



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *10/12/2014* par :

Pierre MONMARCHÉ

**Hypocoercivité : approches alternatives
et applications aux algorithmes stochastiques**

JURY

FRANÇOIS BOLLEY	Membre du Jury
PATRICK CATTIAUX	Président du Jury
DJALIL CHAFAÏ	Membre du Jury
ARNAUD GUILLIN	Membre du Jury
TONY LELIÈVRE	Rapporteur
FLORENT MALRIEU	Rapporteur
LAURENT MICLO	Directeur de thèse

École doctorale et spécialité :

EDMITT

Unité de Recherche :

Institut de Mathématiques de Toulouse (UMR 5219)

Directeur de Thèse :

Laurent MICLO

Rapporteurs :

Tony LELIÈVRE et Florent MALRIEU

*à Xavier,
qui n'aurait certes pas fait des probabilités.*

Table des matières

1	Introduction	13
1.1	Cadre général	13
1.2	L'ergodicité quantifiée	15
1.3	Le cas coercif classique	19
1.4	L'hypocoercivité	29
2	Modèle jouet hypocoercif	35
2.1	Motivations et résultat principal	35
2.2	Calcul exact de la norme	38
2.3	Du discret au continu	44
2.4	Du continu au mouvement brownien	47
2.5	Généralisations	48
2.6	Conclusion	50
2.7	Appendice	51
3	Méthode d'ordre 3	55
3.1	Un échantillonneur déterministe par morceau	55
3.2	Third order inequality	62
3.3	Numerical studies	71
3.4	Quelques remarques conclusives	73
3.5	Appendix	74
4	PDMP contractifs	77
4.1	Contexte et stratégie générale	77
4.2	Exponential decay	81
4.3	Functional inequalities for PDMP	85
4.4	Exemples	92
5	Processus RTP complètement dégénéré	107
5.1	Main results	108
5.2	Escape time at low temperature	111
5.3	Necessary and sufficient conditions for the cooling schedule	117
5.4	Non-minimal rate	132

Préambule

Remerciements

Tout au long d'une thèse, parce que ce n'est après tout qu'un morceau condensé de vie, on accumule des liens, des dettes, on est chaque jour un peu plus redevable envers ceux qui nous entourent. Que ces quelques lignes soient pour moi l'occasion, qui ne se présente pas assez souvent si explicitement, de remercier mes créanciers.

Une fois n'est pas coutume, je ne garderai pas le meilleur pour la fin et remercie sans attendre et chaleureusement mon directeur, le professeur Miclo. Non seulement ton sujet de thèse était passionnant et très bien calibré, Laurent, mais surtout, et même si tu étais le premier à plaisanter sur mon "autonomie", j'ai vraiment beaucoup retiré au contact de ton insatiable curiosité et ta motivation sans faille (notamment face à la lecture pas toujours évidente, je le sais, de ma prose).

C'est un grand plaisir pour moi que Tony Lelièvre et Florent Malrieu aient accepté de rapporter ma thèse (même remarque sur la prose) ; certains de leurs articles, de leurs exposés ou des discussions que nous avons pu avoir sont parmi les principales sources d'inspiration et de motivation de ma thèse. J'ai également eu la chance de rencontrer François Bolley, Djalil Chafaï (sur la page personnelle de qui tout doctorant devrait flâner un ou deux après-midi par an ; et cela devrait être validable comme formation doctorale) et Arnaud Guillin à plusieurs reprises, et je suis très heureux qu'ils aient accepté de faire parti de mon jury, au milieu d'une semaine pourtant déjà bien chargée. Je remercie également Patrick Cattiaux pour sa bonne humeur et sa vitalité, et pour avoir accepté, bien qu'il ait repéré l'entourloupe, de présider le jury.

Ces années à l'institut de mathématiques de Toulouse ne se seraient pas si bien passées sans l'aide de tous ses membres (scientifiques, administratifs et techniques), qui méritent évidemment ma plus grande gratitude.

Ma volonté de persévérer dans l'enseignement et la recherche en maths a véritablement germé à mon arrivée à Rennes, et n'a fait que se renforcer d'année en année, et ce pour deux raisons : la première, ce sont les cours que j'ai suivis, de celui d'un Michel (Pierre) à l'autre (Ledoux), en passant par ceux de François Castella, San Vu Gnoc ou Aldéric Joulin, et bien d'autres. Je garde un excellent souvenir des enseignements à l'ENS de Rennes qui ont su stimuler ma curiosité pendant trois ans ; le M2 à Toulouse m'a lancé dans la thèse. Comme je l'écris sur mes lettres de motivation (c'est vous dire ma franchise, non ?), je me considère en dette envers l'enseignement supérieur – au-delà des trois ans d'engagement qui me reste.

La seconde raison, évidemment, c'est tous ceux que j'ai rencontrés pendant ces années dans le domaine. Ils ont été nombreux à l'ENS grâce à qui les maths ont subitement cessé d'être

une simple matière scolaire, et encore plus nombreux à faire de mes trois ans à Rennes une tranche de vie épanouie. La promo de maths 2008 a ici une place particulière (ah, cette prépa agreg. . .).

Les rennais amènent naturellement aux visanais. Ceci dit, pour parler des visanais, il faudrait écrire une nouvelle thèse. Mais alors, que mettre dans la nouvelle section de remerciements ? Et après, pas moyen de fermer le système ; ce serait un travail de Sisyphe. Alors, en vrac, merci à Coko, Jerem, François, Alex, Fanny, Raton, Anne, Lolasse et Laulau, Viko, Pette, LLD, Mael, Judith, Coline, Manon, (1+Mad)Guigui, Alice, Xavier, Nounou, Oli, Cécile, Titi et Tibo, Simon, Clémence, Ludo, Hélianthe, Julien, Marie, Chachou, Jozy, Perrine, Ronan et Renan. . . ; certains avec qui je parle de maths entre deux Manouchka ou sur facebook, et d'autres avec qui je parle d'autres choses que de maths entre deux Hanabi ou autour d'une guitare.

Et puis, la jeunesse toulousaine. Y a-t-il réellement dans le monde des labos où il fait meilleur vivre une thèse ? Merci à Benoît sans qui je serais encore pire, à Nil sans qui je serais encore moins, à Laurent pour la rage, le tennis et la perceuse, aux éternelles organisatrices Malika et Mélanie pour leurs éternelles organisations ; merci à Maël pour les vannes et les pas vannes ; merci aux Claires (B,C,D) et aux Anto $(n + ine \times (B,C,P,V))$ pour m'avoir permis de gagner un peu de place dans ces remerciements, et pour pas mal d'autres choses aussi (en particulier, quant au dernier, pour les Ralph Azam ; d'ailleurs pendant que j'y pense, qu'en est-il du dernier tome ?), à Bertrand qui est à l'origine de ma soutenance anticipée ; merci aux précurseurs (Mathieu, Anne-Charline, Magali, Adil, Daniel) pour la sagesse des anciens (?) aux mecs-du-deuxième (Raphaël, Stéphane, Kévin, Sofiane, Yury ; hop, tous dans le même panier) pour les vannes sur Hulk ou du même acabit ; merci à mes anciens cobureaux (Olfa, Mawaki, Vuty) pour le calme (et bon courage, vous y êtes presque !), et aux nouveaux (Guillaume, Laure et Ana) pour les signes de vie (désolé pour toutes ces coinches refusées, la conjoncture n'était juste pas des meilleurs ces temps-ci. . .).

C'est aussi pour moi l'occasion de remercier mes parents ; après tout, bien que ma vocation professionnelle ne soit apparue qu'à partir de l'ENS, c'est-à-dire du moment où j'ai quitté le nid familial, il aurait été peu probable que j'arrive jusque à ce bac + 8 de mathématique si je n'avais pas, dès l'école primaire (peut-être même avant), abordé le monde, l'école en général et les cours de mathématiques en particulier avec la curiosité qu'ils m'ont léguée. Le petit cul a fait du chemin (et il n'est plus si petit, mais c'est un autre débat).

Merci à Zach Weinersmith, dont la lecture aura démarré la moindre de mes journées de travail pendant ces deux ans. Merci à Zalouest'h d'en être arrivé là (et à Gaspard en particulier : voilà, t'es content, t'es dans les remerciements, t'es même le seul du groupe, tiens, ça te fera les pieds), et aux Frères Pouettards pour les pouets ; c'est tout de même important d'avoir parfois des interlocuteurs pour qui x est moins une variable de \mathbb{R}^n qu'une catégorie de vidéos.

Évidemment, et même s'ils sont déjà remerciés, explicitement ou non, à plusieurs reprises ci-dessus, je ne peux que le répéter : merci Nil, merci Laurent, merci Simon, pour tout – et ça fait beaucoup. Et puis merde à la vaisselle.

Enfin, après réflexion, j'avais bel et bien gardé le meilleur pour la fin. Je n'ai pas trouvé de pitrerie stylistique pour le dire : merci, Anne.

Vue d'ensemble de la thèse

Prenons comme point de départ le recuit simulé. Cet algorithme, ou plutôt cette heuristique, proposée en 1983 par Kirkpatrick, Gelatt et Vecchi, s'inspire d'une méthode traditionnelle de métallurgie (le recuit), et plus généralement du principe selon lequel un matériau initialement chauffé doit être refroidi lentement pour, *in fine*, minimiser son énergie interne. En géologie, la vitesse de refroidissement est ce qui sépare un cristal de péridot bien arrangé d'un verre volcanique amorphe. Lorsque les atomes d'un matériau sont soudainement gelés, ils se figent instantanément, chacun minimisant à la va-vite l'énergie de liaison avec ses voisins, ce qui aboutit à une configuration proche du point de départ, peu efficace. De même l'algorithme de descente du gradient, qui consiste à optimiser une fonction U , disons de \mathbb{R}^d dans \mathbb{R} , en suivant le flot

$$X'_t = -\nabla U(X_t),$$

va aboutir au minimum local de U dont le bassin d'attraction contenait les paramètres initiaux X_0 . Le principe du recuit simulé est de perturber cette dynamique déterministe par un bruit aléatoire dont l'intensité est contrôlée par un paramètre de température T (par analogie avec l'agitation thermique des atomes), par exemple en

$$dX_t = -\nabla U(X_t)dt + \sqrt{T}dB_t$$

où $(B_t)_{t \geq 0}$ est un mouvement brownien standard (ce qu'on appelle parfois l'équation de Fokker-Planck). Lorsque T est fixé, le processus de Markov solution de cette équation différentielle stochastique (EDS) s'échappe presque sûrement de tout bassin d'attraction, ce qui permet d'explorer l'espace de paramètres. Diminuer la température rend de plus en plus difficile les mouvements localement contre-productifs, et le processus finit par se fixer à un minimum local, qu'on espère global.

Les intérêts d'une telle approche sont les suivants : l'exploration aléatoire de l'espace des paramètres est très peu sensible à la dimension de celui-ci, ou à la régularité du potentiel U (j'ai présenté ici le cas d'un processus à temps continu pour un potentiel régulier, mais l'implémentation se fait par des chaînes de Markov sur un espace d'états discret ; d'ailleurs le recuit simulé est particulièrement adapté à la résolution de questions difficiles de combinatoire, telles que le problème du voyageur de commerce). La méthode est simple à implémenter, robuste et demande peu de mémoire.

Les faiblesses de l'heuristique sont connues : d'abord le choix du schéma de température est laissé à l'utilisateur, et est donc essentiellement empirique (concrètement, la température est généralement fixée de façon adaptative). Ensuite le temps de calcul est extrêmement long. Les schémas de température logarithmiques qui assureraient, dans la théorie, la convergence presque sûre vers un minimum global, sont même d'une lenteur rédhibitoire pour l'application concrète. On préfère donc des schémas de température géométriques (cf. [23]), dont les performances sont aléatoires, ou encore des schémas adaptatifs (cf. [50]).

La principale cause de la longueur des temps de calcul est le phénomène de métastabilité : à faible température, le processus passe un temps très long dans certaines régions (dites métastables) avant de parvenir à franchir une barrière énergétique ou entropique et de poursuivre son exploration. Cette difficulté est loin d'être l'apanage du recuit simulé ; on la trouve par exemple dans de nombreuses réactions chimiques issues de l'industrie ou de la biologie. Une solution pour contourner le problème consiste à biaiser le processus pour le forcer à explorer de nouvelles régions (marches aléatoires auto-évitantes [14], *adaptive biasing force technique*

[68]...). Ces méthodes ont une efficacité réelle. Néanmoins elles nécessitent de garder en mémoire la trajectoire passée.

Une approche alternative (éventuellement complémentaire) consiste à trouver des dynamiques markoviennes (sans mémoire donc) plus efficace que celle de Fokker-Planck ci-dessus. La convergence de l'algorithme repose sur le fait qu'à T fixé, la mesure invariante du processus est $e^{-U(x)/T}dx$, la mesure de Gibbs associée à U . Cette mesure invariante n'est pas modifiée si on ajoute à la force déterministe $-\nabla U$ un biais b , tant qu'il est sans divergence, c'est-à-dire tant que $\nabla \cdot (be^{-U/T}) = 0$ (cf [67]). On peut également considérer des processus position-vitesse, comme la dynamique de Langevin :

$$\begin{aligned} dX_t &= Y_t dt \\ dY_t &= -\nabla U(X_t)dt - Y_t dt + \sqrt{2T}dB_t, \end{aligned}$$

dont la mesure invariante, $e^{-\left(U(x)+\frac{y^2}{2}\right)/T} dx dy$, possède la première marginale ad hoc. Ces processus cinétiques ont une mémoire instantanée (la vitesse en plus de la position) et en conséquence une inertie plus grande que le processus réversible initial, ce qui, on l'espère (et on le constate empiriquement), doit les aider à sortir des bassins d'attraction. On peut même considérer des mémoires instantannées plus générales qu'une simple vitesse (comme dans [48]).

Si ces exemples se sont déjà avérés efficaces en pratique, ils posent cependant des problèmes théoriques. En effet ces dynamiques sortent du cadre d'application des méthodes classiques pour quantifier la vitesse de convergence d'un processus vers son équilibre. Or cette quantification est cruciale pour calibrer (au moins en théorie) le schéma de température. L'étude de ces processus qu'on pourrait taxer de dégénérés embrasse un cadre bien plus général que la simple application au recuit simulé. Après tout, toutes les équations un tant soit peu issues de la mécanique newtonienne sont cinétiques ; ce n'est qu'un exemple. Dans les années 2000 de premiers résultats sont apparus dans ce champs que Villani a depuis baptisé l'hypocoercivité ([86] ; le terme lui-même venant de Gallay). Nous reviendrons sur cette notion de façon plus détaillée hors de cette introduction, mais je veux ici simplement souligner que, malgré les nombreux progrès faits en la matière, les méthodes actuelles demeurent peu précises, démontrant souvent une convergence exponentielle, mais sans estimer raisonnablement le bon ordre de grandeur du taux. Or, répétons-le, c'est bien une information quantitative dont nous avons besoin.

Il y avait donc deux objectifs à l'aube de cette thèse. D'une part, définir de nouvelles dynamiques de simulation de mesure de Gibbs et donner des conditions (théoriques a priori, pratiques dans l'idéal) sur le schéma de température pour qu'un algorithme de recuit simulé basé sur cette dynamique converge vers les minima globaux. D'autre part développer de nouvelles méthodes hypocoercives dans l'optique d'une quantification plus fine que ne le permettait l'état de l'art d'alors ; ce qui devait aider à réaliser le premier objectif.

L'organisation de ce manuscrit suit dans les grandes lignes la chronologie de la thèse dans la mesure où, si la façon dont elle s'est développée n'a pas été linéaire, elle a du moins été logique.

Il fallait au départ trouver une dynamique qui explore l'espace plus efficacement que la marche isotrope. L'un des premiers principes qui vient à l'esprit est de ne pas repasser deux fois au même endroit : on a déjà parlé des marches auto-évitantes, qui nécessitent, une fois implémentées, beaucoup de mémoire. Un compromis entre ces deux extrêmes est de se rappeler uniquement où se trouvait la marche à l'instant précédent, et de pénaliser le retour en arrière.

Ceci aboutit aux marches d'ordre 2 de Diaconis *et al.* ([35, 36]) : sur le tore discret $\mathbb{Z}/N\mathbb{Z}$, Diaconis et Miclo montrent que le trou spectral de ces marches “persistantes” est supérieur à celui de la marche réversible, ce qui, *si la marche d'ordre 2 était réversible*, signifierait une convergence à l'équilibre plus rapide (ces notions seront détaillées dans la suite de l'introduction). Le premier travail de ma thèse a donc constitué en une analyse plus fine de l'équivalent continu des marches persistantes, le processus du télégraphe intégré sur le cercle (ce qui a donné lieu à la publication de [75]).

De l'étude de ce modèle jouet il ressort plusieurs choses. D'abord la convergence est belle et bien asymptotiquement plus rapide pour le processus irréversible que pour le brownien (équivalent continu de la marche simple). Cependant ce n'est pas le cas pour des temps courts ; si l'on note $d(t)$ la distance (en norme L^2 d'opérateur, détails plus loin) entre la loi de X_t , le télégraphe intégré, et la mesure uniforme sur le cercle, alors $d(0) - d(t)$ est de l'ordre de t^3 quand t est proche de 0, alors que c'est de l'ordre de t pour le brownien. D'autre part, si la convergence de $d(t)$ vers 0 est bien (asymptotiquement) exponentielle, elle présente des oscillations. Ce phénomène a aussi été observé numériquement pour l'équation de Boltzmann ([45]) et théoriquement pour d'autres processus cinétiques ([47]).

La simulation de la loi uniforme sur le cercle par une marche aléatoire n'étant pas ce qu'il y a de plus pertinent en pratique, il s'agissait ensuite d'adapter le principe du télégraphe intégré pour une mesure de Gibbs quelconque. On est donc amené à étudier le processus RTP (pour *run and tumble process*), qui est un processus cinétique (X, Y) dont la vitesse Y est un processus de saut pur sur la sphère unité de \mathbb{R}^d . Évidemment, contrairement au cas du cercle, les calculs ne sont plus explicites, et contrairement au cas de la dynamique réversible de Fokker-Planck, que la mesure invariante satisfasse une inégalité de Poincaré ne permet pas de conclure que $d(t)$ vérifie une inégalité différentielle de la forme $d'(t) \leq -\lambda d(t)$. Au vu du comportement de d dans le cas du cercle, on peut néanmoins envisager qu'elle vérifie une inégalité d'ordre 3, de la forme

$$d'''(t) + \lambda_1 d''(t) + \lambda_2 d'(t) + \lambda_3 d(t) \leq 0.$$

Il s'avère que c'est faux. Cependant on peut bel et bien trouver des constantes a et b et une fonction $t \mapsto c(t)$ telle que

$$(\partial_t + a) \left(((\partial_t + b)^2 + c) d \right) (t) \leq 0,$$

de telles sortes, de plus, que cette inégalité implique une décroissance exponentielle. Cela est également vrai dans le cas de la dynamique de Langevin cinétique, et ce résultat a été présenté dans l'article [76]. Les intérêts et les défauts de la méthode apparaîtront plus clairement au fil de ce manuscrit ; dans ce rapide survol de la thèse, constatons simplement qu'elle ne fournissait, quantitativement, que des informations d'une précision du même ordre que les méthodes existantes. C'est-à-dire une précision insuffisante quant à la fin de notre programme. D'autre part elle nécessitait que le taux de saut du processus soit minoré par une constante positive, ce qui revient à demander que le processus ait, en quelque sorte, un comportement dissipatif résiduel. Or notre objectif initial était justement de réduire au maximum ce comportement.

Pour gagner en précision, on pouvait donc se demander s'il n'y avait pas une “bonne” quantité, \bar{d} plutôt que d , à dériver, une approche déjà développée par Villani dans [85]. Avant de se lancer au petit bonheur à dériver tout et n'importe quoi, il fallait, encore une fois, essayer de comprendre ce qui se passait sur des exemples simples. Le RTP, en un sens, a deux bonnes raisons d'être hypocoercif. D'abord il est cinétique, ensuite il est déterministe

par morceau (c'est donc un PDMP, pour *piecewise deterministic Markov process*). D'où l'idée d'appliquer la méthode précédente à des PDMP plus simples, tels que le processus du TCP (pour *Transmission Control Protocol*), issu de la modélisation informatique. Une méthode classique, qu'on baptise parfois "du Γ_2 ", consiste à dériver le long du semi-groupe un objet appelé carré du champs qui, dans le cas par exemple d'une diffusion de Fokker-Planck, coïncide avec le carré du gradient. Pour le TCP, dériver le carré du champs ne donne rien, en revanche dériver le carré du gradient conduit, par des arguments similaires au cas classique, à une convergence exponentielle vers l'équilibre. Ceci a au passage nécessité de démontrer que les mesures invariantes des PDMP considérés satisfaisaient des inégalités de Poincaré, alors qu'on ne possède pas de formules analytiques maniables pour leurs densités. À l'heure où j'écris ces lignes, ce travail est soumis à publication.

À ce point de la thèse, ce travail sur les PDMP a soulevé un certain nombre de pistes sur des questions théoriques d'hypocoercivité. Mais, avec un peu de recul, cela avait peu de chance, en tout cas à court terme, de régler la question du RTP dont le taux de saut s'annule ; processus, en un sens, particulièrement dégénéré. En effet ce processus est complètement déterministe sur des intervalles de temps entiers, eux-mêmes déterministes ; or les inégalités différentielles ci-dessus sont très liées à la création d'entropie, et donc en quelque sorte à la quantité d'aléa injecté dans le système à *tout instant*. En poussant le raisonnement, cela voulait également dire que même pour le RTP non dégénéré, avec une méthode qui reposerait essentiellement sur l'aléa résiduel, minimal, injecté à chaque instant, on ne pouvait pas espérer obtenir le bon ordre de grandeur du taux de convergence.

Fort heureusement, si le RTP dégénéré apparaît recalcitrant s'agissant d'une méthode systématique de type inégalité fonctionnelle, sa dynamique le long d'une trajectoire se décrit assez aisément (au moins en dimension 1). Par exemple lorsqu'il se trouve dans un minimum local, la probabilité qu'une tentative d'en sortir soit couronnée de succès est explicite et ne dépend que de la barrière énergétique à franchir. De là une asymptotique à basse température du temps de sortie du bassin d'attraction d'un minimum local, qui mène à traiter la question d'un processus qui sauterait d'un bassin à un autre en des temps exponentiels (de paramètre dépendant des profondeurs des puits). Une adaptation du travail d'Hajek sur le cas discret ([53]) permet de démontrer une condition nécessaire et suffisante pour le schéma de température d'un recuit simulé basé sur le RTP dégénéré pour assurer la convergence en probabilité vers les minima globaux du potentiel.

