{and}\qquad t_{mix}^n([-a_n,a_n]) \;=\; \log\left(na_n\right).$$

Proof. We adapt the work done in Section 3 of [6] to calculate the Hellinger distance:

$$\mathrm{Hellinger}^2(\mu_t^x,\pi^n) \;=\; 1-\exp\left(-\frac{\theta_n}{2\sigma_n^2}\frac{|x|^2e^{-2\theta_n t}}{2-e^{-2\theta_n t}} + \frac{n}{4}\log\left(4\frac{1-e^{-2\theta_n t}}{2-e^{-2\theta_n t}}\right)\right).$$

□

Fisher Information at mixing time is too broad

Now that we have the mixing time, we will see that the inequality involving the Fisher information is far too broad. Subsequently, we will see that, in fact, we can stop before introducing the varentropy.

Lemma 66 (Fisher Information at mixing time for OU)**.**

$$\mathrm{I}_n^*(t_{mix}^n) \;=\; \frac{4n\sigma_n^2}{\theta_n\,|x|^2\left(\theta_n\,|x|^2 - 2\sigma_n^2\right)} + \frac{2}{\theta_n\sigma_n^2}.$$

The case where $\theta_n = 1$ and $\sigma_n^2 = \frac{2}{n}$ is given by:

$$\mathrm{I}_n^*(t_{mix}^n) \;=\; \frac{8}{|x|^2\left(|x|^2-\frac{4}{n}\right)} + n.$$

Proof. Straightforward. □

Let's see that Fisher condition is not verified for any OU process. In fact, even in the case where $\theta_n = 1$ and $\sigma_n^2 = \frac{2}{n}$, we obtain the following very strong condition on the initial conditions:

$$\mathrm{I}_n^*(t_{\mathrm{mix}}^n) \ll t_{\mathrm{mix}}^n \qquad \Longleftrightarrow \qquad n \ll \log |x|.$$

This implies that this criterion is not very relevant for one of the simplest diffusions, as it is known to exhibit a cutoff for any given initial conditions. In the following subsection, we will identify where this condition fails.

Refinement of the cutoff criterion using entropy

Recall that we have:

$$t_{\mathrm{mix}}^n(\varepsilon) - t_{\mathrm{mix}}^n(1-\varepsilon) \leqslant \frac{2}{\varepsilon \lambda_n} \left(1 + \mathrm{H}_n^* \left(t_{\mathrm{mix}}^n(1-\varepsilon)\right)\right).$$

It then suffies to show the product-like condition and that:

$$\mathrm{H}_n^* \left(t_{\mathrm{mix}}^n(1-\varepsilon)\right) \ll \lambda_n t_{\mathrm{mix}}^n(1-\varepsilon).$$

Lemma 67 (Entropy at mixing time for OU)**.** *Suppose that the initial conditions satisfy* $|x|^2 \geqslant \frac{2\sigma_n^2}{\theta_n}$. *Then we have that:*

$$\mathrm{H}_n^*(t_{mix}^n) = -\frac{n}{2} \log\left(1 - \frac{2\sigma_n^2}{\theta_n |x|^2}\right) + 2 - \frac{n\sigma_n^2}{\theta_n |x|^2}.$$

The case where $\theta_n = 1$ *and* $\sigma_n^2 = \frac{2}{n}$ *is given by:*

$$\mathrm{H}_n^*(t_{mix}^n) = -\frac{n}{2} \log\left(1 - \frac{4}{n |x|^2}\right) + 2 - \frac{2}{|x|^2}.$$

Then, since $|x|^2 \geqslant \frac{2\sigma_n^2}{\theta_n}$, we have that:

$$\mathrm{H}_n^*(t_{\mathrm{mix}}^n) \ll \lambda_n^2 t_{\mathrm{mix}}^n.$$

Thus, one might wonder if Chebyshev's inequality is too loose to provide a sufficient condition for a cutoff.

3.5. **Conclusion and perspectives**

In this study, we observed that the approach by Salez based on varentropy, although effective for certain types of processes, has limitations when applied to diffusions. This method requires very specific initial conditions, making it less applicable in more general situations. A deeper analysis shows that in many cases, stopping at entropy alone, without needing to control the variance of the logarithm, would suffice to ensure a cutoff.

Nevertheless, Salez's work provides a solid foundation to search for a similar but more general criterion, by leveraging the various functional inequalities available. The following perspectives arise directly from this reflection:

1. Explore the cutoff phenomenon for general ergodic diffusions in compact spaces by following the varentropy method developed in [23]. For example: Brownian motion on the torus $\mathbb{T}^n$ or on the sphere $\mathbb{S}^n$.
2. Investigate the control of $\nabla \log p_t(x, y)$ and its connection with Harnack and Li–Yau estimates.

References

[1] D. Aldous and P. Diaconis, *Shuffling cards and stopping times*, Am. Math. Mon., 93:333–348, 1986.

[2] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer, *Sur les inégalités de Sobolev logarithmiques*, volume 10, Société Mathématique de France, Paris, 2000.