Chapitre 1

Introduction

Les démonstrations se font à l'aide de notions, et non de notations.

Carl Friedrich Gauß

En général, dans un papier, entre l'abstract et le non-sense, il y a l'introduction.

Dominique Bakry

1.1 Cadre général

1.1.1 Quelques notations

Précisons d'abord quelques notations usuelles de probabilités qui se retrouveront tout au long de ce manuscrit. Si Y est une variable aléatoire (v.a.), $\mathcal{L}(Y)$ désignera sa loi et $\mathbb{E}[Y]$ son espérance. Si la v.a. Z n'est pas définie en amont et si ν est une loi de probabilité, $\mathbb{E}_{Z \sim \nu}[f(Z)]$ ou, plus simplement, $\mathbb{E}_\nu[f(Z)]$ désignera l'espérance de $f(Z)$ lorsque Z a pour loi ν . En certaines occasions il sera cependant plus confortable de considérer une loi ν comme un opérateur sur un espace fonctionnel, auquel cas on adoptera plus volontier la notation νf pour cette même espérance. En revanche si $(X_t)_{t \geq 0}$ est un processus de Markov, $\mathbb{E}_\nu[f(X_t)]$ désignera l'espérance de $f(X_t)$ lorsque X_0 est de loi ν (et naturellement si $\nu = \delta_x$ la masse de Dirac en x , on la note simplement $\mathbb{E}_x[f(X_t)]$). La loi normale de moyenne m et variance σ sera noté $\mathcal{N}(m, \sigma)$, et la loi exponentielle de paramètre λ , $\mathcal{E}(\lambda)$, et la loi uniforme sur un ensemble V , \mathcal{U}_V . Enfin, de manière générale une barre transversale signifiera un conditionnement par une tribu, une v.a. ou un événement : ainsi $\mathbb{E}_{X \sim \mathcal{N}(0,1)}[X \mid X > 0]$ désignera-t-il par exemple l'espérance d'une gaussienne standard conditionnée à être positive.

1.1.2 Semi-groupes de Markov

Rappelons quelques définitions, notations et propriétés qui nous seront utiles. Si $(X_t)_{t \geq 0}$ est un processus de Markov à valeur dans un espace polonais V , on définit une famille d'opérateurs associé $(P_{s,t})_{0 \leq s \leq t}$ par

$$P_{s,t}f(x) = \mathbb{E}_{X_s=x}[f(X_t)],$$

agissant a priori sur les observables $f \in \mathcal{M}_b(V)$, les fonctions mesurables bornées sur V . Dans tous les cas de ce manuscrit, V sera muni de sa tribu borélienne et d'une mesure μ (en général l'unique mesure invariante du processus, le cas échéant), et l'action s'étendra à $L^2(\mu)$. Le générateur de $(P_{s,t})_{0 \leq s \leq t}$ sera noté $t \mapsto L_t$, défini par

$$L_t f = \lim_{s \rightarrow 0_+} \frac{P_{t,t+s} f - f}{s},$$

pour l'ensemble des f telles que cette limite existe dans $L^2(\mu)$, appelé le domaine de L_t et noté $\mathcal{D}(L_t)$.

En toute rigueur on ne parle de semi-groupe que dans le cas où le processus est homogène, auquel cas $P_{s,t}$ ne dépend que de $t - s$ et sera simplement noté $P_t = P_{0,t}$, et son générateur ne dépend pas du temps et sera simplement noté L . Notons que, par dualité, l'action à droite de P_t sur $\mathcal{M}_b(V)$ implique une action à gauche sur $\mathcal{P}(V)$ les probabilités de V : si $\nu \in \mathcal{P}(V)$, la loi νP_t est définie par

$$\forall f \in \mathcal{M}_b(V), \quad (\nu P_t) f = \nu(P_t f) = \mathbb{E}_{X_0 \sim \nu} [f(X_t)].$$

Les semi-groupes Markoviens que nous considérerons seront toujours de Feller, c'est-à-dire qu'en plus de la propriété de semi-groupe $P_{t+s} = P_t P_s$ et de la contraction sur les fonctions bornées $\|P_t f\|_\infty \leq \|f\|_\infty$, la propriété de continuité $\|P_t f - f\|_\infty \xrightarrow{t \rightarrow 0} 0$ sera satisfaite pour les fonctions f continues. De plus l'espace d'états V sera toujours soit fini, soit une variété différentielle régulière. Dans le second cas, les fonctions régulières (c'est-à-dire infiniment différentiables) dont toutes les dérivées sont dans $L^2(\mu)$ formeront, à chaque occasion de ce manuscrit, une famille stable par $(P_t)_{t \geq 0}$ et L , contenant les constantes, et dense dans $L^2(\mu)$. Autrement dit l'hypothèse de l'existence d'une algèbre standard \mathcal{A} (pour reprendre les termes de [5, définition 2.4.2]) sera toujours faite. Les observables f seront toujours considérés dans cette algèbre \mathcal{A} , de sorte que les opérations successives qu'elles subiront restent licites. Autrement dit, un peu cavalièrement, on ne s'occupera pour ainsi dire pas de questions de régularité et/ou d'intégrabilité, ce afin de mieux se concentrer sur les problèmes de convergences, qui constituaient le cœur de la thèse.

Rappelons les propriétés suivantes :

Proposition 1.1.

1. $P_t 1 = 1$ et $L 1 = 0$.
2. pour tout $t \geq 0$ et $f \in \mathcal{A}$,

$$\partial_t (P_t f) = P_t L f = L P_t f.$$

(a priori l'égalité a lieu dans $L^2(\mu)$; elle est ponctuelle quand la convergence de $\frac{1}{t} (P_t f - f)$ l'est).

3. les trois assertions suivantes sont équivalentes (proposition 9.2 p. 239 de [44]) :
 - μ est invariante pour $(X_t)_{t \geq 0}$, au sens $\mathcal{L}_\mu(X_t) = \mu$ pour tout $t \geq 0$.
 - $\mu P_t = \mu$ pour tout t .
 - $\mu L = 0$.
4. Sont également équivalentes :
 - μ est réversible pour $(X_t)_{t \geq 0}$, au sens $\mathcal{L}_{X_T \sim \mu}(X_{T-s}) = \mathcal{L}_{X_0 \sim \mu}(X_s)$ pour tout $T \geq s \geq 0$.

- P_t est symétrique dans $L^2(\mu)$ pour tout $t \geq 0$.
- L est symétrique dans $L^2(\mu)$.

D'un point de vue spectral, la première propriété implique que les constantes sont propres pour L et P_t respectivement pour les valeurs propres 0 et 1, et la seconde qu'une fonction propre pour L pour la valeur propre θ est propre pour P_t pour la valeur propre $e^{\theta t}$.

Pour en revenir au cas inhomogène, l'égalité de Chapman-Kolmogorov $P_{s,t}P_{u,s} = P_{u,t}$ reste vraie, et

$$\partial_t (P_{s,t}f) = P_{s,t}L_t f.$$

Sauf mention contraire, les processus de Markov que nous considérerons seront par défaut homogènes.

1.1.3 Mesure de Gibbs

Considérons un espace mesuré (V, \mathcal{F}, dm) ; concrètement, V sera discret et dm la mesure de comptage, ou bien V sera une variété riemannienne régulière et dm sa mesure de Lebesgue. Soit U une fonction de V dans \mathbb{R} minorée et telle que $e^{-U} \in L^1(dm)$. La loi de Gibbs associée à U est alors la mesure de probabilité de densité $Z^{-1}e^{-U}$ par rapport à dm , où $Z = \int e^{-U} dm$ est la constante de normalisation. Dans ce cadre U sera appelé un potentiel.

De manière plus générale, si $\varepsilon > 0$, on appellera loi de Gibbs associée à U à température ε la mesure de probabilité

$$\mu_\varepsilon = \frac{e^{-\frac{U(x)}{\varepsilon}}}{\int_V e^{-\frac{U}{\varepsilon}} dm} dm.$$

L'intérêt de cette mesure, du point de vue de l'optimisation de fonctions, est que, si \mathcal{N} est un voisinage de l'ensemble des minima globaux de U ,

$$\mathbb{P}_{\mu_\varepsilon} (X \in \mathcal{N}) \xrightarrow{\varepsilon \rightarrow 0} 1.$$

Si en particulier U n'admet qu'un unique minimum global, μ_ε converge faiblement vers la masse de Dirac en ce minimum (on sait en fait décrire la loi limite dans des cas bien plus généraux; voir par exemple dans [73]).

1.2 L'ergodicité quantifiée

Cohabitent un certain nombre de définitions naturelles de l'ergodicité d'un processus de Markov admettant une unique mesure invariante μ . L'une d'elle est la convergence faible pour tout x de $\mathcal{L}_x(X_t) = \delta_x P_t$ vers μ ; ou de façon équivalente la convergence ponctuelle pour tout f bornée et continue de $P_t f$ vers la constante μf . Plutôt que de se cantonner à cette définition qualitative, on considérera différentes quantités d_t de la forme $d(\delta_x P_t, \mu)$ - en voyant μ comme une mesure de probabilité - ou $d(P_t, \mu)$ - en voyant μ comme un opérateur sur un espace fonctionnel - positives, dont la convergence vers 0 implique la convergence faible précédente, ou bien une convergence en norme d'opérateurs. Ce ne seront pas nécessairement de vraies distances définies sur des espaces de mesures.

Dans le domaine de l'analyse et des EDP ces quantités sont parfois appelées des fonctionnelles de Lyapunov; alors que dans le monde probabiliste ce terme désigne autre chose (comme dans [9] par exemple).

On dira que la convergence est exponentielle s'il existe $\lambda > 0$ et une fonction à croissance au plus polynomiale $C(t)$ tels que $d_t \leq C(t)e^{-\lambda t}$, auquel cas C sera désigné comme le préfacteur et λ le taux de cette (borne de la) convergence. Le taux optimal, ou taux théorique, est le supremum des λ tel qu'une telle inégalité ait lieu.

Passons en revue les trois grandes familles de fonctionnelles qui apparaîtront dans cette thèse.

1.2.1 Les normes d'opérateurs

D'après l'inégalité de Hölder, et de par l'invariance de μ , pour tout $t \geq 0$, pour tout $p \geq 1$ et pour tout $f \in L^p(\mu)$,

$$\|P_t f\|_p \leq \|f\|_p,$$

autrement dit P_t agit continuellement sur $L^p(\mu)$ (et en est même une contraction). On peut donc considérer la norme p de l'opérateur $P_t - \mu$,

$$\|P_t - \mu\|_p = \sup \{ \|P_t f - \mu f\|_p, \|f\|_p = 1 \}.$$

On s'intéressera plus particulièrement au cas $p = 2$ (et en l'absence de confusion possible on notera simplement $\|P_t - \mu\|$), qui permet de bénéficier de toute l'artillerie hilbertienne et de la théorie spectrale.

Plus généralement, lorsque V est une variété régulière de dimension d , on peut s'intéresser à des normes de Sobolev plus générales. En notant

$$W_m^p(\mu) = \left\{ f \in L^p(\mu), \forall \alpha \in \mathbb{N}^d \text{ tel que } \sum_{i=1}^d \alpha_i < m, \partial_\alpha f \in L^p(\mu) \right\}$$

les espaces de Sobolev de (V, μ) , si P_t agit continuellement sur W_m^p , on peut considérer

$$\|P_t - \mu\|_{W_m^p(\mu)} = \sup \{ \|P_t f - \mu f\|_{W_m^p(\mu)}, \|f\|_{W_m^p(\mu)} = 1 \}.$$

1.2.2 Les distances de Wasserstein

On considère un espace polonais V muni d'une distance r , et pour $p > 0$ l'on note $\mathcal{P}_p(V)$ les probabilités de $p^{\text{ième}}$ moment fini, c'est-à-dire les lois ν sur V telles qu'il existe un $x_0 \in V$ tel que

$$\mathbb{E}_\nu [(r(Z, x_0))^p] < \infty$$

(ce qui ne dépend a posteriori pas de x_0). On définit la distance de Wasserstein p entre deux probabilités ν_1, ν_2 de $\mathcal{P}_p(V)$ par

$$W_p(\nu_1, \nu_2) = \inf_{X \sim \nu_1, Y \sim \nu_2} \mathbb{E} [(r(X, Y))^p].$$

Pour $p \geq 1$, c'est bel et bien une distance. Lorsque (X, Y) est de loi π dont la première marginale (c'est-à-dire la loi de X) est ν_1 et la seconde ν_2 , on dit que (X, Y) (ou, de façon équivalente, π) est un couplage de ν_1 et ν_2 . L'un des avantages de la distance de Wasserstein est que la construction de tout couplage en fournit automatiquement une majoration.

La convergence en distance de Wasserstein est équivalente à la convergence en loi plus la convergence des p premiers moments (théorème 6.9 p. 108 de [87])

Rappelons la formulation duale de la distance de Wasserstein W_1 (qu'on appellera parfois W , la distance de Wasserstein), par le théorème de Kantorovich-Rubinstein (on n'utilisera pas de théorème de dualité de type Monge-Kantorovich pour des distances de Wasserstein générales; le lecteur intéressé pourra consulter [82, 13]). Si l'on note

$$\|f\|_{Lip} = \sup_{(x,y) \in V^2} \left\{ \frac{|f(x) - f(y)|}{r(x,y)} \right\}$$

la norme Lipschitz, alors

$$W_1(\nu_1, \nu_2) = \sup \{(\nu_1 - \nu_2)f, \|f\|_{Lip} \leq 1\}.$$

Dans le cas particulier où $r(x,y) = 1_{x \neq y}$, on parle de la distance en variation totale, qui possède plusieurs représentations :

$$\begin{aligned} d_{vt}(\nu_1, \nu_2) &= \inf_{X \sim \nu_1, Y \sim \nu_2} \mathbb{P}(X \neq Y) \\ &= \sup \left\{ (\nu_1 - \nu_2)f, \|f\|_\infty \leq \frac{1}{2} \right\} \\ &= \sup_{A \subset V} |\mathbb{P}_{\nu_1}(Z \in A) - \mathbb{P}_{\nu_2}(Z \in A)| \\ &= \frac{1}{2} \|f_1 - f_2\|_{L^1(\nu)} \quad \text{si } \nu_i = f_i \nu, i = 1, 2. \end{aligned}$$

La convergence en variation totale est très forte, et on la considérera assez peu. Notons que, contrairement au cas des autres distances de Wasserstein, $d_{vt}(\delta_x, \delta_y) = 1$ quelle que soit la position relative de x et y ; ce qui fait dire que la distance en variation totale ne voit pas du tout la géométrie de l'espace sous-jacent, dès que celle-ci est non triviale.

1.2.3 Les entropies

Si Φ est une fonction convexe d'un intervalle $I \subset \mathbb{R}$ dans \mathbb{R} et $f \in \mathcal{A}$ alors, d'après l'inégalité de Jensen, pour une fonction à valeur dans I ,

$$\text{Ent}_{\mu, \Phi} f = \mathbb{E}_\mu [\Phi(f)] - \Phi(\mathbb{E}_\mu [f]),$$

appelée la Φ -entropie de f par rapport à μ , est une quantité positive (en l'absence de confusion possible, on omettra l'indice μ). De plus si la convexité de Φ est stricte, $\text{Ent}_\Phi f = 0$ si et seulement si f est constante. Puisque, pour les processus ergodiques, $P_t f$ converge vers la constante μf , on peut décider de mesurer la vitesse de cette convergence à l'aune de

$$d(P_t, \mu) = \sup_{Im(f) \subset I, \text{Ent}_\Phi f \neq 0} \frac{\text{Ent}_\Phi P_t f}{\text{Ent}_\Phi f}.$$

Pour la fonction $\Phi(x) = x^2$ sur \mathbb{R} , on retrouve la variance

$$\text{Var}_\mu f = \mathbb{E}_\mu [f^2] - (\mathbb{E}_\mu [f])^2 = \mathbb{E}_\mu [(f - \mathbb{E}_\mu [f])^2] = \|f - \mu f\|_{L^2(\mu)}^2.$$

La dernière écriture s'interprète d'un point de vue hilbertien comme suit : μf étant le projeté orthogonal dans $L^2(\mu)$ de f sur la droite des constantes, $f - \mu f$ est son projeté orthogonal sur

1^\perp l'orthogonal des constantes, c'est-à-dire les fonctions d'intégrale nulle par rapport à μ . On retombe sur la norme 2 d'opérateur : puisque P_t et μ coïncident sur les constantes, on voit que

$$\begin{aligned}\|P_t - \mu\|_2^2 &= \sup_{f \in L^2(\mu) \setminus \{0\}} \frac{\|P_t f - \mu f\|_2^2}{\|f\|_2^2} \\ &= \sup_{f \in 1^\perp} \frac{\|P_t f - \mu f\|_2^2}{\|f\|_2^2} \\ &= \sup_{f \in 1^\perp} \frac{\text{Var}_\mu P_t f}{\text{Var}_\mu f} = d(P_t, \mu).\end{aligned}$$

L'autre cas classique est $\Phi(x) = x \log(x)$ de $[0, \infty)$ dans \mathbb{R} , qui donne l'entropie de Boltzmann usuelle pour les fonctions positives,

$$\text{Ent}_B f = \mathbb{E}[f \log f] - \mathbb{E}[f] \log \mathbb{E}[f].$$

Pour une étude générale des Φ -entropies, on renvoie aux textes [26, 17] de Chafaï, d'une part, Bolley et Gentil de l'autre. Quant à cette thèse, elle n'a considéré que des entropies particulières, correspondant aux fonctions

$$\tilde{\Phi}_p(x) = \frac{x^p - x}{p - 1}$$

pour $p \in (1, 2]$, c'est-à-dire

$$\text{Ent}_{\tilde{\Phi}_p} = \frac{\mathbb{E}[f^p] - (\mathbb{E}[f])^p}{p - 1}.$$

Plus précisément on posera

$$\text{Ent}_p f = \text{Ent}_{\tilde{\Phi}_p} \left(f^{\frac{2}{p}} \right) = \frac{\mathbb{E}[f^2] - \left(\mathbb{E} \left[f^{\frac{2}{p}} \right] \right)^p}{p - 1}.$$

Ces p -entropies, introduites par Beckner ([12]), interpolent entre la variance et l'entropie. En effet, si $1 < p < q < 2$,

$$\text{Var} f = \text{Ent}_2 f \leq \text{Ent}_q f \leq \text{Ent}_p f \xrightarrow{p \rightarrow 1} \text{Ent}_B(f^2).$$

On notera $\text{Ent}_1 f = \text{Ent}_B(f^2)$. Notons que la convergence en entropie de Boltzmann entraîne la convergence en variation totale, en vertu de l'inégalité de Csiszár-Kullback-Pinsker : si $\mu f = 1$,

$$\frac{1}{2} \|f - 1\|_{L^1(\mu)}^2 \leq \text{Ent}_B f.$$

Si, plutôt qu'à la convergence de $P_t f$ vers μf , on s'intéresse à celle de νP_t vers μ , on peut étendre la définition d'entropie par rapport à μ aux mesures de la façon suivante

$$H(\nu|\mu) = \begin{cases} \text{Ent}_B \left(\frac{d\nu}{d\mu} \right) & \text{si } \nu \prec \mu \\ +\infty & \text{sinon.} \end{cases}$$

L'inégalité de Csiszár-Kullback-Pinsker se lit alors

$$2d_{vt}^2(\nu, \mu) \leq H(\nu|\mu).$$

1.3 Le cas coercif classique

Tout au long de cette section, on pourra se fixer les idées en pensant au processus de Fokker-Planck sur \mathbb{R}^d , solution de l'équation différentielle stochastique

$$dX_t = -\nabla U(X_t) + \sqrt{2}dB_t, \quad (1.1)$$

de générateur

$$Lf(x) = -\nabla U(x) \cdot \nabla f(x) + \Delta f(x). \quad (1.2)$$

1.3.1 La variance et Poincaré

On commence par s'intéresser au cas de la variance, c'est-à-dire à la norme 2, pour l'étude de laquelle on dispose, comme on va le voir, des outils de la théorie spectrale. Considérons un processus markovien $(X_t)_{t \geq 0}$ sur \mathbb{R}^d dont on supposera qu'il possède une unique mesure invariante μ . Notons $(P_t)_{t \geq 0}$ et L le semi-groupe et le générateur associés.

Définition 1.2. On définit le carré du champ Γ par

$$\forall f, g \in \mathcal{A}, \quad 2\Gamma(f, g) = L(fg) - fLg - gLf.$$

C'est une forme bilinéaire symétrique, et l'on notera $\Gamma(f) = \Gamma(f, f)$ la forme quadratique associée. Dans la mesure où

$$\Gamma(f) = \lim_{t \rightarrow 0} \frac{P_t(f^2) - (P_t f)^2}{2t},$$

l'inégalité de Cauchy-Schwarz prouve que $\Gamma f \geq 0$.

Définition 1.3. On dira que μ satisfait une inégalité de Poincaré de constante c pour Γ si

$$\forall f \in \mathcal{A}, \quad \mu \left[(f - \mu f)^2 \right] \leq c \mu [\Gamma(f)].$$

Autrement dit, en notant $\mathcal{E}_\mu(f) = \int \Gamma(f) d\mu$, si

$$\forall f \in \mathcal{A}, \quad \text{Var}_\mu f \leq c \mathcal{E}_\mu(f). \quad (1.3)$$

On appelle $\mathcal{E}_\mu(f)$ l'énergie de f par rapport à μ (lorsqu'il n'y a pas d'ambiguïté sur Γ ; hors du cadre classique, on verra que cela peut prêter à confusion). Remarquons que, dans la mesure où par invariance $\mu L = 0$,

$$\begin{aligned} \mu \Gamma(f) &= -\mu f Lf \\ &= -\langle f, Lf \rangle_{L^2(\mu)} \quad (\text{produit scalaire dans } L^2(\mu)). \end{aligned}$$

La positivité de Γ entraîne que les parties réelles des valeurs propres de L sont nécessairement négatives. Si de plus μ est réversible pour $(P_t)_{t \geq 0}$, c'est-à-dire si L est symétrique dans $L^2(\mu)$, son spectre est réel. Dans ce cas l'inégalité de Poincaré, qu'on lit maintenant

$$\forall f \in \mathcal{A}, \quad \langle f, Lf \rangle \leq -\frac{1}{c} \|f - \mu f\|^2,$$

est équivalente au fait que les seules fonctions propres de L associées à 0 sont les constantes, et que $\sigma(L)$, le spectre de L , privé de 0, c'est-à-dire le spectre de L sur l'orthogonale dans $L^2(\mu)$ des constantes, est inclus dans $(-\infty, -c^{-1}]$; on l'appelle donc également inégalité de trou spectral (ce trou étant la distance entre la plus grande valeur propre de L , à savoir 0, et les autres).

Ce que nous avons jusqu'ici appelé la "méthode classique" pour quantifier la convergence à l'équilibre *via* des inégalités fonctionnelles réside, en sa version la plus simple, dans l'énoncé suivant :

Théorème 1.4 (2.5.5 de [5]). *L'inégalité de Poincaré (1.3) est équivalente à ce que pour tout $f \in \mathcal{A}$,*

$$\text{Var}_\mu P_t f \leq e^{-\frac{2t}{c}} \text{Var}_\mu f.$$

Démonstration. Commençons par le sens direct. Quitte à considérer $g = f - \mu f$, on peut se ramener à une fonction d'intégrale nulle. De $\partial_t P_t f = L P_t f$, de la différentielle d'une forme quadratique et par invariance on déduit

$$\begin{aligned} \partial_t \|P_t f\|^2 &= 2 \langle P_t f, L P_t f \rangle \\ &= -2\mu \Gamma(P_t f) \\ &\leq -\frac{2}{c} \|P_t f\|^2 \end{aligned}$$

Sachant que $P_0 f = f$, le lemme de Gromwall conclut.

Pour démontrer la réciproque, on constate en faisant tendre t vers 0 que $\|P_t f\|^2 \leq e^{-\frac{2}{c}t} \|f\|^2$ implique que

$$2 \langle f, L f \rangle \leq -\frac{2}{c} \|f\|^2,$$

ce qui est exactement (1.3). □

Notons que la démonstration de la réciproque utilise le fait que les deux fonctions $\|P_t f\|^2$ et $e^{-\frac{2}{c}t} \|f\|^2$ ont la même valeur en 0; en particulier cet argument ne permet pas de déduire une inégalité de Poincaré à partir d'une inégalité de la forme

$$\text{Var}_\mu P_t f \leq C e^{-\frac{2t}{c}} \text{Var}_\mu f.$$

si $C > 1$. De fait, une telle inégalité n'entraîne pas forcément d'inégalité de Poincaré, comme les nombreux cas hypocoercifs que nous verrons plus tard le démontrerons. On peut tout de même remarquer ceci :

Proposition 1.5. *Si le processus est réversible et s'il existe $\lambda > 0$ et $C > 1$ tels que*

$$\text{Var}_\mu P_t f \leq C e^{-\lambda t} \text{Var}_\mu f,$$

alors, en réalité,

$$\text{Var}_\mu P_t f \leq e^{-\lambda t} \text{Var}_\mu f.$$

Démonstration. Dans le cas réversible on dispose des représentations spectrales suivantes :

$$L = - \int_{\sigma(L)} \lambda dE_\lambda \quad P_t = \int_{\sigma(L)} e^{-\lambda t} dE_\lambda,$$

où E_λ est une résolution de l'identité (cf [81] p. 341, et le théorème 13.37 p. 360). Si l'on note $\lambda_0 = \inf(\sigma(-L) \setminus \{0\})$ le trou spectral de L , on en déduit que

$$\|P_t\|_{L^2(\mu) \setminus 1^\perp} = e^{-\lambda_0 t}$$

(en effet si f est une fonction propre de L associée à 0, $\text{Var}_\mu f = \text{Var}_\mu P_t f \leq C e^{-\lambda t} \text{Var}_\mu f$ implique que $\text{Var}_\mu f = 0$, autrement dit f est nécessairement une constante). Ceci, avec l'hypothèse de l'existence de C et λ , implique que $\lambda \leq \lambda_0$, et finalement que

$$\text{Var}_\mu P_t f \leq e^{-\lambda_0 t} \text{Var}_\mu f \leq e^{-\lambda t} \text{Var}_\mu f.$$

□

Toujours dans le cas réversible, Cattiaux, Guillin et Zitt ont montré dans [24] l'équivalence, pour les diffusions hypoelliptiques, entre l'inégalité de Poincaré, l'existence d'une fonction de Lyapunov (au sens probabiliste), l'intégrabilité des temps d'atteinte de domaines bornés et la convergence exponentiellement rapide en variation totale; les trois dernières notions apparaissent dans une approche à la Meyn et Tweedie ([72]) et la construction de couplages. Dans la mesure où ce domaine est vaste et où mon travail ne s'y est pas aventuré, j'invite le lecteur curieux à consulter cette dernière référence.

1.3.2 L'entropie et log-Sobolev

Avant d'adapter les raisonnements précédents au cas de l'entropie, rappelons la notion suivante :

Définition 1.6. *On dit que $(X_t)_{t \geq 0}$ est une diffusion si pour toute fonction $\theta \in \mathcal{C}^\infty(\mathbb{R}^d, \mathbb{R})$ et toutes $f = (f_1, \dots, f_d) \in \mathcal{A}^d$, si $\theta(f)$ est dans le domaine de L ,*

$$L(\theta(f)) = \sum_i (\partial_i \theta(f)) Lf_i + \sum_{ij} (\partial_i \partial_j \theta(f)) \Gamma(f_i, f_j).$$

Cela revient à dire que L est un opérateur différentiel de degré 2 sans terme constant (au moins dans toute carte, dans le cas d'une variété différentielle), ou encore que la forme de Dirichlet $\mathcal{E}(f)$ correspondante est locale (cf. [3]). On dira sans distinction que $(X_t)_{t \geq 0}$ est une diffusion, que $(P_t)_{t \geq 0}$ est un semi-groupe de diffusion ou que L est un générateur de diffusion.

Définition 1.7. *On dit que μ satisfait une inégalité de log-Sobolev (ou de Gross) de constante c pour Γ si*

$$\text{Ent}_{\mu,1} f \leq 2c \mathcal{E}_\mu(f). \quad (1.4)$$

Les arguments étant peu ou prou les mêmes que pour la variance, on rappelle juste ce résultat sans sa démonstration :

Théorème 1.8 (2.6.3 de [5]).

— Si μ est réversible pour $(P_t)_{t \geq 0}$, l'inégalité (1.4) implique pour tout $f \in \mathcal{A}$ la décroissance

$$\text{Ent}_{\mu,1}(P_t f) \leq e^{-\frac{2t}{c}} \text{Ent}_{\mu,1} f. \quad (1.5)$$

— Si μ est invariante et que $(P_t)_{t \geq 0}$ est une diffusion, l'inégalité (1.4) et la décroissance (1.5) pour tout $f \in \mathcal{A}$ sont équivalentes.

1.3.3 Autour des inégalités fonctionnelles

Pour qui n'a jamais entendu parlé d'inégalités fonctionnelles, les deux sections précédentes, présentées comme une méthode pour établir une vitesse explicite de convergence à l'équilibre, peuvent laisser sur sa faim, voire passer pour de la poudre aux yeux. En effet, qu'avons-nous accompli dans le cas de la variance? On s'est intéressé à une grandeur $d_f(t)$ dépendant d'une fonction f , on l'a dérivée par rapport au temps et on a décrété que $-\frac{1}{2}d'_f(t)$ s'appelait une énergie. Ensuite on a invoqué un grand mathématicien du passé pour donner son nom à l'hypothèse suivante : *on suppose qu'il existe c , indépendant de f , tel que $d_f(t) \leq -\frac{c}{2}d'_f(t)$* . Et l'on a conclu par le lemme de Gromwall. Autrement dit on s'est contenté de mettre toutes les difficultés sous le tapis, ou plutôt sous la question délicate de savoir si, oui ou non, une inégalité de Poincaré est satisfaite, et si oui avec quelle constante (car rappelons-le, nous cherchons des résultats explicites et dans l'idéal précis).

Certes, on a uniquement déplacé le problème. Mais il faut bien voir que le monde des inégalités fonctionnelles n'est pas du tout cantonné à cette question du taux de décroissance de la variance, et que c'est au contraire un cadre riche dont les outils vont nous servir. Nous allons présenter dans ce qui suit certains des résultats les plus classiques qui nous seront utiles par la suite. La préséance est une nouvelle fois donnée à la variance et à l'inégalité de Poincaré, ce qui constitue un cas simple en comparaison à la myriade de sophistications diverses qui ont été développées. La raison en est la suivante : cette étude du cas "classique" n'a qu'un but illustratif et servira surtout à souligner les difficultés inhérentes aux cas hypocoercifs ; difficultés déjà suffisamment intéressantes dans le seul cas (simple, on l'a dit) de la variance. Le chapitre 4 sera une courte exception à ce principe, où des inégalités de Beckner seront considérées.

Théorème 1.9 (3.2.1 et 3.2.2 de [5]). *Si μ_1 et μ_2 satisfont l'inégalité de Poincaré (1.3) pour des constantes et carrés du champs respectifs c_1, c_2, Γ_1 et Γ_2 , alors la mesure produit $\mu_1 \otimes \mu_2$ satisfait une inégalité de Poincaré de constante $\max(c_1, c_2)$ pour $\Gamma_1 + \Gamma_2$.*

Le même résultat a cours en remplaçant l'inégalité de Poincaré (1.3) par l'inégalité de log-Sobolev (1.4).

Cette propriété de tensorisation n'est pas anodine. En particulier elle n'est pas satisfaite par les inégalités de Sobolev classiques de la forme, pour $p > q$,

$$\forall f \in W_m^q(\mu) \quad \|f\|_p \leq c \|f\|_{W_m^q(\mu)}.$$

Les inégalités usuelles correspondent à $\Gamma f = |\nabla f|^2$. Dans la suite de cette section on se restreindra à ce cas. En d'autres termes, on dira que μ satisfait une inégalité de Poincaré de constante c si

$$\|f - \mu f\|_2^2 \leq c \|\nabla f\|_2^2. \quad \mathcal{P}(c)$$

On voit immédiatement que l'image par une fonction k -lipschitzienne d'une loi satisfaisant $\mathcal{P}(c)$ satisfait $\mathcal{P}(k^2 c)$:

$$\begin{aligned} \mathbb{E}_\mu [f^2(g(Z))] - (\mathbb{E}_\mu [f(g(Z))])^2 &\leq c \mathbb{E}_\mu [|\nabla(f \circ g)|^2(Z)] \\ &\leq ck^2 \mathbb{E}_\mu [|\nabla f|^2(g(Z))]. \end{aligned}$$

Sur la droite réelle les travaux de Muckenhoupt mènent à la caractérisation suivante :

Théorème 1.10 (6.2.2 de [5]). Soit ν une mesure de probabilité sur \mathbb{R} admettant une densité g strictement positive sur un intervalle $I = (a, b)$, et soit m la médiane de ν , définie par $\nu([-\infty, m]) = 0.5$. On pose

$$B_\alpha = \max \left(\sup_{x \in (\alpha, b)} \left(\int_x^b g(t) dt \int_\alpha^x \frac{1}{g(t)} dt \right), \sup_{x \in (a, \alpha)} \left(\int_a^x g(t) dt \int_x^\alpha \frac{1}{g(t)} dt \right) \right).$$

Alors ν satisfait $\mathcal{P}(c)$ si et seulement si $B_m < \infty$.

De plus dans ce cas, la constante c optimale satisfait

$$\frac{1}{2} \inf_{\alpha \in I} B_\alpha \leq \frac{1}{2} B_m \leq c \leq 4 \inf_{\alpha \in I} B_\alpha \leq 4 B_m.$$

Un critère équivalent existe pour l'inégalité de log-Sobolev sur \mathbb{R} (théorème 6.3.4 de [5]). Ceci permet de démontrer le critère suivant :

Théorème 1.11 (6.4.3 de [5]). soit U de classe C^2 sur \mathbb{R} et μ la mesure de Gibbs associée. Supposons qu'en-dehors d'un compact $|U'| > 0$, et que $U''(U')^{-2} \rightarrow 0$ à l'infini. Alors :

1. La loi μ satisfait une inégalité de Poincaré si et seulement si $|U'|$ est minorée par une constante strictement positive en-dehors d'un compact.
2. La loi μ satisfait une inégalité de log Sobolev si et seulement si, en-dehors d'un compact,

$$\frac{U}{(U')^2} + \frac{\log |U'|}{(U')^2}$$

est bornée.

En particulier si $U(x) = |x|^\alpha$ pour $\alpha > 0$, μ satisfait une inégalité de Poincaré si et seulement si $\alpha \geq 1$ et de log-Sobolev si $\alpha \geq 2$. On connaît les constantes optimales de Poincaré pour $\alpha = 1$ et $\alpha = 2$:

Théorème 1.12 (1.5.1 et 6.4.1 de [5]).

- La mesure gaussienne standard sur \mathbb{R}^d , $\mathcal{N}(0, I)$, de densité $(2\pi)^{-d/2} e^{-\frac{|x|^2}{2}}$ satisfait $\mathcal{P}(1)$.
- La loi de Laplace sur \mathbb{R} de densité $\frac{1}{2} e^{-|t|}$ satisfait $\mathcal{P}(4)$.

En particulier la loi exponentielle de paramètre λ , en tant qu'image de la loi de Laplace par la fonction $x \mapsto \lambda^{-1}|x|$, satisfait $\mathcal{P}(4\lambda^{-2})$.

On peut obtenir, à partir de ces exemples de bases, de nombreuses inégalités de Poincaré par perturbation. Si ν est une loi de probabilité, on appelle ν_ψ , définie par

$$\nu_\psi f = \frac{1}{\nu(e^{-\psi})} \nu [f e^{-\psi}],$$

la perturbation de ν par le potentiel ψ . On rappelle les résultats suivants, qui seront utiles dans la suite de notre étude :

Théorème 1.13 (3.1.18 de [80]). Soit ν satisfaisant $\mathcal{P}(c)$ et ψ bornée. Alors ν_ψ satisfait $\mathcal{P}(c')$ avec

$$c' = e^{(\max \psi - \min \psi)} c.$$

Théorème 1.14 (2.7 de [2]). *Soit ν satisfaisant $\mathcal{P}(c)$ et ψ telle que $\|\nabla\psi\|_\infty^2 = c^{-1}\eta$ avec $\eta < \frac{5-\sqrt{17}}{5}$. Alors ν_ψ satisfait $\mathcal{P}(c')$ avec*

$$c' = 2c \left(\frac{1-\eta}{2-\eta} - 2\eta \right)^{-1}$$

Quand il s'agit de démontrer des inégalités fonctionnelles, le défaut principal de l'approche perturbative est de fournir des résultats loins d'être précis. Par exemple si l'on voit la mesure ν sur \mathbb{R} de densité $e^{-\frac{x^2}{2}-x}$ (renormalisée) comme une perturbation de la gaussienne standard, les résultats de perturbation vont donner pour ν une inégalité pour une constante moins bonne que 1, alors que la densité de ν est de la forme $e^{-U(x)}dx$ avec $U''(x) \geq 1$ et satisfait donc une inégalité de Poincaré de constante 1 d'après la proposition 1.15 ci-après, appliquée au générateur $L = -\nabla U \cdot \nabla + \Delta$.

Dans le chapitre 4, on démontrera un résultat perturbatif plus adapté à la situation que les théorèmes 1.13 et 1.14, mais qui aura le même défaut quant à la précision.

1.3.4 Le critère de Bakry-Emery

Rappelons maintenant un critère qui fait intervenir des inégalités de Poincaré locales et des commutations semi-groupe/gradient. On considère un semi-groupe $(P_t)_{t \geq 0}$ de générateur L et de carré du champ Γ , et l'on pose

$$\Gamma_2 f = \frac{1}{2} L \Gamma f - \Gamma(f, Lf),$$

et $\Gamma_2(f, g)$ la forme bilinéaire associée obtenue par polarisation. Si $\rho \in \mathbb{R}$, on dit que L vérifie le critère de courbure ρ (ou critère Γ_2 de constante ρ) si pour tout $f \in \mathcal{A}$

$$\Gamma_2 f \geq \rho \Gamma f.$$

Proposition 1.15. *Sont équivalentes les trois assertions suivantes :*

1. L satisfait le critère de courbure ρ
2. $\forall f \in \mathcal{A}, t > 0, \Gamma(P_t f) \leq e^{-2\rho t} P_t(\Gamma f)$
3. $\forall f \in \mathcal{A}, t > 0, P_t(f^2) - (P_t f)^2 \leq \frac{1-e^{-2\rho t}}{\rho} P_t(\Gamma f)$
4. $\forall f \in \mathcal{A}, t > 0, P_t(f^2) - (P_t f)^2 \geq \frac{e^{2\rho t}-1}{\rho} \Gamma(P_t f)$.

Avant d'en donner quelques éléments de démonstration, remarquons ceci. La troisième assertion est une inégalité de Poincaré uniformément vérifiée par les mesures $\delta_x P_t$, d'où son nom de locale (idem pour la quatrième, l'inégalité de Poincaré inversée). Si P_t est ergodique et $\rho > 0$, on obtient en faisant tendre t vers l'infini une inégalité de Poincaré de constante ρ^{-1} pour la mesure invariante.

Démonstration. Remarquons que, lorsque Q est une forme quadratique, en posant

$$\psi(s) = P_s [Q(P_{t-s} f)],$$

on calcule

$$\begin{aligned} \psi'(s) &= P_s L [Q(P_{t-s} f)] - 2P_s [Q(P_{t-s} f, LP_{t-s} f)] \\ &= 2P_s \left[\left(\frac{1}{2} L Q(\cdot) - Q(\cdot, L \cdot) \right) P_{t-s} f \right] \end{aligned}$$

En particulier pour $Qf = \Gamma f$, en supposant le critère Γ_2 satisfait,

$$\psi'(s) = 2P_s\Gamma_2P_{t-s}f \geq 2\rho P_s\Gamma P_{t-s}f = 2\rho\psi(s).$$

Ceci montre que (1) entraîne (2). Maintenant en supposant (2) et en considérant $Qf = f^2$, on obtient

$$\begin{aligned} P_t(f^2) - (P_t f)^2 &= \int_0^t \psi'(s) ds \\ &= 2 \int_0^t P_s \Gamma (P_{t-s} f) ds \\ &\leq 2 \int_0^t e^{-2\rho(t-s)} P_s P_{t-s} \Gamma f ds \\ &= \frac{1 - e^{-2\rho t}}{\rho} P_t \Gamma f. \end{aligned}$$

Ce qui montre que (2) implique (3); et par le même calcul mais en utilisant

$$P_s \Gamma (P_{t-s} f) \geq e^{2\rho s} \Gamma (P_s P_{t-s} f)$$

on montre que (2) implique (4). Enfin, pour montrer que (3) ou (4) implique (1), on peut faire un développement de Taylor à l'ordre 2 des deux termes de l'inégalité de Poincaré locale (cf. [5] p.82). \square

Dans le cas des diffusions, des raisonnements semblables fournissent des résultats analogues pour l'entropie plutôt que la variance (cf. [5] p. 83) et même pour toutes les Φ -entropies (cf. [17]).

Un exemple typique d'application : considérons le générateur sur \mathbb{R}^d

$$Lf = -\nabla U \cdot \nabla f + \Delta f, \tag{1.6}$$

de mesure invariante e^{-U} . En notant H_g la hessienne de la fonction g et $\|A\|$ la norme de Hilbert-Schmidt de la matrice A , c'est-à-dire

$$\|A\|^2 = \sum_{i,j=1}^d A_{i,j}^2,$$

on calcule

$$\begin{aligned} \Gamma f &= |\nabla f|^2 \\ \Gamma_2 f &= \|H_f\|^2 + (\nabla f)^T H_U \nabla f. \end{aligned}$$

En supposant que la courbure de U (c'est-à-dire les valeurs propres de H_U) est minorée par $\rho \in \mathbb{R}$, on voit que L satisfait le critère de courbure ρ .

1.3.5 Le recuit simulé réversible

Le processus de Fokker-Planck $(X_t)_{t \geq 0}$, dont le générateur est donné par (1.6), a pour mesure invariante la mesure de Gibbs associée à un potentiel U . Pour le simuler, il s'agit

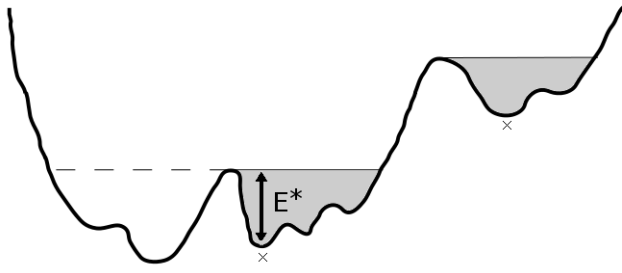


FIGURE 1.1 – La profondeur d’un minimum local, et la profondeur critique.

simplement de savoir estimer en chaque temps le gradient de U en X_t . Un candidat possible pour un algorithme de recuit simulé (continu) est donc le processus de générateur

$$L_\varepsilon f(x) = -\nabla U(x) \cdot \nabla f(x) + \varepsilon \Delta f(x), \quad (1.7)$$

ou plutôt le processus inhomogène de générateur $L_t = L_{\varepsilon_t}$ lorsque $t \mapsto \varepsilon_t > 0$ est un schéma de température à préciser. On note μ_ε la loi de Gibbs associée à U à température ε .

Hypothèse 1.16. *On suppose que U est C^2 , que U et $|\nabla U|$ vont à l’infini à l’infini, et que*

$$\inf (|\nabla U|^2 - \Delta U) > -\infty.$$

Sous cette hypothèse, le spectre de L_ε est discret ([78, p.120]), on dispose donc d’un trou spectral $\lambda_\varepsilon > 0$ (on peut aussi invoquer le théorème 1.11). Chiang, Hwang, et Sheu démontrent dans [28], par des méthodes de grandes déviations à la Freidlin et Wentzel, que $-\varepsilon \log \lambda_\varepsilon$ converge quand ε tend vers 0 vers une constante positive E^* .

Cette constante est décrite comme suit : on dit qu’on peut atteindre x à partir de y au niveau d’énergie E s’il existe un chemin ϕ continu de x à y avec $U(\phi(s)) \leq E$ pour tout $s \in [0, 1]$. La profondeur d’un minimum local non global x_0 est le plus petit niveau d’énergie E tel qu’un point z avec $U(z) < U(x_0)$ est atteignable depuis x_0 à niveau $U(x_0) + E$. Enfin la profondeur critique E^* est la profondeur maximale parmi tous les minima locaux non globaux de U (cf. figure 1.1) .