[3] D. Bakry, R. Gill, and S. Molchanov, *Lectures on probability theory. Ecole d'Eté de Probabilités de Saint-Flour XXII-1992. Summer School, 9th-25th July, 1992, Saint-Flour, France*, vol. 1581, 1994.

[4] D. Bakry, *Remarques sur les semigroupes de Jacobi*, Séminaire de Probabilités XXX, 236:23–39, 1996. Hommage à P. A. Meyer.

[5] D. Bakry, I. Gentil, and M. Ledoux, *Analysis and geometry of Markov diffusion operators*, volume 348 of Grundlehren der mathematischen Wissenschaften, Springer, Cham, 2014.

[6] J. Boursier, D. Chafaï, and C. Labbé, *Universal cutoff for Dyson Ornstein Uhlenbeck process*, *Probab. Theory Relat. Fields*, vol. 185, no. 1-2, pp. 449-512, Sep. 2022, doi: 10.1007/s00440-022-01158-5.

[7] D. Chafaï, *Entropies, convexity, and functional inequalities: on Φ-entropies and Φ-Sobolev inequalities*, *J. Math. Kyoto Univ.*, vol. 44(2), pp. 325-363, 2004.

[8] D. Chafaï, *Binomial-Poisson entropic inequalities and the $M/M/\infty$ queue*, *ESAIM Probab. Stat.*, vol. 10, pp. 317-339, 2006.

[9] D. Chafaï and J. Lehec, *On Poincaré and Logarithmic Sobolev Inequalities for a Class of Singular Gibbs Measures*, in *Geometric Aspects of Functional Analysis*, Springer International Publishing, 2020, pp. 219–246.

[10] G.-Y. Chen and L. Saloff-Coste, *The cutoff phenomenon for ergodic Markov processes*, *Electron. J. Probab.*, vol. 13, no. 3, pp. 26-78, 2008.

[11] G.-Y. Chen and L. Saloff-Coste, *The L^2-cutoff for reversible Markov processes*, J. Funct. Anal., vol. 258, no. 7, pp. 2246–2315, 2010.

[12] P. Diaconis, *The cutoff phenomenon in finite Markov chains*, *Proc. Nat. Acad. Sci. U.S.A.*, vol. 93(4), pp. 1659-1664, 1996.

[13] P. Diaconis and L. Saloff-Coste, *Logarithmic Sobolev inequalities for finite Markov chains*, *Ann. Appl. Probab.*, vol. 6(3), pp. 695-750, 1996.

[14] F. J. Dyson, *A Brownian-motion model for the eigenvalues of a random matrix*, *J. Math. Phys.*, vol. 3, pp. 1191-1198, 1962.

[15] A. Engoulatov, *A universal bound on the gradient of logarithm of the heat kernel for manifolds with bounded Ricci curvature*, *J. Funct. Anal.*, vol. 238(2), pp. 518-529, 2006.

[16] I. Gentil, *Logarithmic Sobolev inequality for diffusion semigroups*, arXiv e-prints, arXiv:1009.3421, 2010.

[17] A. Grigor'yan, *Heat Kernel and Analysis on Manifolds*, vol. 47, Providence, RI: American Mathematical Society (AMS); Somerville, MA: International Press, 2009.

[18] M. Lassalle, *Polynômes de Hermite généralisés*, *C. R. Acad. Sci. Paris Ser. I Math.*, 313(9):579-582, 1991.

[19] M. Lassalle, *Polynômes de Jacobi généralisés*, *C. R. Acad. Sci. Paris Ser. I Math.*, 312(6):425-428, 1991.

[20] M. Lassalle, *Polynômes de Laguerre généralisés*, *C. R. Acad. Sci. Paris Ser. I Math.*, 312(10):725-728, 1991.

[21] S. Li, X.-D. Li, and Y.-X. Xie, *On the law of large numbers for the empirical measure process of generalized Dyson Brownian motion*, *J. Stat. Phys.*, 181(4):1277-1305, 2020.

[22] P.-L. Meliot, *The cut-off phenomenon for Brownian motions on compact symmetric spaces*, *Potential Analysis*, 40(4):427-509, 2014.

[23] J. Salez, *Cutoff for non-negatively curved Markov chains*, preprint 2102.05597v1, 2021.

[24] L. Saloff-Coste, *Precise estimates on the rate at which certain diffusions tend to equilibrium*, *Mathematische Zeitschrift*, 217(1):641-677, 1994.

[25] L. Saloff-Coste, *Aspects of Sobolev-type inequalities*, volume 289, Cambridge University Press, 2002.

[26] L. Saloff-Coste, *On the convergence to equilibrium of Brownian motion on compact simple Lie groups*, *J. Geom. Anal.*, 14(4):715-733, 2004.

[27] C. Villani, *Optimal transport. Old and new*, volume 338, Berlin: Springer, 2009.

Samuel Chan-ashing, ENS Rennes - Sorbonne Université, 75005 Paris, France
e-mail: samuel.chan-ashing@ens-rennes.fr