Les auteurs en déduisent un premier résultat de convergence, ensuite amélioré par Royer (cf. [79]) pour aboutir à ceci :

Théorème 1.17 (1.1 de [79]). *Si $\varepsilon_t = \frac{c}{\ln t}$ pour $t \geq t_0 > 1$ avec $c > E^*$, alors la convergence suivante a lieu : pour tout $x \in \mathbb{R}$,*

$$d_{vt}(\delta_x P_{t_0, t}, \mu_{\varepsilon_t}) \xrightarrow{t \rightarrow \infty} 0$$

En particulier ceci implique que la probabilité que le processus se trouve dans un voisinage des minima globaux de U tend vers 1.

Un peu plus tard, Holley, Kusuoka et Stroock proposent dans [57, 58] une méthode plus analytique, basée sur les inégalités fonctionnelles, qui va plus particulièrement nous intéresser. Le cadre en est un peu restreint, au sens où l’espace d’état est une variété compacte. Avant de la présenter (sommairement) précisons que, si cette première méthode est apparentée à la méthode pour démontrer la décroissance exponentielle de la variance à partir d’une inégalité

de Poincaré, un analogue a été développé par Miclo ([73], sous l'hypothèse 1.16) qui permet de sortir du cadre compact, plutôt basé sur la décroissance exponentielle de l'entropie (ou énergie libre) à partir d'une inégalité de log-Sobolev (voir aussi le travail de Zitt [88] sur les potentiels à croissance lente à l'infini).

Dans le cadre compact les auteurs montrent une asymptotique plus précise du trou spectral de $\tilde{L}_\varepsilon = \varepsilon^{-1}L_\varepsilon$ que le principe de grande déviation précédent :

Lemme 1.18 (1.14 de [58]). *Il existe des constantes $c, C > 0$ (dépendant éventuellement de U mais pas de ε) telles que*

$$ce^{-\frac{E^*}{\varepsilon}} \leq \tilde{\lambda}_\varepsilon \leq Ce^{-\frac{E^*}{\varepsilon}}.$$

Supposons que $m_0 = \mathcal{L}(X_0)$ soit absolument continue par rapport à μ_{ε_0} , de densité f_0 . Dans ce cas (lemme 1.6 de [57]) $m_t = m_0 P_{0,t}$ est absolument continue par rapport à μ_{ε_t} , et sa densité f_t satisfait

$$\partial_t f_t = \tilde{L}_t f_t + \left(\frac{1}{\varepsilon_t}\right)' (U - \mu_{\varepsilon_t}[U]) f_t.$$

En conséquence,

$$\begin{aligned} \partial_t \|f_t - 1\|_{L^2(\mu_{\varepsilon_t})}^2 &= -2\mathcal{E}_{\mu_{\varepsilon_t}}(f_t) + 2\left(\frac{1}{\varepsilon_t}\right)' \langle (U - \mu_{\varepsilon_t}[U]) f_t, f_t - 1 \rangle_{L^2(\mu_{\varepsilon_t})} \\ &= -2\mathcal{E}_{\mu_{\varepsilon_t}}(f_t) + 2\left(\frac{1}{\varepsilon_t}\right)' \int (U - \mu_{\varepsilon_t}[U]) \left((f_t - 1)^2 + (f_t - 1) \right) d\mu_{\varepsilon_t} \\ &\leq \left(-2ce^{-\frac{E^*}{\varepsilon_t}} + 4\left(\frac{1}{\varepsilon_t}\right)' \|U\|_\infty \right) \|f_t - 1\|_{L^2(\mu_{\varepsilon_t})}^2 + 4\left(\frac{1}{\varepsilon_t}\right)' \|U\|_\infty \|f_t - 1\|_{L^2(\mu_{\varepsilon_t})} \end{aligned}$$

Pour que cette inégalité différentielle fournisse un contrôle de $\|f_t - 1\|_{L^2(\mu_{\varepsilon_t})}^2$, il faut que $\left(\frac{1}{\varepsilon_t}\right)'$ soit au plus de l'ordre de $e^{-\frac{E^*}{\varepsilon_t}}$. En supposant que $\varepsilon_t = \frac{E^*}{\log(1+\gamma t)}$, $u(t) = \|f_t - 1\|_{L^2(\mu_{\varepsilon_t})}^2$ satisfait l'inégalité différentielle

$$u'(t) \leq \frac{1}{1+\gamma t} \left[\left(-2c + \frac{4\gamma\|U\|_\infty}{E^*} \right) u(t) + \frac{4\gamma\|U\|_\infty}{E^*} \sqrt{u(t)} \right].$$

Ceci fournit (pour γ assez petit) un contrôle uniforme en temps de $\|f_t - 1\|_{L^2(\mu_{\varepsilon_t})}^2$; ce qui n'est pas suffisant pour conclure, mais nous a permis d'illustrer quelques unes des idées qui permettent de démontrer le résultat suivant :

Théorème 1.19 (2.7 de [58]). *Supposons qu'il existe une constante $\kappa \geq E^*$ telle*

$$\left(\frac{1}{\varepsilon_t}\right)' \leq \kappa^{-1}(1+t). \tag{1.8}$$

Alors il existe $A > 0$ et $p > 2$ tels que pour tout $f \in \mathcal{A}$ et pour tout $t > 0$,

$$\|f - \mu_{\varepsilon_t} f\|_{L^p(\mu_{\varepsilon_t})} \leq A(1+t)\mathcal{E}_{\mu_{\varepsilon_t}}(f). \tag{1.9}$$

Et d'autre part

Théorème 1.20 (2.2 de [58]). *S'il existe $A, \kappa > 0$ et $p > 2$ tels que (1.8) et (1.9), alors il existe $T, K > 0$ tels que pour tout $t \geq T$ et pour toute loi initiale $dm_0 = f_0 d\mu_{\varepsilon_0}$, la densité $f_t = \frac{d(m_0 P_{0,t})}{d\mu_{\varepsilon_t}}$ est bornée par K . En conséquence il existe une constante B telle que pour tout $\delta > 0$,*

$$\mathbb{P}(U(X_t) > \min U + \delta) \leq B \frac{(\log(1+t))^d}{\delta(1+t)^{\frac{\delta}{\kappa}}}$$

où d est la dimension.

Notez que la conjonction de ces deux résultats donne une décroissance de la probabilité d'erreur plus lente que $(1+t)^{-\frac{\delta}{E^*}}$, ce qui n'est pas très bon ; une probabilité faible d'erreur demandera de longs temps de calcul.

La borne supérieure sur le trou spectral permet d'autre part de montrer que cette condition sur le schéma de température est optimale. Supposons que $\varepsilon_t = \frac{c}{\ln(1+t)}$ avec $c < E^*$. On peut considérer un minimum local non global x_0 de profondeur E_0 strictement supérieure à c , et C_{x_0} l'ensemble des points qui sont atteignables depuis x_0 à niveau $U(x_0) + E_0$. Dans ce cas, conséquence des lemmes 3.5 et 3.7 de [58]

$$\inf_{x \in C_{x_0}} \mathbb{P}_x(X_t \in C_{x_0} \text{ pour tout } t > 0) > 0$$

qui implique

$$\inf_{x \in C_{x_0}} \mathbb{P}_x(U(X_t) \geq U(x_0) \text{ pour tout } t > 0) > 0.$$

Une étude plus poussée des schémas rapides (ou sous-admissibles) de température est faite dans [74], par des méthodes de semi-groupes.

Ce résultat appelle un certain nombre de remarques. D'abord il faut en souligner le caractère théorique : la constante E^* est en pratique inconnue, impossible de l'utiliser pour effectivement prescrire un schéma de température. Quand bien même on aurait une idée de l'ordre de grandeur de E^* , un schéma de l'ordre de $E^*/\log(1+t)$, on l'a dit, demanderait des temps de calculs trop longs pour être implémenté ; on utilise plutôt des schémas géométriques, de la forme $t^{-\alpha}$. Dès lors l'optimalité de la vitesse $E^*/\log(1+t)$ s'interprète ainsi : les performances des schémas pratiques sont aléatoires. C'est donc un résultat négatif. Il est également théorique au sens où il considère un résultat asymptotique en temps infini (du fait que la constante B par exemple du théorème 1.20 est inconnue) ; les temps d'exécutions, eux, seront finis.

Ces différentes critiques trouvent une réponse dans le travail de Catoni (cf. [23], au moins pour les chaînes discrètes), qui montre qu'à horizon de temps fini, il existe des schémas géométriques pertinents (en un sens différent, évidemment, du théorème 1.20 ou du théorème 1.17). Et néanmoins cela ne change pas le fait que la convergence du processus est indubitablement lente, à cause des difficultés du processus à sortir des régions métastables à basse température. Parmi les différentes réponses envisagées en préambule à ce constat, la dynamique réversible de (1.1) peut être remplacée par une dynamique cinétique dont on espère que l'inertie accélère la sortie des bassins d'attraction. Ainsi dans [83] est-il considéré l'équation de Langevin cinétique,

$$\begin{aligned} dX_t &= Y_t dt \\ dY_t &= -Y_t dt - U'(X_t) dt + \sqrt{2} dB_t, \end{aligned}$$

de générateur

$$L = y\partial_x - (y + U'(x))\partial_y + \partial_y^2$$

et de mesure invariante $\mu = e^{-U(x)-y^2} dx dy$. L'étude empirique de [83] indique que la simulation de la mesure de Gibbs associée à U *via* le processus cinétique est plus efficace qu'avec le processus de Fokker-Planck ou la marche simple associée à une étape d'acceptation/rejet (l'algorithme de Metropolis-Hastings le plus classique).

On pourrait croire qu'il s'agit dès lors de suivre pour ce processus la méthode classique présentée le long de ce chapitre pour obtenir un taux de convergence du processus à température fixé, et un critère de convergence sur les schémas de température; d'autant que μ est la tensorisation de la mesure de Gibbs associée à U et d'une gaussienne standard, on connaît donc l'asymptotique de son trou spectral à basse température. Tout le travail semble avoir déjà été fait.

... ce qui nous amène à la section suivante.

1.4 L'hypocoercivité

1.4.1 L'exemple de Langevin cinétique

On considère sur \mathbb{R} un potentiel U sympathique (mettons : régulier et quadratique à l'infini), et le générateur

$$L = y\partial_x - (y + U'(x))\partial_y + \partial_y^2, \quad (1.10)$$

ergodique de mesure invariante $\mu = e^{-U(x)-y^2} dx dy$. Dans ce cas $\Gamma f = (\partial_y f)^2$ (la dérive déterministe de la dynamique, c'est-à-dire l'opérateur différentiel d'ordre 1 du générateur, n'apparaît jamais dans Γ ; et ici l'aléa ne porte que sur la variable de vitesse y). En particulier si $f \in \mathcal{A}$, sans être constante, ne dépend que de x , $\Gamma f = 0$ et

$$\partial_t \left(\|P_t f - \mu f\|_2^2 \right)_{|t=0} = -2\mu \Gamma f = 0.$$

Impossible de trouver une constante $c > 0$ telle que

$$\|f - \mu f\|_2^2 \leq c\mu \Gamma f,$$

autrement dit μ ne satisfait pas d'inégalité de Poincaré pour Γ . Notons qu'on réserve l'écriture $\mathcal{E}_\mu f$ pour l'énergie classique, c'est-à-dire $\mathcal{E}_\mu f = \mu |\nabla f|^2$ (pour laquelle ici, μ satisfait bel et bien une inégalité de Poincaré).

Quant au critère Γ_2 , on calcule

$$\Gamma_2 f = \left(\partial_y^2 f \right)^2 + (\partial_y f)^2 - \partial_y f \partial_x f,$$

qui ne contrôle pas Γf du fait du dernier terme. Autrement dit la courbure de Bakry-Emery du processus est $-\infty$.

D'après le théorème 1.4, en l'absence d'inégalité de Poincaré, il ne peut pas exister de taux $\lambda > 0$ tel que pour tout $f \in \mathcal{A}$,

$$\|P_t f - \mu f\|_2^2 \leq e^{-\lambda t} \|f - \mu f\|_2^2.$$

Plutôt que la norme de $L^2(\mu)$, une idée usuelle en analyse serait alors de considérer le problème dans $\mathcal{H}^1(\mu)$,

$$\|g\|_{\mathcal{H}^1}^2 = \|g\|_2^2 + \|\nabla g\|_2^2 = \|g\|_2^2 + \|\partial_x g\|_2^2 + \|\partial_y g\|_2^2.$$

On calcule donc, en notant $f_t = P_t f - \mu f$,

$$\begin{aligned} \partial_t \left(\|\partial_y f_t\|_2^2 \right) &= -2\|\partial_y^2 f_t\|_2^2 - 2\|\partial_y f_t\|_2^2 + 2 \langle \partial_x f_t, \partial_y f_t \rangle \\ \partial_t \left(\|\partial_x f_t\|_2^2 \right) &= -2\|\partial_x \partial_y f_t\|_2^2 - 2 \langle U'' \partial_x f_t, \partial_y f_t \rangle \\ \Rightarrow \partial_t \left(\|f_t\|_{\mathcal{H}^1}^2 \right) &= -2\|\partial_y^2 f_t\|_2^2 - 4\|\partial_y f_t\|_2^2 - 2\|\partial_x \partial_y f_t\|_2^2 \\ &\quad + 2 \langle \partial_x f_t, \partial_y f_t \rangle - 2 \langle U'' \partial_x f_t, \partial_y f_t \rangle . \end{aligned}$$

Là encore tout disparaît si f_t ne dépend que de x , ce qui implique pour les mêmes raisons que pour la norme 2 qu'une décroissance exponentielle

$$\|f_t\|_{\mathcal{H}^1} \leq e^{-\lambda t} \|f_0\|_{\mathcal{H}^1}$$

est impossible. Mais si maintenant on dérive le terme croisé $\langle \partial_x f_t, \partial_y f_t \rangle$, on obtient

$$\partial_t \left(\langle \partial_x f_t, \partial_y f_t \rangle \right) = -2 \langle \partial_x \partial_y f_t, \partial_y^2 f_t \rangle + \|\partial_x f_t\|_2^2 - \langle \partial_x f_t, \partial_y f_t \rangle - \langle U'' \partial_y f_t, \partial_y f_t \rangle$$

Si désormais f_t ne dépend que de x , il reste malgré tout le terme $\|\partial_x f_t\|_2^2$. On pose donc, pour des paramètres $a, b, c > 0$,

$$|g|^2 = \|g\|_2^2 + a\|\partial_x g\|_2^2 + b\|\partial_y g\|_2^2 - 2c \langle \partial_x g, \partial_y g \rangle .$$

Tant que $c < \sqrt{ab}$, c'est une norme hilbertienne sur $\mathcal{H}^1(\mu)$, équivalente à la norme usuelle, et en supposant toujours U quadratique à l'infini,

$$\begin{aligned} \partial_t \left(|f_t|^2 \right) &= -2 \left(a\|\partial_x \partial_y f_t\|_2^2 + b\|\partial_y^2 f_t\|_2^2 - 2c \langle \partial_x \partial_y f_t, \partial_y^2 f_t \rangle \right) \\ &\quad - (2 + 2b - 2c\|U''\|_\infty) \|\partial_y f_t\|_2^2 - c\|\partial_x f_t\|_2^2 + 2 \langle (b - aU'' + c) \partial_x f_t, \partial_y f_t \rangle \\ &\leq \left(-2 - 2b + 2c + \frac{\|b - aU'' + c\|_\infty^2}{\gamma} \right) \|\partial_y f_t\|_2^2 + (\gamma - c) \|\partial_x f_t\|_2^2 \end{aligned}$$

pour tout $\gamma > 0$. Ne serait-ce qu'en posant $a = b = \gamma = c/2 = \varepsilon$ et en faisant tendre ε vers 0 (mais, comme dans la suite, tout pourrait être fait explicitement), on en déduit l'existence de $\kappa > 0$ tel que

$$\partial_t \left(|f_t|^2 \right) \leq -\kappa \|\nabla f_t\|_2^2.$$

On invoque alors le fait que μ satisfait une inégalité de Poincaré *pour le gradient complet*, c'est-à-dire qu'il existe une constante l telle que

$$\|f_t\|_2^2 \leq l \|\nabla f_t\|_2^2,$$

pour déduire

$$\partial_t \left(|f_t|^2 \right) \leq -\frac{\kappa}{2} \|\nabla f_t\|_2^2 - \frac{\kappa}{2l} \|f_t\|_2^2.$$

Par équivalence de $|g|$ et $\|g\|_{\mathcal{H}^1}$, on dispose ainsi de $\lambda > 0$ tel que

$$\partial_t (|f_t|^2) \leq -\lambda |f_t|^2.$$

Ce qui, avec à nouveau l'équivalence des normes, entraîne :

$$\|f_t\|_{\mathcal{H}^1} \leq c_1 |f_t| \leq c_1 e^{-\lambda t} |f_0| \leq c_1 c_2 e^{-\lambda t} \|f_0\|_{\mathcal{H}^1}.$$

Ceci nous permet enfin d'introduire la notion phare du titre de cette thèse : supposons qu'une fonctionnelle de Lyapunov d_t mesure en un sens (norme d'opérateur, distance de Wasserstein, etc.) la distance entre P_t et μ .

Définition 1.21. *Si d_t converge vers 0 quand t tend vers l'infini, on dira que cette convergence est hypocoercive s'il existe $\lambda > 0$ et $C \geq 1$ tels que*

$$d_t \leq C e^{-\lambda t} d_0,$$

et strictement hypocoercive si une telle inégalité ne peut avoir lieu pour $C = 1$.

Avant de donner quelques perspectives “historiques” (le terme est peut-être un peu fort pour un domaine dont on considère que les premiers résultats datent des années 2000) sur l'hypocoercivité, ajoutons une dernière considération à notre exemple introductif. À partir de la convergence en norme \mathcal{H}^1 , on peut obtenir un résultat similaire pour la norme L^2 , grâce à des estimées de régularité hypoelliptiques. La diffusion de Fokker-Planck n'est pas elliptique puisque le brownien n'intervient que sur la variable de vitesse (la matrice de diffusion n'est donc pas inversible) mais elle satisfait le critère d'Hörmander ([59, 52]) sur les crochets de Lie et admet donc à tout instant positif une densité strictement positive et régulière par rapport à la mesure de Lebesgue. On peut en fait montrer une estimation quantitative de la forme suivante (cf. [86, théorème A.8, p.143]) : il existe une constante C telle que pour tout $f \in \mathcal{A}$,

$$\|\nabla P_t f\|_2^2 \leq \frac{C}{t^{\frac{3}{2}}} \|f\|_2^2.$$

Ainsi grâce à cette inégalité de Poincaré inversée, pour $t \geq 1$ (en notant abusivement C des constantes successives)

$$\begin{aligned} \|P_t f - \mu f\|_2^2 &\leq \|P_t f - \mu f\|_{\mathcal{H}^1}^2 \\ &\leq C e^{-\lambda(t-1)} \|P_1 f - \mu f\|_{\mathcal{H}^1}^2 \\ &\leq C e^{-\lambda t} \|f - \mu f\|_2^2 \end{aligned}$$

L'existence d'un C tel qu'une telle inégalité est également vraie pour $t < 1$, on a donc bien une convergence hypocoercive de la norme 2.

1.4.2 Quelques considérations globales

Le texte fondateur de l'hypocoercivité, ne serait-ce que parce que le terme lui-même y apparaît pour la première fois, est du à Villani ([85], en 2006) ; on y trouve notamment une vue d'ensemble des résultats précurseurs du domaine. La méthode de modification de la norme \mathcal{H}^1 présentée ci-dessus en est tirée. Avant elle, d'autres démonstrations de la convergence

exponentielle dans le cas du processus de Langevin cinétique avaient été données : une étude plutôt probabiliste de Talay ([84]) ; le travail de Hérau et Nier ([56, 54]), par l'analyse spectrale d'opérateurs hypoelliptiques ; la méthode de Desvillettes et Villani ([33, 32]), proche de la méthode classique de la section 1.3, si ce n'est qu'au lieu de montrer une inégalité différentielle d'ordre 1 pour la fonctionnelle de Lyapunov, un système d'inégalités différentielles d'ordre 2 est établi.

La méthode de fonctionnelle de Lyapunov modifiée est ensuite améliorée par Dolbeault, Mouhot et Schmeiser ([38, 39, 22]). À la différence de la méthode de Villani, ces derniers n'ajoutent à la norme de L^2 que des termes d'ordre 0, c'est-à-dire qui ne font pas intervenir de dérivées de $P_t f$, mais simplement un opérateur A construit à partir de projections, continu sur L^2 . De la sorte, $\|f\| + \varepsilon \langle Af, f \rangle$ est équivalent, pour ε assez petit au moins, à la norme 2, et l'on n'a pas recours à des estimées hypoelliptiques.

Dans d'autres directions, à partir de la méthode de Villani, on peut noter le lien esquissé par Bakry, Cattiaux et Guillin [9] avec les fonctions de Lyapunov (au sens probabiliste) ; et d'autre part l'approche de Baudoin [11] qui est à la méthode de Villani ce que le critère Γ_2 est à la méthode de la section 1.3, aboutissant à des inégalités locales (voir aussi le travail de Kontis, Ottobre et Zegarlinski dans la même veine, [60, 61]). Enfin, dans le très récent [6], Arnold et Erb modifient, pour le processus de Langevin cinétique quadratique, non pas la variance mais l'entropie (ce que Villani faisait également dans son mémoire [86], dans un cadre plus général mais de façon moins précise), ce qui fournit, dans la vision duale de convergence des densités, une convergence en norme L^1 plutôt que L^2 .

Des définitions plus générales ou plus vagues sont parfois données de l'hypocoercivité, comme "l'étude du rôle de la partie non-dissipative du générateur dans le processus de dissipation" (dans [85]). La définition 1.21 a pour elle le mérite de la simplicité. On peut faire à son sujet quelques remarques générales.

Soit $(P_t)_{t \geq 0}$ un semi-groupe ergodique de loi invariante μ . Considérons un ensemble E de probabilité stable par P_t et contenant μ , et une fonctionnelle de Lyapunov (au sens analyste) ρ définie sur E , c'est-à-dire, de façon abstraite, une fonction $\rho : E \rightarrow [0, \infty]$ nulle en μ et strictement positive (éventuellement infinie) ailleurs. Supposons qu'il existe une fonction $t \mapsto C(t)$, telle que pour tout $\nu \in E$

$$\rho(\nu P_t) \leq C(t)\rho(\nu).$$

Notons \mathfrak{C} l'ensemble des fonctions telles qu'une telle inégalité a lieu, et $C_{opt}(t)$ l'infimum sur \mathfrak{C} . Supposons qu'il existe E' dense dans E (pour une topologie pour laquelle ρ et P_t soient continues) telle que $\partial_s(\rho(\nu P_s))(0)$ existe pour tout $\nu \in E'$. De manière élémentaire les assertions suivantes sont équivalentes :

1. Il existe $\lambda > 0$ tel que $C_{opt}(t) \leq e^{-\lambda t}$.
2. Il existe $\lambda > 0$ et $C \in \mathfrak{C}$ tels que $C(t) \leq e^{-\lambda t}$.
3. Il existe $\lambda > 0$ tel que pour tout $\nu \in E'$, $\partial_s(\rho(\nu P_s))(0) \leq -\lambda\rho(\nu)$
4. Il existe $C \in \mathfrak{C}$ dérivable en 0 avec $C'(0) < 0$.

Ces propriétés décrivent ce que l'on pourrait appeler le cas coercif. Notons qu'il existe des exemples où les fonctions de \mathfrak{C} sont toutes discontinues en 0 (cf. [8]). Notons que par la

propriété de semi-groupe, $C_{opt}(t+s) \leq C_{opt}(t)C_{opt}(s)$. Si $C \in \mathfrak{C}$ est régulière et satisfait cette propriété, on peut toujours écrire, pour tout $N \geq 1$,

$$C(t) \leq \left(C\left(\frac{t}{N}\right) \right)^N \xrightarrow{N \rightarrow \infty} e^{C'(0)t}.$$

Simplement si $C'(0) = 0$, cela ne donne rien. Mais, en l'absence de coercivité à $t = 0$, on peut se contenter d'une contraction à un temps donné :

Proposition 1.22. *Supposons qu'il existe une fonction $C \in \mathfrak{C}$ localement bornée et un temps $t_* > 0$ tel que $C(t_*) < 1$. Alors il existe des constantes κ et λ telles que $\forall t \geq 0$,*

$$\rho(\nu P_t) \leq \kappa e^{-\lambda t} \rho(\nu),$$

autrement dit la convergence est hypocoercive.

Démonstration. On décompose $t = kt_* + s$ avec $k \in \mathbb{N}$ et $s \in [0, t_*)$, et l'on pose $\tilde{C}(t) = C(t_*)^k C(s)$. Par composition on a bien

$$\rho(\nu P_t) \leq \tilde{C}(t) \rho(\nu).$$

Ensuite

$$\begin{aligned} \tilde{C}(t) &= C(t_*)^{\frac{kt_*+s}{t_*}} \frac{C(s)}{C(t_*)^{\frac{s}{t_*}}} \\ &\leq e^{-\frac{|\log C(t_*)|}{t_*} t} \sup_{s \leq t_*} \left\{ \frac{C(s)}{C(t_*)^{\frac{s}{t_*}}} \right\}. \end{aligned}$$

□

Ceci dit en majorant par $\kappa e^{-\lambda t}$ on perd toute information sur les temps petits ou sur d'éventuelles oscillations, et si le C optimal est de la forme $g(t)e^{-\lambda t}$ avec un préfacteur $g(t)$ polynomial, on ne peut qu'obtenir une majoration du type $\kappa_\varepsilon e^{-(\lambda-\varepsilon)t}$ avec une constante κ_ε qui explose.

Dernière remarque générale : d'après la proposition 1.5, la convergence en norme 2 d'un processus réversible ne peut jamais être strictement hypocoercive.

1.4.3 Conclusion sur les motivations de la thèse

Hypocoercivité : approches alternatives et applications aux algorithmes stochastiques ; c'est l'intitulé de cette thèse. Nous venons d'en définir le premier terme, et nous avons déjà survolé la question des applications en présentant sommairement le recuit simulé. Reste à justifier la volonté de trouver des méthodes alternatives. Dans la section précédente, après tout, on démontre la convergence exponentiellement rapide, à température fixée, du processus de Langevin cinétique vers la mesure de Gibbs associée, à un taux explicite dont l'on pourrait se contenter. Cependant, on constate numériquement que ce taux explicite obtenu n'est pas très précis par rapport au taux optimal, ce qui est particulièrement gênant lorsque l'on veut considérer le comportement de ce taux explicite à basse température.

Cette mauvaise précision, dans la méthode de la modification de la fonctionnelle de Lyapunov que nous avons présentée, tient pour beaucoup des majorations relativement grossières

effectuées. Il est naturel de se demander s'il n'y a pas une fonctionnelle particulière à considérer, ou même un système particulier de fonctionnelles liées entre elles par des inégalités différentielles (comme par exemple dans [33, 32]), dont l'évolution au cours du temps fournirait une bonne estimation (au moins à température basse) du taux de convergence. C'est en tout cas une question suffisamment naturelle pour que le processus de Langevin cinétique, cobaye privilégié, ait subi ces dernières années, et continue à subir, un certain nombre de traitements en ce sens, même dans le cas d'un potentiel quadratique ([33, 6, 11, 38], etc.).

Plutôt que de prendre comme point de départ les méthodes existantes et de tenter de les perturber dans l'espoir d'en améliorer les performances, cette thèse a pris le parti d'en premier lieu étudier très en détail un modèle très simple. La compréhension de la mécanique de la convergence à l'équilibre dans ce cas devait permettre d'isoler certaines idées exploitables dans un cadre plus général. Cette étude est l'objet du chapitre suivant.

Chapitre 2

Modèle jouet hypocoercif

*Les systèmes, comme les constitutions, sont les jouets
avec lesquels s'amuse les personnes graves.*

Jacques Bainville

(Note : ce chapitre est une adaptation de l'article [75].)

2.1 Motivations et résultat principal

Considérons la marche symétrique sur le tore discret $\mathbb{Z}/N\mathbb{Z} = \mathbb{Z}_N$. Elle est réversible pour la mesure uniforme, irréductible et - pour peu que N soit impair - apériodique, et donc ergodique. Cette dynamique est l'ingrédient de base d'un algorithme de Metropolis-Hasting réversible. N'ayant aucune mémoire, elle revient fréquemment sur ses pas.

L'idée développée entre autres par Diaconis dans [35] puis [36] est que l'on devrait pouvoir accélérer la convergence à l'équilibre en pénalisant les retours en arrière. Concrètement, cela revient à connaître à tout instant la position courante X_n mais également la position antérieure X_{n-1} . La suite $(X_k)_{k \geq 0}$ est ainsi une chaîne de Markov d'ordre 2, au sens où

$$\mathcal{L}(X_{n+1} \mid X_0, \dots, X_n) = \mathcal{L}(X_{n+1} \mid X_{n-1}, X_n).$$

En d'autres termes, le couple $(X_{n-1}, X_n)_{n \geq 0}$ est une chaîne de Markov, dont on définit les transitions comme suit : si $X_n = x$ et $X_{n-1} = x - y$ avec $y \in \{-1, 1\}$, alors

$$\begin{aligned} \mathbb{P}(X_{n+1} = x + y) &= \frac{1 + \alpha}{2} \\ \mathbb{P}(X_{n+1} = x - y) &= \frac{1 - \alpha}{2} \end{aligned}$$

avec un paramètre $\alpha \in (0, 1)$. Notons que la vitesse $Y_{n-1} = X_n - X_{n-1}$ est markovienne en elle-même, à valeurs dans $\{-1, 1\}$, de transitions

$$\begin{aligned} \mathbb{P}(Y_{n+1} = Y_n) &= \frac{1 + \alpha}{2} \\ \mathbb{P}(Y_{n+1} = -Y_n) &= \frac{1 - \alpha}{2}. \end{aligned}$$

Lorsque N est impair, la chaîne $(X_n, Y_n)_{n \geq 0}$ définie par ces transitions et par $X_{n+1} = X_n + Y_n$, est irréductible, apériodique, et donc ergodique. Sa loi invariante est simplement

$$\mu_N = \mathcal{U}_{\mathbb{Z}_N} \times \frac{\delta_{-1} + \delta_1}{2}.$$

Diaconis et Miclo prouvent dans [36], pour un cadre légèrement plus général, qu'il existe un paramètre $\alpha > 0$ qui maximise le trou spectral de cette marche (qui est donc plus grand que celui de la marche réversible, correspondant à $\alpha = 0$). Notons

$$M_\alpha f(x, y) = \mathbb{E}(f(X_1, Y_1) | X_0 = x, Y_0 = y)$$

la matrice de transition de la chaîne, $\sigma(M_\alpha)$ son spectre, et

$$1 - \lambda_\alpha = 1 - \sup(|\sigma(M_\alpha) \setminus \{1\}|)$$

son trou spectral. Ici, dans le cas d'un nombre fini d'états, quelque soit la norme $\|\cdot\|$ considérée sur $L^2(\mu_N) = \mathbb{R}^N$, on sait que

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \log(\|M_\alpha^n - \mu_N\|) = \log(\lambda_\alpha).$$

Autrement dit la vitesse asymptotique de la convergence à l'équilibre peut bel et bien être améliorée avec un paramètre $\alpha > 0$. Dans ce cas particulier, les calculs étant explicites, on montre que ce taux exponentiel de convergence $\log(\lambda_\alpha)$ est de valeur absolue maximale (et donc la vitesse asymptotique est la meilleure) pour $\alpha_{opt} = \frac{1 - \sin(\pi/N)}{1 + \sin(\pi/N)}$ pour lequel $\lambda_{opt} = \sqrt{\alpha_{opt}}$ (cf. [36]). En comparaison, pour la marche isotrope ($\alpha = 0$) on a $\lambda_0 = \cos(\pi/N)$; et l'on a bien

$$\cos(\pi/N) = \sqrt{(1 - \sin(\pi/N))(1 + \sin(\pi/N))} \geq \sqrt{\frac{1 - \sin(\pi/N)}{1 + \sin(\pi/N)}}.$$

Cependant cette information asymptotique est relativement faible. D'autre part le lien entre le trou spectral et le taux asymptotique de convergence n'est, en toute rigueur, pas aussi limpide pour un espace d'états infini, lorsque le processus n'est pas réversible. On va donc calculer explicitement la distance en norme 2 entre M_α^n et μ_N ou plutôt, afin de clarifier les calculs en se débarrassant d'artefacts liés au monde discret, entre leurs équivalents continus, que nous allons maintenant définir.

Notons $\mathbb{T} = \mathbb{R}/(2\pi\mathbb{Z})$ le tore. Pour tout $N \geq 1$ on peut considérer la marche persistante $(X_n^N, Y_n^N)_{n \geq 0}$ sur le tore discret, pour un paramètre α_N donné. Plutôt que $X_n^N \in \mathbb{Z}_N$, intéressons-nous plutôt à $U_t^N = \frac{2\pi}{N} X_n^N \in \mathbb{T}$ si $t = n \frac{2\pi}{N}$ que l'on prolonge de façon affine à $t \geq 0$ et $V_t^N = Y_n^N$ si $t \in \frac{2\pi}{N}[n, n+1[$. Ainsi à tout instant t , la dérivée à droite de U_t^N satisfait, quelque soit N ,

$$\lim_{s \rightarrow 0} \frac{U_{t+s}^N - U_t^N}{s} = V_t^N.$$

D'autre part le temps entre deux sauts de la vitesse Y^N suit une loi géométrique de paramètre $\frac{1 - \alpha_N}{2}$; si ce paramètre est de l'ordre de $\frac{1}{N}$, il converge donc en loi vers une variable exponentielle. Plus précisément, si la suite $(\alpha_N)_{N \geq 1}$ est telle que

$$\frac{N}{2\pi} \times \frac{1 - \alpha_N}{2} \xrightarrow{N \rightarrow +\infty} a > 0,$$

alors quand N tend vers l'infini, V_t^N converge en loi (en tant que processus) vers $Y_t = (-1)^{N_t}$ où $(N_t)_{t \geq 0}$ est un processus de Poisson d'intensité a . Le processus Y est appelé processus du télégraphe. Ainsi U_t^N converge vers $X_t = \int_0^t Y_s ds$; on l'appelle parfois le processus du télégraphe intégré, ou circulaire.

On s'intéresse donc au processus Markovien $(X_t, Y_t)_{t \geq 0}$. Il est ergodique de mesure invariante $\mu = \mathcal{U}_{\mathbb{T}} \times \frac{\delta_1 + \delta_{-1}}{2}$, la mesure uniforme sur $\mathbb{T} \times \{-1, 1\}$. Il n'est pas réversible mais sa dynamique est plutôt simple : X_t avance à vitesse 1 dans un sens et, à des temps de loi $\mathcal{E}(a)$, fait volte-face, et reprend sa route en sens inverse, toujours à vitesse 1. Notons $(P_t^a)_{t \geq 0}$ le semi-groupe associé sur $L^2(\mu)$ (on omettra le paramètre a en l'absence d'ambiguïté). Son générateur, bien défini sur les fonctions régulières de $L^2(\mu)$, est

$$\mathcal{L}_a f(x, y) = y \partial_x f(x, y) + a(f(x, -y) - f(x, y)).$$

Pour quantifier la distance entre P_t et μ on s'intéressera à la norme 2 d'opérateur, dont on rappelle la définition :

$$\|P_t - \mu\| = \sup \left\{ \|P_t f - \mu f\|_{L^2(\mu)}, \|f\|_{L^2(\mu)} = 1 \right\}.$$

Comme ce serait le cas pour des opérateurs de dimension finie, on suspecte que

$$\lim_{t \rightarrow +\infty} \frac{1}{t} \log \|P_t - \mu\| = -\lambda \quad (2.1)$$

avec

$$\lambda = \inf \{ -\Re(\theta), \theta \text{ valeur propre de } \mathcal{L}_a \} \quad (2.2)$$

On va vérifier que ceci est juste, mais on cherche surtout des résultats plus quantitatifs, en estimant précisément la norme $\|P_t - \mu\|$ en tout temps $t \geq 0$, car en pratique des renseignements asymptotiques tels que (2.1) ne sont pas très exploitables. Voilà l'essentiel des résultats obtenus (illustrés par les figures 2.1, 2.2 et 2.3) sur ce modèle :

Théorème 2.1. *Pour $a \geq 1$, on a $\lambda = a - \sqrt{a^2 - 1}$ et pour $a \leq 1$, $\lambda = a$. Plus précisément, pour tout $t > 0$,*

— *Si $a > 1$ alors, en notant $\omega = \sqrt{a^2 - 1}$ et $\gamma = e^{-2\omega t}$,*

$$\begin{aligned} \|P_t - \mu\| &= e^{(-a + \sqrt{a^2 - 1})t} \sqrt{1 + \frac{2}{\omega^2 \left(\frac{1+\gamma}{1-\gamma}\right) + a \sqrt{1 + \omega^2 \left(\frac{1+\gamma}{1-\gamma}\right)^2} - 1}} \\ &= 1 - \frac{t^3}{3} + o_{t \rightarrow 0}(t^3) \\ &\underset{t \rightarrow +\infty}{\sim} \frac{a^2}{a^2 - 1} e^{\lambda t} \end{aligned}$$

— *Si $a = 1$ alors*

$$\begin{aligned} \|P_t - \mu\| &= e^{-t} \sqrt{1 + \frac{2}{\sqrt{1 + \frac{1}{t^2}} - 1}} \\ &= 1 - \frac{t^3}{3} + o_{t \rightarrow 0}(t^3) \\ &\underset{t \rightarrow +\infty}{\sim} 2te^{-t} \end{aligned}$$

— Si $a < 1$ alors

$$\begin{aligned}\|P_t - \mu\| &= e^{-at} \sqrt{g(t)} \\ &= 1 - \frac{at^3}{3} + o_{t \rightarrow 0}(t^3)\end{aligned}$$

avec g telle que

$$\begin{aligned}\limsup_{t \rightarrow +\infty} g(t) &= \frac{1+a}{1-a} \\ \liminf_{t \rightarrow +\infty} g(t) &= 1\end{aligned}$$

et, en notant $\nu = 2\sqrt{1-a^2}$, si $t \in [0, \frac{\pi}{\nu}]$ alors

$$g(t) = \left(1 + \frac{2}{\sqrt{\frac{\nu^2}{a^2} \frac{1}{2(1-\cos(\nu t))} + 1} - 1} \right)$$

Ce résultat est analogue à celui de l'étude [47] de Gadat et Miclo sur des diffusions cinétiques (l'équation de Langevin cinétique pour un potentiel quadratique, et le processus d'Ornstein-Uhlenbeck intégré sur le tore). On retrouve en particulier la décroissance initiale en t^3 , le fait que le choix optimal de a (au sens : qui donne la plus grande vitesse asymptotique de convergence) voit la constante devant le terme exponentiel exploser, et que ce paramètre sépare deux régimes, dont l'un où le spectre est complexe (plus précisément, à part 0, plus aucune valeur propre n'est réelle) et des oscillations apparaissent.

Le théorème 2.1 sera démontré au cours de la partie 2.2. La partie 2.3 s'attache au lien entre le modèle discret de la marche persistante et son analogue continu du télégraphe intégré. Lorsque la fréquence de changement de vitesse devient grande ce processus continu tend vers le mouvement brownien, ce qui est étudié en partie 2.4. La partie 2.5 quant à elle discute des généralisations de ces premiers résultats à des potentiels quelconque et à la dimension supérieure. Enfin, l'appendice regroupe quelques lemmes techniques utilisés dans le reste du chapitre.

2.2 Calcul exact de la norme

Remarquons une fois encore que si le processus était réversible, le travail serait simple puisque \mathcal{L} serait diagonalisable en base orthonormée (dans $L^2(\mu)$). Ce n'est pas le cas ici mais on va tout de même pouvoir décomposer l'espace en plans stables orthogonaux ce qui nous ramènera à calculer des normes d'opérateurs en dimension 2, qu'il faudra ensuite comparer entre elles.

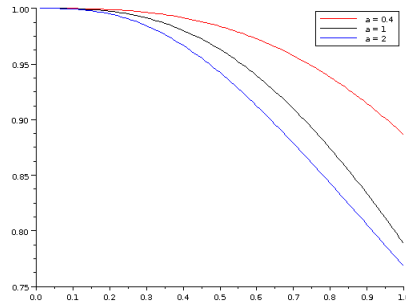


FIGURE 2.1 – Norme du semi-groupe pour différentes valeurs de a au cours du temps (ici $t \in [0, 1]$). Au début la décroissance est d’autant plus rapide que a est grand.

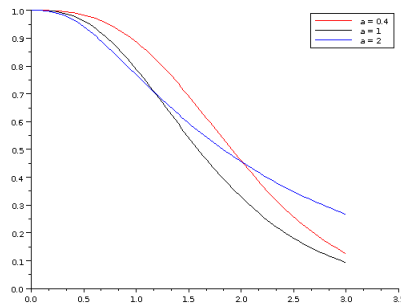


FIGURE 2.2 – Cependant la tendance finit par s’inverser (ici $t \in [0, 3]$)

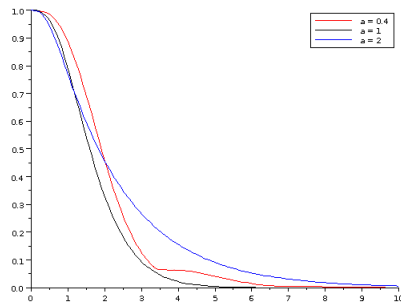


FIGURE 2.3 – La meilleure vitesse asymptotique est obtenue pour $a = 1$. Pour $a < 1$ il arrive que la dérivée de la norme s’annule presque (ici $t \in [0, 5]$)

Lemme 2.2. Les plans $V_n = \{f : (x, y) \mapsto e^{inx}g(y), g \in \mathbb{C}^{\{-1,1\}}\}$, pour $n \in \mathbb{Z}$, sont invariants par \mathcal{L} , orthogonaux et totaux dans $L^2(\mu)$. L’action de P_t sur V_n est donnée par e^{tK_n} où

$$K_n g(y) = iny g(y) + a(g(-y) - g(y))$$

Démonstration. L’orthogonalité et le caractère total découlent directement de ceux de $(x \mapsto e^{inx})_{n \in \mathbb{N}}$ dans $L^2(\mathbb{T})$. On s’assure ensuite directement que pour $f(x, y) = e^{inx}g(y)$ on a bel et bien $\mathcal{L}f(x, y) = e^{inx}K_n g(y)$. \square

On est donc ramené à calculer la norme d’une matrice 2×2 . Notons

$$R(t, a, n) \stackrel{def}{=} \|P_t^a - \mu\|_{V_n}^2.$$

Notons que pour tout $n \neq 0$ on a $V_n \subset Ker(\mu)$. Le cas $n = 0$ est un peu à part et facile à régler : K_0 est diagonalisable avec deux valeurs propres, 0 (associées aux constantes, que l'on retranche ici) et $-2a$. Ainsi

$$R(t, a, 0) = e^{-4at}.$$

Cette restriction ne réalisera en fait jamais la norme globale (sauf $t = 0$ bien sûr) : en effet on va voir que, quelque soit a , \mathcal{L} possède des valeurs propres de parties réelles $-a$; ainsi sur une droite propre pour une telle valeur propre $\|P_t\| = e^{-at} > e^{-2at}$. D'autre part $K_n = \bar{K}_{-n}$ et on se restreindra donc dans la suite à $n > 0$. Finalement,

$$\|P_t - \mu\| = \sup_{n \geq 1} \left(\|P_t\|_{V_n} \right) = \sup_{n \geq 1} \left(\sqrt{R(t, a, n)} \right)$$

Calcul des normes des restrictions

Lemme 2.3. *Si $a > n$ alors pour tout $t > 0$*

$$R(t, a, n) = e^{-2(a-\sqrt{a^2-n^2})t} \times \left(1 + \frac{2}{\omega^2 \left(\frac{1+\gamma}{1-\gamma}\right) + \frac{a}{n} \sqrt{1 + \omega^2 \left(\frac{1+\gamma}{1-\gamma}\right)^2 - 1}} \right)$$

avec $\omega = \sqrt{\left(\frac{a}{n}\right)^2 - 1}$ et $\gamma = e^{-2\sqrt{a^2-n^2}t}$.

Démonstration. Les deux valeurs propres de K_n , réelles, sont $\lambda_1 = -a + n\omega > \lambda_2 = -a - n\omega$. On calcule que (e_1, e_2) sont des vecteurs propres correspondants unitaires ils vérifient $|\langle e_1, e_2 \rangle| = \frac{n}{a}$ (les vecteurs propres sont « d'autant plus orthogonaux » que a est loin de n), on peut donc choisir (e_1, e_2) unitaires tels que $\langle e_1, e_2 \rangle = \frac{n}{a}$. En posant $u = re^{i\theta}e_1 + e_2$ on a ainsi

$$\begin{aligned} e^{tK_n}u &= re^{i\theta}e^{\lambda_1 t}e_1 + e^{\lambda_2 t}e_2 \\ \|u\|^2 &= r^2 + 1 + 2r\frac{n}{a}\cos(\theta) \\ \|e^{tK_n - \lambda_1 t}u\|^2 &= r^2 + \gamma^2 + 2r\gamma\frac{n}{a}\cos(\theta) \\ &= \|u\|^2 + (\gamma - 1) \times \left[\gamma + 1 + 2r\frac{n}{a}\cos(\theta) \right]. \end{aligned}$$

En conséquence

$$\begin{aligned} \frac{\|e^{tK_n - \lambda_1 t}u\|^2}{\|u\|^2} &= \frac{r^2 + \gamma^2 + 2r\gamma\frac{n}{a}\cos(\theta)}{r^2 + 1 + 2r\frac{n}{a}\cos(\theta)} \\ &= \gamma + \frac{r^2 + \gamma^2 - \gamma r^2 - \gamma}{r^2 + 1 + 2r\frac{n}{a}\cos(\theta)} \end{aligned}$$

quantité qui, à r fixé, est monotone en $\cos(\theta)$. Les valeurs extrémales sont donc obtenues avec $\cos(\theta) = 1$ (quitte à prendre $r < 0$). On a alors

$$\begin{aligned}\frac{\|e^{tK_n - \lambda_1 t} u\|^2}{\|u\|^2} &= 1 + (\gamma - 1) \times \frac{\gamma + 1 + 2r \frac{n}{a}}{r^2 + 1 + 2r \frac{n}{a}} \\ &= 1 - 2 \frac{n}{a} (1 - \gamma) \times \frac{(r + \frac{n}{a}) - \frac{n}{a} + \frac{a}{2n} (1 + \gamma)}{(r + \frac{n}{a})^2 + 1 - (\frac{n}{a})^2}.\end{aligned}$$

D'après le lemme 2.15, les valeurs extrémales sont

$$\frac{\|e^{tK_n - \lambda_1 t} u\|^2}{\|u\|^2} = 1 - \frac{\left(\frac{n}{a}\right)^2 (1 - \gamma)}{\left(\frac{n}{a}\right)^2 - \left(\frac{1+\gamma}{2}\right) \pm \sqrt{\left(\frac{1+\gamma}{2}\right)^2 - \gamma \left(\frac{n}{a}\right)^2}}.$$

Le maximum est obtenu pour $\pm = -$, et l'on obtient

$$\begin{aligned}\|e^{tK_n - \lambda_1 t}\|^2 &= 1 + \frac{\left(\frac{n}{a}\right)^2 (1 - \gamma)}{\left(\frac{1+\gamma}{2}\right) - \left(\frac{n}{a}\right)^2 + \sqrt{\left(\frac{1+\gamma}{2}\right)^2 - \gamma \left(\frac{n}{a}\right)^2}} \\ &= 1 + \frac{2}{\omega^2 \left(\frac{1+\gamma}{1-\gamma}\right) + \frac{a}{n} \sqrt{1 + \omega^2 \left(\frac{1+\gamma}{1-\gamma}\right)^2} - 1}.\end{aligned}$$

□

Lemme 2.4. Si $a < n$ alors pour tout $t > 0$

$$R(t, a, n) = e^{-2at} \times \left(1 + \frac{2}{\sqrt{\frac{\nu_n^2}{a^2} \frac{1}{2(1-\cos(\nu_n t))} + 1} - 1} \right)$$

avec $\nu_n = 2\sqrt{n^2 - a^2}$.

Démonstration. Dans ce cas les valeurs propres de K_n sont complexes conjuguées, $\lambda_1 = \bar{\lambda}_2 = \lambda = -a + i\sqrt{n^2 - a^2}$, de partie réelle a . On trouve des vecteurs propres normés associés e_1 et e_2 vérifiant $\langle e_1, e_2 \rangle = \frac{a}{n}$ (là encore le produit scalaire des vecteurs propres tend vers 0 à mesure que a et n s'éloignent).

Posons $u = e_1 + r e^{i\theta} e_2$ avec $r \in \mathbb{R}$ et $\theta \in]-\pi, \pi]$. On a alors $e^{tK_n} u = e^{\lambda t} (e_1 + r e^{i\theta} e^{-2i\sqrt{n^2 - a^2}} e_2)$, et ainsi

$$\begin{aligned}\|u\|^2 &= r^2 + 1 + 2r \frac{a}{n} \cos(\theta) \\ \|e^{tK_n - t\lambda} u\|^2 &= r^2 + 1 + 2r \frac{a}{n} \cos(\theta - 2t\sqrt{n^2 - a^2})\end{aligned}$$

Par le lemme 2.15 on obtient que le rapport entre les deux est extrémal pour $r = \pm 1$, on est donc ramené à

$$\|e^{tK_n - t\lambda}\|^2 = \sup_{\theta \in \mathbb{T}} \frac{\alpha_n + \cos(\theta - \nu_n t)}{\alpha_n + \cos(\theta)}$$

avec $\alpha_n = \frac{n}{a} > 1$. Le lemme 2.16 de l'appendice conclut. \square

Lemme 2.5. *Si $a = n$ alors pour tout $t > 0$*

$$R(t, a, n) = e^{-2at} \times \left(1 + \frac{2}{\sqrt{1 + \frac{1}{n^2 t^2} - 1}} \right)$$

Démonstration. Dans ce cas $-n$ est valeur propre double de K_n . Considérons la base $g_1(y) = 1 + iy$ et $g_2(y) = \frac{1}{n}$ de $\mathbb{C}^{\{-1,1\}}$. La matrice de K_n dans cette base est alors un bloc de Jordan, d'exponentielle $e^{-nt} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$. En renormalisant g_1 et g_2 , on obtient des vecteurs de base unitaires e_1 et e_2 avec $\langle e_1, e_2 \rangle = \frac{1}{\sqrt{2}}$, $e^{tK_n} e_1 = e^{-nt} e_1$ et $e^{tK_n} e_2 = e^{-nt} (e_2 + \sqrt{2} n t e_1)$. En posant $u = (x + iy)e_1 + e_2$, on a ainsi

$$\begin{aligned} e^{tK_n} u &= e^{-nt} (u + \sqrt{2} n t e_1) \\ \|u\|^2 &= x^2 + y^2 + 1 + \sqrt{2} x \\ \|e^{tK_n + nt} u\|^2 &= \|u\|^2 + 2n^2 t^2 + 2\sqrt{2} n t \left(x + \frac{1}{\sqrt{2}} \right) \end{aligned}$$

Le rapport $\frac{\|e^{tK_n + nt} u\|^2}{\|u\|^2}$ est donc optimal pour $y = 0$. Reste à choisir x .

$$\frac{\|e^{tK_n + nt} u\|^2}{\|u\|^2} = 1 + 2\sqrt{2} n t \times \frac{x + \frac{1}{\sqrt{2}} + \frac{nt}{\sqrt{2}}}{\left(x + \frac{1}{\sqrt{2}}\right)^2 + \frac{1}{2}}$$

D'après le lemme 2.15, les valeurs extrêmes sont

$$\frac{\|e^{tK_n + nt} u\|^2}{\|u\|^2} = 1 + \sqrt{2} n t \times \frac{1}{-\frac{nt}{\sqrt{2}} \pm \sqrt{\frac{n^2 t^2}{2} + \frac{1}{2}}}$$

et le maximum est obtenu pour $\pm = +$, ce qui donne le résultat escompté. \square

Remarquons qu'on aurait pu obtenir ce résultat par continuité à partir des cas $a \leq n$.

Comparaison des $R(t, a, n)$

Il s'agit maintenant de comparer les normes de ces restrictions entre elles. Un développement limité en $t = 0$ montre que $R(t, a, n) = 1 - \frac{n^3}{3} t^3 + o(t^3)$ pour $a \geq n$ et $R(t, a, n) = 1 - \frac{an^2}{3} t^3 + o(t^3)$ pour $a \leq n$, ce qui laisse penser qu'au moins au début $R(t, a, 1)$ prévaut (autrement dit que l'erreur décroît lentement sur V_1 les fonctions de grande longueur d'onde en x). D'autre part, si $a > 1$, c'est aussi sur V_1 que se trouve la droite propre associée à la

valeur propre de \mathcal{L} de plus grande partie réelle, c'est donc également $R(t, a, 1)$ qui devrait prévaloir asymptotiquement. En fait nous allons voir que, pour l'essentiel, seule compte cette norme sur V_1 . Notons que les expressions calculés pour $R(t, a, n)$ permettent d'étendre leur définition à n non entier et qu'alors $n \in]0, +\infty[\mapsto R(t, a, n)$ est continue.

Dans un premier temps, on peut dériver $R(t, a, n)$ pour $n \in]0, a[$. Le lemme 2.18 de l'annexe montre que cette dérivée est négative et ainsi $\max_{1 \leq n < a} R(t, a, n) = R(t, a, 1)$ pour tout $t > 0$. Par continuité on a même $\max_{1 \leq n \leq a} R(t, a, n) = R(t, a, 1)$. Ainsi a-t-on réglé les cas $a \geq 1$ du théorème 2.1 puisqu'alors $\|P_t - \mu\| = \max_{n \in \mathbb{Z}^*} R(t, a, n) = R(t, a, 1)$.

Le cas des $n > a$ est un peu plus délicat, pour qui

$$R(t, a, n) = e^{-ta} \sqrt{g_n(t)}$$

avec, si $\nu_n = 2\sqrt{n^2 - a^2}$,

$$g_n(t) = 1 + \frac{2}{\sqrt{\frac{\nu_n^2}{a^2} \frac{1}{2(1-\cos(\nu_n t))} + 1} - 1}$$

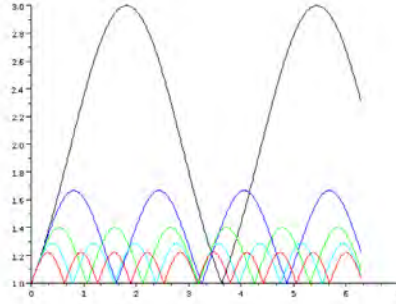


FIGURE 2.4 – Plus l est grand plus l'amplitude et la longueur d'onde de g_l sont faibles.

qui est $2\pi/\nu_n$ périodique. Calculer le supremum des g_n pour tout t est à peu près impossible du fait des périodes incommensurables (cf. figure 2.4) ; c'est un phénomène de quasi-périodicité. Cependant on peut penser (d'après le développement limité en 0) qu'en temps petit la norme prépondérante correspond à n minimal et qu'elle le reste jusqu'à ce que g_n atteigne son maximum. C'est effectivement le cas, comme on va le montrer dans un instant. Ensuite le sup des g_k oscillera entre ce maximum et 1.

Lemme 2.6. *Si $k < n$ alors pour tout $t \in [0, \frac{\pi}{\nu_k}]$ on a $g_k(t) \geq g_n(t)$.*

Démonstration.

$$\begin{aligned} g_n(t) \leq g_k(t) &\Leftrightarrow 1 + \frac{2}{\sqrt{\frac{\nu_n^2}{a^2} \frac{1}{2(1-\cos(\nu_n t))} + 1} - 1} \leq 1 + \frac{2}{\sqrt{\frac{\nu_k^2}{a^2} \frac{1}{2(1-\cos(\nu_k t))} + 1} - 1} \\ &\Leftrightarrow \frac{1 - \cos(\nu_n t)}{\nu_n^2} \leq \frac{1 - \cos(\nu_k t)}{\nu_k^2} \end{aligned}$$

Ces deux termes sont égaux et de dérivées égales en $t = 0$, pour les comparer il suffit donc de comparer leurs dérivées secondes. Or, si $\nu_n \geq \nu_k$ alors $\cos(\nu_n t) \leq \cos(\nu_k t)$ pour $t \in \left[0, \frac{\pi}{\nu_n}\right]$, et donc $g_n(t) \leq g_k(t)$ pour ces t . Puisque g_k est croissante sur $\left[0, \frac{\pi}{\nu_k}\right]$ on a pour $t \in \left[\frac{\pi}{\nu_n}, \frac{\pi}{\nu_k}\right]$

$$g_k(t) \geq g_k\left(\frac{\pi}{\nu_n}\right) \geq g_n\left(\frac{\pi}{\nu_n}\right) \geq g_n(t).$$

On achève en constatant que ν_n est croissante en n . \square

Lemme 2.7. *Si $n > a$ alors pour tout $t > 0$ on a $R(t, a, n) \leq R(t, a, a)$.*

Démonstration. D'après le lemme précédent, pour tout $\varepsilon > 0$ on a $R(t, a, n) \leq R(t, a, a + \varepsilon)$ pour $t \leq \frac{\pi}{\nu_{a+\varepsilon}}$; or $\nu_{a+\varepsilon} \xrightarrow{\varepsilon \rightarrow 0} 0$ et la continuité de R conclut. \square

En particulier si $a \geq 1$ pour tout t on aura $\|P_t - \mu\| = R(t, a, 1)$, ce qui démontre les deux tiers de la propriété 2.1. Pour $a < 1$ on peut comparer plus finement les g_n :

Lemme 2.8. *Soit $g(t) = \sup_{n \in \mathbb{N}} g_n(t)$. Si $t \leq \frac{\pi}{\nu_1}$ alors $g(t) = g_1(t)$, et d'autre part*

$$\limsup_{t \rightarrow +\infty} g(t) = \frac{1+a}{1-a} (= \sup g)$$

$$\liminf_{t \rightarrow +\infty} g(t) = 1 (= \inf g)$$

Démonstration. La première assertion a déjà été démontrée, et le résultat pour la limite sup découle directement de la périodicité de g_1 . Pour la limite inf, considérons $\varepsilon > 0$, et soit $N \in \mathbb{N}$ tel que $\frac{1+\frac{\varepsilon}{N}}{1-\frac{\varepsilon}{N}} \leq 1 + \varepsilon$. On a ainsi, pour tout $k \geq N$ et pour tout $t > 0$, $g_k(t) \leq 1 + \varepsilon$. Le lemme 2.17 de l'appendice nous donne un temps où les fonctions restantes (en nombre fini) sont simultanément proches de leur minimum : soit $\delta > 0$ tel que $\forall n < N$ et $\forall k \in \mathbb{Z}$,

$$\left| t - \frac{2k\pi}{\nu_n} \right| \leq \delta \Rightarrow g_n(t) \leq 1 + \varepsilon.$$

Soient $t > 0$ et $k_1, \dots, k_{N-1} \in \mathbb{N}$ tels que $\left| \frac{2\pi}{\nu_n} k_n - t \right| < \delta$ pour tout $n < N$; on obtient que $g_n(t) \leq 1 + \varepsilon$ pour tout $n < N$, et donc pour tout $n \in \mathbb{N}$. Soit ε_0 le minimum sur $]1, t+1]$ de $g - 1$ (fonction continue). Si $\varepsilon_0 = 0$ alors g est périodique et son minimum est sa limite inf. Sinon on peut recommencer l'argument ci-dessus pour obtenir un temps t_2 tel que pour tout $n \in \mathbb{N}$ on ait $g_n(t_2) \leq 1 + \varepsilon_0/2$, donc nécessairement $t_2 > t + 1$; finalement en itérant le procédé on peut trouver des temps arbitrairement grand où g est arbitrairement proche de 1, ce qui conclut. \square

Ce lemme finit de démontrer le théorème 2.1.

2.3 Du discret au continu

L'étude du télégraphe intégré a permis de mieux comprendre la convergence en temps petit, et nous pouvons maintenant faire le lien avec la marche discrète de [36]. D'abord constatons que des calculs identiques aux précédents nous permettent de calculer la norme de M_α^n , $n \in \mathbb{N}$. Pour $k \in \llbracket 1, N \rrbracket$ on notera $e^{2ik\pi/N} = C_k + iS_k$, $\alpha_l = \frac{1-|S_l|}{1+|S_l|}$, $C_0^2 = \frac{4\alpha}{(1+\alpha)^2}$ et $S_0^2 = \left(\frac{1-\alpha}{1+\alpha}\right)^2$.

Lemme 2.9. Les plans $W_k = \{(x, y) \mapsto e^{2ik\pi x/N} g(y), g \in \mathbb{C}^{\pm 1}\}$ sont stables par M . Notons $R_N(n, \alpha, k) \stackrel{\text{def}}{=} \|M_\alpha^n - \mu\|_{W_k}^2$.
— si $\alpha < \alpha_k$ alors

$$R_N(n, \alpha, k) = \lambda_+^{2n} \times \left(1 + \frac{2}{\omega^2 \left(\frac{1+\gamma}{1-\gamma}\right) + \frac{S_0}{S_k} \sqrt{1 + \omega^2 \left(\frac{1+\gamma}{1-\gamma}\right)^2 - 1}} \right)$$

$$\text{avec } \lambda_\pm = \sqrt{\alpha} \left(\frac{|C_k|}{C_0} \pm \sqrt{\left(\frac{C_k}{C_0}\right)^2 - 1} \right), \gamma = \left(\frac{\lambda_-}{\lambda_+}\right)^n \text{ et } \omega^2 = \left(\frac{S_0}{S_k}\right)^2 - 1.$$

si $\alpha > \alpha_k$ alors

$$R_N(n, \alpha, k) = \alpha^n \times \left(1 + \frac{2}{\sqrt{2 \frac{\left(\frac{S_k}{S_0}\right)^2 - 1}{1 - \cos(2n\psi)} + 1} - 1} \right)$$

$$\text{où } \tan \psi = \sqrt{\left(\frac{C_0}{C_k}\right)^2 - 1}.$$

si enfin $\alpha = \alpha_k$ alors

$$R_N(n, \alpha, k) = \alpha^n \times \left(1 + \frac{2}{\sqrt{1 + \frac{C_0^2}{S_0^2 n^2} - 1}} \right)$$

Démonstration. La démarche et les calculs sont quasiment les mêmes que dans le cas continu et n'amènent aucune difficulté nouvelle. \square

Rappelons qu'en introduction, pour passer du modèle discret au continu, plutôt que $X_n^N \in \mathbb{Z}_N$ nous avons regardé $U_t^N = \frac{2\pi}{N} X_n^N \in \mathbb{T}$ si $t = n \frac{2\pi}{N}$, et $V_t^N = Y_n$ si $t \in \frac{2\pi}{N}[n, n+1[$. Remarquons que pour $u = \frac{2\pi}{N}x$ on peut réécrire $e^{i\frac{2k\pi}{N}x} = e^{iku}$, l'espace V_k correspond donc à W_k :

Lemme 2.10. Pour tout $t > 0$ et $k \in \mathbb{Z}$, si $\alpha^{(N)} \in [0, 1]$ est tel que $\frac{N}{2\pi} \times \frac{1-\alpha^{(N)}}{2} \xrightarrow{N \rightarrow +\infty} a$ alors

$$R_N \left(\left\lfloor \frac{Nt}{2\pi} \right\rfloor, \alpha^{(N)}, k \right) \xrightarrow{N \rightarrow +\infty} R(t, a, k).$$

Démonstration. On le vérifie sans difficulté particulière sur les expressions analytique données dans le lemme 2.9 et la partie 2.2. \square

Cependant, contrairement au cas continu, dans la marche discrète la plus grande valeur propre (associée au $|\cos(\frac{2k\pi}{N})|$ maximal) ne correspond pas à $k = 1$ mais à $k = \pm \lfloor \frac{N}{2} \rfloor$. Pour avoir la convergence des normes globales d'opérateurs il faut ignorer les deux plans $W_{\pm \lfloor \frac{N}{2} \rfloor}$. En un sens le caractère fini des positions prises par la particule entraîne l'existence d'observables qui convergent mal, ce qui disparaît à la limite des processus, mais pas dans le passage à la limite des normes.

Lemme 2.11. *Pour tout $t > 0$ et $k \in \mathbb{Z}$, si $\alpha^{(N)} \in [0, 1]$ est tel que $\frac{N}{2\pi} \times \frac{1-\alpha^{(N)}}{2} \xrightarrow{N \rightarrow +\infty} a$ alors*

$$R_N \left(\left\lfloor \frac{Nt}{2\pi} \right\rfloor, \alpha^{(N)}, \left\lfloor \frac{N}{2} \right\rfloor - k \right) \xrightarrow{N \rightarrow +\infty} R \left(t, a, k + \frac{1}{2} \right)$$

Démonstration. Les calculs sont les mêmes que précédemment ; le $1/2$ apparaît avec

$$\sin \left(\frac{2\pi}{N} (\lfloor \frac{N}{2} \rfloor - k) \right) = \sin \left(\pi - \frac{2\pi}{N} (\lfloor \frac{N}{2} \rfloor - k) \right) = \sin \left(\frac{2\pi}{N} (k + \frac{1}{2}) \right)$$

□

Le travail de comparaison des $R(t, a, n)$ englobait déjà les n non-entiers, et en notant

$$P_t^N f(u, v) \stackrel{\text{def}}{=} \mathbb{E}(f(U_t^N, V_t^N) | U_0^N = u, V_0^N = v)$$

on obtient *in fine*

Proposition 2.12. *Si $\frac{N}{2\pi} \times \frac{1-\alpha^{(N)}}{2} \xrightarrow{N \rightarrow +\infty} a \geq \frac{1}{2}$ alors*

$$\|P_t^N - \mu_N\| \xrightarrow{N \rightarrow +\infty} R \left(t, a, \frac{1}{2} \right).$$

D'autre part si l'on note $\mathcal{V}_N = W_{\pm \lfloor \frac{N}{2} \rfloor}^\perp$ et si $a \geq 1$ alors

$$\|P_t^N - \mu_N\|_{\mathcal{V}_N} \xrightarrow{N \rightarrow +\infty} \|P_t^a - \mu\|.$$

Les convergences sont uniformes en t .

Démonstration. Tout est déjà démontré sauf le caractère uniforme en t ; les fonctions en présence étant toutes décroissantes, il découle du théorème de Dini. □

Remarque 2.13. *Notons que, grosso modo, les choses se passent bien également pour $a < \frac{1}{2}$ dans le premier cas et pour $a < 1$ dans le second mais avec de très légères subtilités : par exemple, dans le deuxième cas et pour reprendre les notations de la partie 2.2, la fonction $g(t)$ limite n'est pas le supremum des $g_n(t)$ pour n entier mais pour n entier ou demi-entier, ce qui peut éventuellement légèrement changer la valeur exacte de la norme lors d'un « creux » de $R(t, a, 1)$.*

Un constat particulier sur ce défaut de convergence du discret vers le continu est que si l'on prend pour tout N la probabilité optimale (au sens du trou spectral maximal) de changer de sens dans la marche persistante, alors on converge vers un taux $1/2$ de saut pour Y_t , qui n'est pas optimal pour le processus continu, et qui donne le même taux exponentiel $1/2$ de convergence que le mouvement brownien sur le tore.

Cependant le phénomène de décroissance initiale en t^3 , lui, n'est pas affecté par cette subtilité ; c'est normal car son origine n'est pas dans la prise du supremum des normes des restrictions mais, déjà localement, sur chacun des plans W_k . Une interprétation possible est que prendre, au lieu d'un processus réversible, l'intégrale d'un processus réversible retarde initialement l'effet de mélange du hasard ; ou bien que la particule commence par se déplacer de façon déterministe et brouille donc moins bien les pistes qu'une diffusion au moins initialement.

Si l'on compare (en oubliant le défaut de convergence) la marche simple et la marche persistante pour $a = 1$ à la limite, pour un nombre n d'itérations fixé, l'écart L^2 à l'équilibre de la marche réversible est environ $1 - \frac{t}{2}$ avec $t = n \left(\frac{2\pi}{N}\right)^2$ et celle de la non réversible est $1 - \frac{t^3}{3}$ avec $t = n \frac{2\pi}{N}$, qui devient meilleure que la précédente pour $n \approx \sqrt{\frac{3}{4\pi}} N$: c'est le nombre d'itérations à partir duquel la marche irréversible est plus proche de la mesure uniforme que la réversible.

2.4 Du continu au mouvement brownien

Lorsque $a \rightarrow +\infty$, la vitesse du processus continu saute de plus en plus vite de -1 en 1 ; à la limite, les vitesses en deux temps distincts devraient donc être décorrélatées. Le processus devrait en conséquence être l'intégrale d'un bruit blanc, autrement dit un mouvement brownien. Avec la bonne renormalisation, c'est effectivement le cas :

Proposition 2.14. $X^a = (X_{ta})_{t>0}$ converge en loi vers un mouvement brownien standard sur \mathbb{T} quand $a \rightarrow +\infty$.

Démonstration. Notons $\tilde{Y}_t = (-1)^{N_t}$ où N_t est un processus de Poisson de paramètre 1. Ainsi X suit la même loi que $\int_0^t \tilde{Y}_{as} ds$

$$\begin{aligned} X^a(t) &\stackrel{\mathcal{L}}{=} \int_0^{ta} \tilde{Y}_{as} ds \\ &= \frac{1}{a} \int_0^{ta^2} \tilde{Y}_u du \end{aligned}$$

ce qui nous ramène à l'exemple 3 p. 360 de [44] où l'on nous indique la marche à suivre.

Détaillons : on montre d'abord que $M(t) = \tilde{Y}_t + 2 \int_0^t \tilde{Y}_u du$ est une martingale. En effet le nombre de changement de signes de \tilde{Y}_t dans une période $t - s$ suit une loi de Poisson de paramètre $t - s$, et ainsi

$$\begin{aligned} \mathbb{P}(\tilde{Y}_t = \tilde{Y}_s) &= \sum_{k \text{ pair}} \frac{(t-s)^k}{k!} e^{-(t-s)} = \cosh(t-s) e^{-(t-s)} \\ \mathbb{P}(\tilde{Y}_t = -\tilde{Y}_s) &= \sum_{k \text{ impair}} \frac{(t-s)^k}{k!} e^{-(t-s)} = \sinh(t-s) e^{-(t-s)} \end{aligned}$$

Ainsi $\mathbb{E}(\tilde{Y}_t | \mathcal{F}_s) = \tilde{Y}_s e^{-2(t-s)}$, et

$$\begin{aligned} \mathbb{E}(M(t) | \mathcal{F}_s) &= \tilde{Y}_s e^{-2(t-s)} + 2 \int_0^s \tilde{Y}_u du + 2 \int_s^t \tilde{Y}_s e^{-2(u-s)} du \\ &= \tilde{Y}_s + 2 \int_0^s \tilde{Y}_u du \\ &= M(s) \end{aligned}$$

Si l'on montre la convergence de la martingale $\frac{1}{n} M(n^2 t) = 2X_t^n + \frac{1}{n} \tilde{Y}_t$ vers le brownien, on aura celle de X^n ; or la première s'obtient de la convergence des crochets. La variation quadratique de $\int_0^s \tilde{Y}_u du$, processus 1-lipschitzien, est nulle, donc

$$\langle M \rangle_t = \lim_{\delta \rightarrow 0}^{\mathbb{P}} \sum_{t_i \in \pi} (\tilde{Y}_{t_{i+1}} - \tilde{Y}_{t_i})^2$$

où la limite en proba a lieu lorsque le pas δ de la partition π de $[0, t]$ tend vers 0. Notons Z_t le nombre de saut de \tilde{Y} sur cet intervalle.

$$\begin{aligned} \mathbb{P}\left(\sum_{t_i \in \pi} (\tilde{Y}_{t_{i+1}} - \tilde{Y}_{t_i})^2 \neq 4Z_t\right) &\leq \mathbb{P}(\text{deux sauts sont distants de moins de } \delta) \\ &\xrightarrow{\delta \rightarrow 0} 0 \end{aligned}$$

Ainsi $\langle M \rangle_t = 4Z_t$, et $\langle \frac{1}{n}M(n^2 \cdot) \rangle_t = \frac{4}{n^2}Z_{n^2t} \xrightarrow{n \rightarrow +\infty} 4t$ (par la loi des grands nombres), ce qui donne la convergence de $\frac{1}{2}\frac{1}{n}M(n^2t)$ (et donc de X_t^n) vers le mouvement brownien standard (cf [44]). \square

Qu'en est-il de la norme ? Celle du modèle irréversible converge-t-elle vers celle du brownien ? C'est effectivement le cas. Le générateur du mouvement brownien est $\frac{1}{2}\partial_x^2$, diagonalisable dans la base orthonormée des $x \mapsto e^{inx}$ pour les valeurs propres $-\frac{n^2}{2}$. Rappelons la norme du semi-groupe associé à (X_t, Y_t) sur le plan V_n , quand $a > n$:

$$\|P_t\|_{V_n}^2 = e^{2\lambda_1 t} \left(1 + \frac{2}{\frac{\omega^2}{n^2} \left(\frac{1+\gamma}{1-\gamma}\right) + \frac{a}{n} \sqrt{1 + \frac{\omega^2}{n^2} \left(\frac{1+\gamma}{1-\gamma}\right)^2} - 1} \right)$$

avec $\lambda_1 = -a + \sqrt{a^2 - n^2}$, $\omega = \sqrt{a^2 - n^2}$ et $\gamma = e^{-2t\sqrt{a^2 - n^2}}$. On observe que $a\lambda_1 \rightarrow -\frac{1}{2}n^2$, $\omega \rightarrow +\infty$ et que $\gamma^a \rightarrow 0$ quand $a \rightarrow +\infty$; Au final, en notant P_t^a le semi-groupe associé à (X_{at}, Y_{at}) , on récupère

$$\|P_t^a\|_{V_n} \xrightarrow{a \rightarrow +\infty} e^{-\frac{1}{2}n^2 t},$$

ce qui est la norme du semi-groupe Q_t associé au mouvement brownien sur la droite $\text{vect}\{x \mapsto e^{inx}\}$. En particulier la convergence pour $n = 1$ donne la convergence de la norme globale $\|P_t^a - \mu\| \rightarrow \|Q_t - \lambda\|$.

2.5 Généralisations

2.5.1 Avec un potentiel général

En fait le cas précédent, où la mesure invariante pour X_t est la loi uniforme sur le cercle, est immédiatement généralisable à des processus admettant pour loi limite n'importe quelle mesure de la forme $\nu = e^{-V(x)} dx$. En effet, considérons comme précédemment $Y_t \in \{-1, 1\}$ qui, avec taux a , change de signe. Soit $X_t \in \mathbb{T}$ la solution de

$$dX_t = Y_t e^{V(X_t)} dt.$$

Autrement dit X_t représente la position d'une particule se déplaçant à vitesse (déterministe) inversement proportionnelle à la densité $e^{-V(x)}$ (les zones « peu intéressantes » sont parcourues plus vite) et changeant de sens de parcours selon des temps exponentiels. Montrons qu'alors la mesure invariante pour (X_t, Y_t) est $\mu = \nu \otimes \mathcal{U}_{\{-1,1\}}$, et que la norme 2 du semi-groupe associé se calcule exactement comme précédemment. Le générateur markovien associé au processus est

$$\mathcal{L}f(x, y) = e^{V(x)}y\partial_x f(x, y) + a(f(x, -y) - f(x, y))$$

Et l'on vérifie

$$\begin{aligned} & \nu \otimes \mathcal{U}_{\{-1,1\}}[\mathcal{L}f(x, y)] \\ &= \int_{x \in \mathbb{T}} \int_{y=\pm} [e^{V(x)}y\partial_x f(x, y) + a(f(x, -y) - f(x, y))] e^{-V(x)} dx dy \\ &= \int_{y=\pm} y \left(\int_{x \in \mathbb{T}} \partial_x f(x, y) dx \right) dy \\ &= 0. \end{aligned}$$

Normalisons dans la suite V pour avoir $\nu(\mathbb{T}) = 2\pi$. Considérons pour $n \in \mathbb{N}$, $g_n(x) = \exp\left(in \int_0^x e^{-V(u)} du\right)$ et des fonctions de la forme $f(x, y) = g_n(x)h(y)$. Alors

$$\begin{aligned} \mathcal{L}f(x, y) &= g_n(x) [inyh(y) + a(h(-y) - h(y))] \\ &= g_n(x) K_n h(y) \end{aligned}$$

avec K_n précédemment défini dans le cas uniforme. On parvient donc là encore à décomposer l'espace en plans stables V_n , et ces plans sont à nouveau orthogonaux entre eux dans $L^2(\mu)$:

$$\begin{aligned} \langle g_n, g_k \rangle_{L^2(\mu)} &= \int_0^{2\pi} \exp\left(i(n-k) \int_0^x e^{-V(u)} du\right) e^{-V(x)} dx \\ &= \int_0^{2\pi} e^{i(n-k)u} du \\ &= 2\pi \delta_{nk} \end{aligned}$$

Finalement, si P_t^V est le semi-groupe associé au processus (et P_t est toujours celui associé au potentiel nul), on a exactement

$$\|P_t^V - \mu\|_{L^2(\mu)} = \|P_t - \lambda \otimes \mathcal{U}_{\{-1,1\}}\|_{L^2(\lambda \otimes \mathcal{U}_{\{-1,1\}})}$$

Remarquons que lorsque V n'est connu qu'à une constante près, modifier cette constante ne fait que dilater le temps et modifier le paramètre a .

2.5.2 En dimension supérieure

Remarquons que, dans l'optique d'un algorithme de Monte-Carlo non réversible, les résultats s'adaptent à la dimension supérieure. Ainsi en définissant Y_t^1, \dots, Y_t^d et X_t^1, \dots, X_t^d comme précédemment, dans le cas où $V(x) = \sum V_i(x_i)$, on construit un semi-groupe P_t^V sur \mathbb{T}^d de mesure invariante $\mu = e^{-V(x)} dx \otimes \mathcal{U}_{\{-1,1\}}^{\otimes d}$ et de norme

$$\|P_t^V\|_{L^2(\nu)} = \prod_{i=1}^d \|P_t^{V_i}\|_{L^2(e^{-V_i(x_i)} dx_i \otimes U_{\{-1,1\}})}$$

On aurait pu imaginer un autre processus, construit en gardant l'idée d'une particule dont la vitesse scalaire dépendrait de façon déterministe de la position mais dont la direction changerait aléatoirement à taux constant. Cela donnerait un générateur du type :

$$\mathcal{L}f(x, y) = e^{V(x)} \mathcal{D}_x f(x, y) \cdot y + a \int_{\mathbb{S}^d} (f(x, z) - f(x, y)) dz$$

pour $f \in \mathcal{C}_0(\mathbb{R}^d \times \mathbb{S}^d)$. Ci-dessus les vitesses sont prises uniformément sur la sphère mais on aurait pu les choisir différemment sans que les remarques à suivre ne s'en trouvent modifiées. La mesure invariante est alors $e^{-V(x)} dx \otimes \mathcal{U}_{\mathbb{S}^d}$, ce qui semble bien parti. Néanmoins, à part pour un potentiel nul, on ne va pas pouvoir se ramener à l'étude d'un opérateur sur les vitesses par la même méthode qu'avant, c'est-à-dire en trouvant des fonctions propres de la famille d'opérateurs $K_y : f(x) \mapsto e^{V(x)} \mathcal{D}_x f(x) \cdot y$ sous la forme $f(x) = e^{u(x)}$, qui permettraient jusqu'ici de se ramener à des opérateurs n'agissant que sur les vitesses. En effet on a alors

$$K_y f(x) = e^{V(x)} f(x) \mathcal{D}_x u(x) \cdot y$$

Il s'agirait donc de trouver une fonction $u : \mathbb{R}^d \rightarrow \mathbb{R}$ de différentielle $x \mapsto e^{-V(x)}(c_1 dx_1 + c_2 dx_2 + \dots + c_d dx_d)$. Or cette 1-forme linéaire n'est pas exacte (n'étant pas fermée), et un tel u ne saurait exister. En comparaison, pour le processus avec des coordonnées indépendantes du début de ce paragraphe, la 1-forme linéaire qui apparaît est $x \mapsto \sum e^{-V_i(x_i)} dx_i$, qui est bel et bien exacte.

2.6 Conclusion

Un potentiel général sur \mathbb{R}^d n'ayant aucune raison d'être la somme de potentiels unidimensionnels, les télégraphes intégrés présentés à l'instant ne sont pas de bons candidats à la simulation des mesures de Gibbs pour un problème pratique. Dans le chapitre suivant nous allons introduire un nouveau processus, le RTP, basé sur une dynamique analogue au télégraphe intégré, qui remplira ce rôle. Il sera impossible de faire des calculs aussi explicites que dans le présent chapitre, et on s'attachera donc à démontrer la convergence exponentielle de la variance vers 0 par des arguments de type hypocoercifs, d'une part proches à la fois de l'approche "système dynamique" de Desvillettes et Villani ([32]) et de la modification de la fonctionnelle de Lyapunov de Villani, Dolbeault, Mouhot et Schmeiser ([38, 85]), et d'autre part basés sur l'idée qu'il faut tenir compte des inhomogénéités temporelles dans la convergence, observées numériquement dans [45] et théoriquement dans ce chapitre.

2.7 Appendice

Lemme 2.15. Si $f(R) = \frac{R-a}{R^2+b}$ avec $b > 0$, alors f admet ses valeurs extrêmes en $R_{\pm} = a \pm \sqrt{a^2+b}$, et ces valeurs sont $f(R_{\pm}) = \frac{1}{2R_{\pm}}$.

Lemme 2.16. Si $g(\theta) = \frac{\alpha+\cos(\theta-s)}{\alpha+\cos(\theta)}$ avec $\alpha > 1$, alors

$$\max_{\theta \in \mathbb{T}} g(\theta) = 1 + \frac{2}{\sqrt{\frac{2(\alpha^2-1)}{1-\cos(s)} + 1} - 1}$$

De plus ce maximum est majoré par $\frac{\alpha+1}{\alpha-1}$, borne atteinte uniquement pour $s = \pi [2\pi]$.

Démonstration. Le premier lemme ne présente aucune difficulté. Pour le second, remarquons tout d'abord pour $s = 0 [2\pi]$ que g est alors constante égale à 1 et son max l'est également, le lemme est donc vrai dans ce cas. Supposons dans la suite que $1 - \cos(s) \neq 0$. Réécrivons maintenant

$$g(\theta) = \frac{\alpha + \cos(\theta + s)}{\alpha + \cos(\theta)} = \cos(s) + \frac{\alpha(1 - \cos(s)) - \sin(\theta) \sin(s)}{\alpha + \cos(\theta)}$$

$g(\theta)$ étant continue périodique il suffit de déterminer ses points critiques. Or $g'(\theta) = 0$ équivaut à

$$\begin{aligned} 0 &= -\cos(\theta) \sin(s) (\alpha + \cos(\theta)) + \sin(\theta) (\alpha(1 - \cos(s)) - \sin(\theta) \sin(s)) \\ &= -\sin(s) + \sin(\theta) (\alpha(1 - \cos(s)) - \alpha \cos(\theta) \sin(s)), \end{aligned}$$

équation affine dont les solutions sont

$$\begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} = \begin{pmatrix} -\beta \\ 0 \end{pmatrix} + t \begin{pmatrix} 1 - \cos(s) \\ \sin(s) \end{pmatrix}$$

pour $t \in \mathbb{R}$ et où l'on note $\beta = \frac{1}{\alpha}$. La condition $\cos^2 + \sin^2 = 1$ équivaut à

$$t^2 - \beta t + \frac{\beta^2 - 1}{1 - \cos(s)}$$

qui admet nécessairement deux solutions réelles puisque g est périodique non constante donc possède au moins deux points critiques. Ces solutions sont données par

$$t(1 - \cos(s)) = \frac{1}{2}\beta(1 - \cos(s)) + \frac{1}{2}\varepsilon\sqrt{2(1 - \cos(s)) - \beta^2 \sin^2(s)}$$

où $\varepsilon = \pm 1$. On obtient ainsi les valeurs extrêmes de g :

$$\begin{aligned}
g(\theta_\varepsilon) &= \cos(s) + \frac{\alpha(1 - \cos(s)) - t \sin^2(s)}{\alpha - \beta + t(1 - \cos(s))} \\
&= \frac{\alpha - t \sin^2(s) - \beta \cos(s) + t \cos(s) - t \cos^2(s)}{\alpha - \beta + t(1 - \cos(s))} \\
&= \frac{\alpha - \beta \cos(s) - t(1 - \cos(s))}{\alpha - \beta + t(1 - \cos(s))} \\
&= \frac{\alpha - \frac{1}{2}\beta(1 + \cos(s)) - \frac{1}{2}\varepsilon\sqrt{2(1 - \cos(s)) - \beta^2 \sin^2(s)}}{\alpha - \frac{1}{2}\beta(1 + \cos(s)) + \frac{1}{2}\varepsilon\sqrt{2(1 - \cos(s)) - \beta^2 \sin^2(s)}}
\end{aligned}$$

Puisque $\alpha > 1 > \beta$, on a $\alpha - \frac{1}{2}\beta(1 + \cos(s)) > 0$ et la valeur ci-dessus est maximale pour $\varepsilon = -1$, et ainsi

$$\begin{aligned}
\max_{\theta \in \mathbb{T}} g(\theta) &= \frac{\alpha - \beta \left(\frac{1+\cos(s)}{2} \right) + \sqrt{\left(\frac{1-\cos(s)}{2} \right) \left(1 - \beta^2 \frac{1+\cos(s)}{2} \right)}}{\alpha - \beta \left(\frac{1+\cos(s)}{2} \right) - \sqrt{\left(\frac{1-\cos(s)}{2} \right) \left(1 - \beta^2 \frac{1+\cos(s)}{2} \right)}} \times \frac{\alpha}{\alpha} \\
&= \frac{\alpha^2 - \left(\frac{1+\cos(s)}{2} \right) + \sqrt{\left(\frac{1-\cos(s)}{2} \right) \left(\alpha^2 - \frac{1+\cos(s)}{2} \right)}}{\alpha^2 - \left(\frac{1+\cos(s)}{2} \right) - \sqrt{\left(\frac{1-\cos(s)}{2} \right) \left(\alpha^2 - \frac{1+\cos(s)}{2} \right)}} \times \frac{\sqrt{\alpha^2 - \frac{1+\cos(s)}{2}}}{\sqrt{\alpha^2 - \frac{1+\cos(s)}{2}}} \\
&= \frac{\sqrt{\alpha^2 - \frac{1+\cos(s)}{2}} + \sqrt{\frac{1-\cos(s)}{2}}}{\sqrt{\alpha^2 - \frac{1+\cos(s)}{2}} - \sqrt{\frac{1-\cos(s)}{2}}} \times \frac{\sqrt{\frac{1-\cos(s)}{2}}}{\sqrt{\frac{1-\cos(s)}{2}}} \\
&= \frac{\sqrt{\frac{2(\alpha^2-1)}{1-\cos(s)} + 1} + 1}{\sqrt{\frac{2(\alpha^2-1)}{1-\cos(s)} + 1} - 1} \\
&= 1 + \frac{2}{\sqrt{\frac{2(\alpha^2-1)}{1-\cos(s)} + 1} - 1}
\end{aligned}$$

□

Lemme 2.17. *Considérons $T_i > 0$ pour $1 \leq i \leq M - 1$ et $\delta > 0$. Alors il existe $t > 0$ et $k_1, \dots, k_{M-1} \in \mathbb{Z}$ tels que pour tout $0 \leq i \leq M - 1$*

$$|k_i T_i - t| < \delta$$

Démonstration. Considérons¹ le réseau de \mathbb{R}^M engendré par les $(0, \dots, 0, T_n, 0, \dots, 0)$ (avec T_n en n^{eme} position) pour $n < M$ et par $(1, 1, \dots, 1)$, de volume fondamental V le produit des

1. L'argument me vient de Jérémy Leborgne; je l'en remercie.

T_n . Ainsi, en considérant le pavé $[-\delta, \delta] \times \dots \times [-\delta, \delta] \times [-V\delta^{-M+1}, V\delta^{-M+1}]$, de volume $V2^d$, on sait par le théorème de Minkowski qu'il contient au moins un point du réseau autre que l'origine. Les $M - 1$ premières coordonnées de ce point sont de la forme $h_n T_n + h_M$ avec $h_j \in \mathbb{Z}$ pour $j \leq M$. Si $\delta < \min(T_1, \dots, T_{M-1}, 1)$ alors aucun de ces coefficients h_j ne peut être nul, et nécessairement h_M est de signe opposé aux autres h_j . Il suffit donc de prendre $k_n = |h_n|$ et $t = |h_M|$. \square

Lemme 2.18. Notons, pour $s > 0$ et $p \in]0, 1[$,

$$h(p) = \frac{p}{1-p^2} \frac{1+e^{-ps}}{1-e^{-ps}}$$

$$\phi(p) = e^{ps} \left(1 + \frac{2}{ph(p) + \sqrt{h(p)^2 + \frac{1}{1-p^2}} - 1} \right).$$

Alors pour tout s , $p \mapsto \phi(p)$ est croissante.

En prenant $p = \sqrt{1 - \left(\frac{n}{a}\right)^2}$ et $s = 2at$ on obtient en particulier que pour $t > 0$ et $a > 0$, $n \in]0, a[\mapsto R(t, a, n)$ est décroissante.

Démonstration. Le calcul de la dérivée est effectué *via* Maple :

```
h:=p->p/(1-p^2)*(1+exp(-s*p))/(1-exp(-s*p)):
phi:=p->exp(p*s)*(1+2/(p*h(p)+sqrt(h(p)^2+1/(1-p^2))-1)):
resultat := simplify(exp(-p*s)*diff(phi(p),p)):
```

Le résultat est de la forme $\frac{\text{numérateur}(p)}{(\text{un terme})^2(p^2-1)(e^{-ps}-1)}$; il s'agit donc de vérifier que le numérateur est positif. À l'instruction

```
solve(numérateur(p)=0,p);
```

la réponse est

```
-RootOf(_Z exp(_Z) + _Z + 2 - 2 exp(_Z))/s
```

Autrement dit le numérateur s'annule en p si $e^p = \frac{2+p}{2-p}$, équation dont la seule solution est $p = 0$: en effet, s'il y avait une autre solution p^* , la dérivée de $e^p \frac{2-p}{2+p}$ s'annulerait entre 0 et p^* , or celle-ci est $\frac{-z^2 e^z}{(2+z)^2}$. Ainsi le numérateur est de signe constant pour $p \in]0, 1[$ et ϕ est monotone. Les limites de ϕ en 0 et 1 sont respectivement $1 + \frac{2}{\sqrt{4s^2+1}-1}$ et e^s , dont l'égalité est équivalente à $2e^s - 2 - s^2 = 0$, d'unique solution $s = 0$; vu leurs équivalents pour $s \rightarrow +\infty$ on a donc $\phi(1) > \phi(0)$ pour $s > 0$, donc ϕ est croissante. \square

Chapitre 3

Méthode d'ordre 3

Third time's a charm.

dicton anglais

Ce chapitre est une adaptation de l'article [76].

3.1 Un échantillonneur déterministe par morceau

Considérons un potentiel U sur \mathbb{R}^d . Notre but est de définir un processus de Markov $(X_t, Y_t)_{t \geq 0}$ jouissant des propriétés suivantes :

- X_t , la position, est dans \mathbb{R}^d , et Y_t , la vitesse, est dans \mathbb{S}^d .
- $dX_t = Y_t dt$.
- Le processus est ergodique, et la première marginale de sa loi μ est prescrite, égale à $e^{-U(x)} dx$.
- Y_t est un processus de saut.

Les deux premières de ces propriétés signifient que (X, Y) est un processus cinétique, comme celui de Langevin. La troisième signifie qu'il permet de simuler une mesure de Gibbs, la tâche pour laquelle il est conçu. La quatrième enfin rend le processus a priori simple à simuler, concrètement, par un ordinateur. Plus précisément entre deux saut de Y , (X, Y) est complètement déterministe. De tels processus de Markov déterministes par morceaux (ou PDMP, pour l'anglais *piecewise deterministic Markov processes*) ont récemment attiré l'attention de différents domaines, pour la raison qu'ils constituent une alternative simple aux diffusions quant il s'agit de modéliser des systèmes stochastiques (nous reviendrons plus en détail sur les PDMP au chapitre suivant ; pour l'heure, signalons simplement la référence [7] pour un aperçu de quelques questions récentes sur le sujet).

En dimension un, la vitesse est 1 ou -1, comme pour le télégraphe intégré du chapitre précédent. Si le processus passe à deux reprises par un même état (x, y) , c'est qu'il a nécessairement fait un demi-tour entre les deux, et est passé par x dans l'autre sens, c'est-à-dire qu'il a visité l'état $(x, -y)$. En moyenne (en temps), $(x, -1)$ et $(x, +1)$ sont donc équitablement visités. L'ergodicité implique alors que la loi invariante est une mesure produit dont la seconde marginale est uniforme sur $\{\pm 1\}$. Afin de réduire notre champ d'investigation, nous ajoutons cette contrainte pour toute dimension, de sorte que la mesure invariante du processus est $\mu(dx, dy) = e^{-U(x)} \sigma(dy) dx$ (où σ désigne la mesure uniforme sur la sphère).

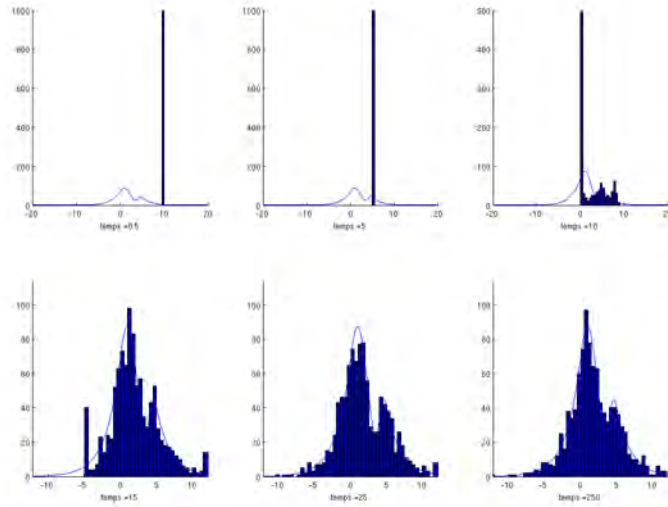


FIGURE 3.1 – Première marginale du processus à différents temps pour une mesure invariante bi-modale $e^{-U(x)}dx$, $(X_0, Y_0) = (7, -1)$ et $a(x, y) = (yU'(x))_+$. Tant que le potentiel décroît le long de la trajectoire, le processus est déterministe. Il s'échappe facilement du minimum local.

Derechef en dimension un, la seule possibilité au moment d'un saut est de multiplier la vitesse par son opposée. Ceci conduit à un générateur infinitésimal de la forme

$$Lf(x, y) = yf'(x, y) + a(x, y) (f(x, -y) - f(x, y))$$

où le taux de saut a est une fonction positive. Dans ce cas l'invariance de la loi μ , c'est-à-dire l'équation $\mu L = 0$, se lit

$$yU'(x) = a(x, y) - a(x, -y),$$

qui est satisfaite si et seulement si a est de la forme $a(x, y) = \frac{1}{2}(yU'(x) + b(x))$ pour une fonction $b(x)$ (remarquons que nécessairement $b(x) = a(x, y) + a(x, -y)$). La positivité de a implique que $b(x) \geq |U'(x)|$. Quand cette inégalité est en fait un égalité, on obtient $a(x, y) = (yU'(x))_+$ (où par $(g)_+$ on signifie la partie positive de g) : c'est le choix qui minimise le taux de saut, autrement dit le comportement dissipatif du processus. D'autre part ce taux de saut est pertinent du point de vue de la simulation. En effet il implique que, tant que $Y_t U'(X_t) \leq 0$, en d'autres termes tant que le processus descend le long du potentiel, aucun saut n'est permis. Au contraire lors d'une ascension, c'est-à-dire lorsque $Y_t U'(X_t) > 0$, le prochain temps de saut T sera tel que

$$E := \int_0^T yU'(x + ys)ds = U(x + T) - U(x)$$

est de loi exponentielle de paramètre 1 $\mathcal{E}(1)$ (on voit sur cette définition de T que la régularité de U , utile pour écrire le générateur, n'est pas nécessaire pour définir le processus). Il est donc suffisant de calculer le potentiel le long de la trajectoire, et de simuler un processus de Poisson. Notez que pour un taux "non minimal" de la forme $a(x, y) = (yU'(x))_+ + a_*$ où $a_* > 0$ est un taux de saut résiduel, présent à tout instant, il suffit simplement d'ajouter à la dynamique minimale des sauts à taux constants a_* (c'est-à-dire de simuler un second processus de Poisson). Une simulation de cette dynamique est représenté dans les figures 3.1 et 3.2.

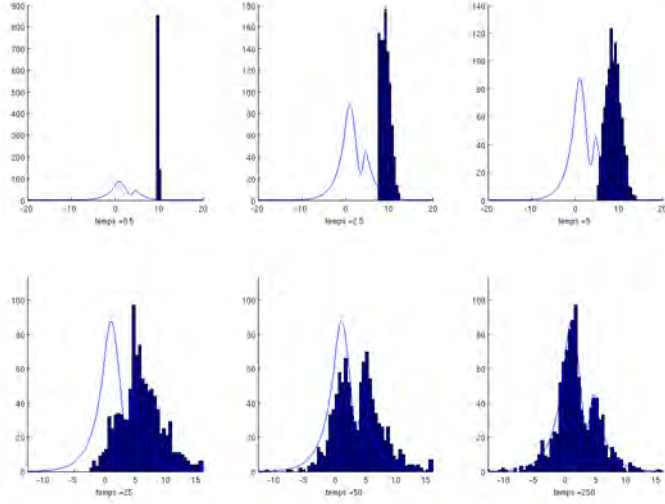


FIGURE 3.2 – Ici $a(x, y) = 1 + (yU'(x))_+$. Autrement dit, contrairement au cas de la figure 3.1, persiste à tout instant un niveau minimal d'aléa : le comportement est plus dissipatif, et le processus s'échappe plus difficilement du minimum local.

En toute dimension, la remarque précédente tient encore pour le taux de saut $a(x, y) = (y \cdot \nabla_x U(x))_+$. Chercher un noyau de transition $p_{x,y}(z)$ sur \mathbb{S}^d tel que μ soit invariante pour le générateur

$$Lf(x, y) = y \cdot \nabla_x f(x, y) + (y \cdot \nabla_x U(x))_+ \left(\int_{\mathbb{S}^d} f(x, z) p_{x,y}(z) \sigma(dz) - f(x, y) \right),$$

revient à résoudre (au sens faible) l'équation

$$\forall (x, y) \in \mathbb{R}^d \times \mathbb{S}^d, \quad y \cdot \nabla_x U(x) = (y \cdot \nabla_x U(x))_+ - \int_{\mathbb{S}^d} (z \cdot \nabla_x U(x))_+ p_{x,z}(y) \sigma(dz).$$

Notons que si, en particulier, $y^* = -\frac{\nabla_x U(x)}{|\nabla_x U(x)|}$, dans la mesure où $(z \cdot \nabla_x U(x))_+ < -y^* \cdot \nabla_x U(x)$ pour tout $z \neq -y^*$, la seule possibilité est $p_{x,-y^*} = \delta_{y^*}$. D'ailleurs le noyau $p_{x,y} = \delta_{-y}$ est une solution du problème pour tout couple (x, y) puisque

$$y \cdot \nabla_x U(x) = (y \cdot \nabla_x U(x))_+ - (-y \cdot \nabla_x U(x))_+.$$

Qui plus est, multiplier la vitesse par -1 n'a pas de coût numérique ; et si l'on a gardé en mémoire pendant la simulation le dernier temps à partir duquel le processus a commencé à grimper contre la pente du potentiel, on peut directement sauter jusqu'à ce point et court-circuiter la descente déterministe. On en arrive donc à considérer le générateur

$$Lf(x, y) = y \cdot \nabla_x f(x, y) + (y \cdot \nabla_x U(x))_+ (f(x, -y) - f(x, y)),$$

qui définit un processus très simple pour lequel μ est invariante. Le problème, évidemment, est qu'il n'est pas du tout ergodique, puisqu'il n'est autorisé qu'à faire demi-tour, coincé pour toujours dans une seule et même direction. Pour résoudre ce problème on peut ajouter un autre générateur markovien, qui n'agirait que sur la variable de vitesse, ergodique sur la sphère de

mesure invariante la loi uniforme. Parmi les différentes possibilités, l'une peut-être des plus simples à implementer serait

$$Lf(x, y) = y \cdot \nabla_x f(x, y) + (y \cdot \nabla_x U(x))_+ (f(x, -y) - f(x, y)) + r \left(\int f(x, z) dz - f(x, y) \right).$$

Autrement dit, on ajoute à la dynamique précédente des sauts (uniformes sur la sphère) de la vitesse à taux constant r .

Bien qu'à ma connaissance ce processus ne semble pas avoir été considéré antérieurement pour échantillonner des mesures de Gibbs, il a déjà été étudié. Introduit dans [77] par Ratanov, il apparaît notamment pour modéliser les déplacements de la bactérie *Escheria coli* (par exemple dans [22, 46]). Dans la littérature on lui donne parfois encore le nom de télégraphe intégré, ou encore de télégraphe intégré biaisé, ou ergodique. Cependant ce terme ne semble plus très adapté au cas où le taux de saut dépend d'un potentiel préexistant (qui plus est multi-dimensionnel) - que ce soit dans notre cas à fin d'optimisation, ou dans la modélisation d'un gradient de nutriments - et on le baptisera RTP pour *run and tumble process*, c'est-à-dire littéralement dans la langue de Molière le processus de course et bascule, en référence au mouvement bactérien.

Outre le modèle jouet du tore muni de la mesure uniforme considéré au chapitre précédent et dans les articles [75, 38], l'ergodicité de ce processus avait déjà été quantifiée par Fontbona, Guérin et Malrieu ([46]) par des méthodes de couplages dans le cas d'un potentiel présentant un unique minimum en 0. Peu de temps après la publication des résultats de ce chapitre dans [76], Calvez, Raoul et Schmeiser l'ont également traité dans [22] avec une méthode de fonction de Lyapunov modifiée dans l'esprit de [38]. Dans chacun de ces articles, le taux de saut est supposé minoré par une constante positive; autrement dit le RTP qu'on appellera optimal ou complètement dégénéré, de taux de saut $a(x, y) = (y \cdot \nabla U(x))_+$, passe entre les mailles du filet.

Deux dynamiques cinétiques

On va s'appliquer à démontrer la décroissance exponentielle de la variance de $P_t f$ par rapport à la mesure invariante dans deux cas cinétiques : le RTP que nous venons d'introduire, et le processus de Langevin cinétique (que l'on désignera dans la suite par kFP pour *kinetic Fokker-Planck*, la dénomination de [32, 54, 56])

$$\begin{cases} dX_t = Y_t dt \\ dY_t = -U'(X_t) dt - Y_t dt + \sqrt{2} dB_t, \end{cases} \quad (3.1)$$

dont on rappelle le générateur

$$Lf = y \partial_x f - U'(x) \partial_y f - y \partial_y f + \partial_y^2 f. \quad (3.2)$$

Plus précisément le RTP considéré aura pour générateur

$$Lf = y \partial_x f + a(x, y) (f_- - f), \quad (3.3)$$

en notant $f_-(x, y) = f(x, -y)$. Comme on le voit on se restreint à la dimension un. L'objectif principal à cela est de garder les calculs à un niveau raisonnable. J'ai appliqué (avec succès)

la méthode ci-dessous au RTP en dimension supérieure, mais ces résultats gagneront à être énoncés lorsqu'une meilleure compréhension des mécanismes en jeux leur fournira un cadre plus abstrait, plus agréable pour la généralisation.

L'unique loi invariante du kFP est $\mu = e^{-U(x)}dx \otimes e^{-\frac{y^2}{2}}dy$, la mesure de Gibbs associé à l'Hamiltonien $U(x) + \frac{y^2}{2}$. Celle du RTP est $\mu = e^{-U(x)}dx \otimes \frac{\delta_1 + \delta_{-1}}{2}(dy)$ avec $U'(x) = a(x, 1) - a(x, -1)$. On normalise U dans les deux cas de sorte que μ soit de masse 1. Les hypothèses de croissance de U que nous ferons dans la suite entraîneront l'ergodicité des deux processus.

Ces deux processus ont plusieurs particularités en commun. L'une d'elle est l'absence de coercivité de la part de la partie déterministe de la dynamique, le potentiel n'étant pas supposé convexe ; autrement dit deux particules couplées avec le même aléa (le même Brownien pour le kFP, le même processus de Poisson pour le RTP) n'ont aucune tendance à se rapprocher. D'un autre côté l'aléa intervient seulement sur la variable de vitesse, de sorte que les processus sont complètement dégénérés au sens de [9] et que leurs courbures de Bakry-Emery sont égales à $-\infty$.

Dans la mesure où l'article [76] (et donc ce chapitre qui en est tiré) adopte un point de vue plutôt analytique que probabiliste sur le problème, rappelons que, dans le monde des EDP, l'étude de telles dynamiques est plutôt vue de la façon duale suivante : en notant \hat{L} le dual L , qui opère sur les mesures (plus précisément ici sur les densités régulières¹), la loi $\mu_t = \mathcal{L}(X_t, Y_t)$ est la solution (faible) de

$$\begin{cases} \partial_t \mu_t = \hat{L} \mu_t \\ \mu_0 = \mathcal{L}(X_0, Y_0). \end{cases}$$

Ainsi

$$P_t f(x, y) = \int f(u, v) \mu_t(du, dv)$$

lorsque $\mu_0 = \delta_{(x,y)}$. Notre but est de quantifier la convergence de μ_t vers μ .

On l'a vu dans le chapitre 1, des taux de convergence exponentielle explicites ont déjà été obtenus pour le kFP ; par fonction de Lyapunov modifiée ([85, 38]), systèmes différentiels ([32]), couplages ([40]) ou analyse spectrale hypoelliptique ([56, 54])... On aurait pu essayer d'appliquer l'une ou l'autre de ces méthodes au RTP (c'est d'ailleurs ce qui sera fait dans [22]).

Cependant ces méthodes ne donnent pas d'asymptotiques précises du taux de convergence à basse température, dans l'esprit par exemple du théorème 1.18 de l'introduction pour le processus de Fokker-Planck réversible. Il fallait donc tenter autre chose. L'idée développée dans le reste de ce chapitre a initialement été proposée par Gadat et Miclo dans [47]. Elle part du constat suivant : on sait bien que la variance $P_t f$ va finir par décroître et tendre 0. Si sa dérivée est nulle, c'est donc qu'une de ses dérivées supérieures doit être négative. De plus les dérivations successives de la variance font intervenir l'algèbre de Lie des parties symétriques et anti-symétriques du générateur, qui intervient dans la théorie hypoelliptique d'Hörmander. Ensuite un calcul de la dérivée troisième de la variance dans le cas du kFP avec potentiel quadratique (toujours dans [47]) fait apparaître le terme $\|\partial_x P_t f\|_2^2$, qui est exactement la "coercivité manquante" qui sépare $\mu \Gamma f = \|\partial_y P_t f\|_2^2$ de $\mathcal{E}_\mu f = \|\nabla P_t f\|_2^2$. Enfin le comportement de la variance dans les cas cinétiques où elle est explicitement calculable (c'est-à-dire dans [47] et dans le chapitre précédent) présente les deux caractéristiques suivantes :

1. on rappelle qu'on travaille sous l'hypothèse de l'existence d'une algèbre de fonctions régulières dense dans $L^2(\mu)$. Dans ce chapitre, pour conserver les notations de [76], elle sera notée \mathcal{D} .

— décroissance initiale de l'ordre de t^3 ,
— possibles oscillations dans le préfacteur,
typiques des fonctions $\phi(t) = e^{-\lambda t} (a + b \cos(\nu t + \theta))$, qui sont les solutions de

$$\left((\partial_t + \lambda)^3 + \nu^2 (\partial_t + \lambda) \right) \phi = 0.$$

Il est donc légitime de se demander s'il n'existerait pas des constantes telles qu'une inégalité différentielle de la forme (en notant F_t la variance de $P_t f$)

$$F_t''' + aF_t'' + bF_t' + cF_t \leq 0.$$

Hélas il n'en est rien dans le cas du kFP gaussien, ni pour le télégraphe intégré, où l'on peut se convaincre que les constantes a , b et c devraient dépendre de n en fonction du plan V_n (définis en section 2.2) sur lequel on se trouve. Plus précisément, dans une écriture sous la forme

$$\left((\partial_t + \lambda)^3 + \nu^2 (\partial_t + \lambda) \right) F_t \leq 0,$$

il faudrait que ν dépende de n , et devienne arbitrairement grand avec n . Mais dès que ν est positif, cela signifie juste que la solution ϕ de l'équation différentielle correspondante oscille de plus en plus vite; le taux exponentiel de convergence étant donné par λ . Un espoir subsiste.

Ceci a motivé les arguments que nous allons maintenant présenter.

Main result

Let $f_t = (P_t - \mu)f$ where $\mu f = \int f d\mu$, so that $\forall t \geq 0$, $\mu f_t = 0$; in other words $f_t \in 1^\perp$ the orthogonal space of the constants in $L^2(\mu)$. Let $F_t = \|f_t\|_{L^2(\mu)}^2$. In the following (cf. Section 3.2) we show that, for both the kFP and the RTP, under some assumptions on the potential U or on the rate a , there exist explicit $\lambda, \eta > 0$ and $\nu_* \in \mathbb{R}$ and a function $t \mapsto \nu_t \geq \nu_*$ such that

$$(\partial_t + \lambda) \left[(\partial_t + \eta)^2 + \nu_t \right] F_t \leq 0. \quad (3.4)$$

Furthermore ν_* is such that the roots of the polynomial $(X + \eta)^2 + \nu_*$ have negative real part, namely either $\nu_* \geq 0$ or $\sqrt{|\nu_*|} < \eta$; to sum up, $\mathcal{R}e(\eta - \sqrt{-\nu_*}) > 0$. Then exponential decay follows from the next result.

Theorem 3.1. *Assume (3.4) holds.*

- if $\nu_* \leq 0$ then $F_t \leq \phi_t$
- if $\nu_* > 0$ then $F_t \leq \phi_t + e^{-\eta t} \sup_{s \leq t} e^{\eta s} (\phi_s - F_s)$; furthermore the length of a time interval

where $F > \phi$ is less than $\frac{\pi}{\nu_*}$

with in both cases ϕ solution of

$$(\partial_t + \lambda) \left[(\partial_t + \eta)^2 + \nu_* \right] \phi_t = 0,$$

$\phi_0 = F_0$, $\phi'_0 = F'_0$ and an explicit ϕ''_0 . In particular,

$$\lim_{t \rightarrow \infty} \frac{1}{t} \ln F_t \leq -\min \left(\lambda, \mathcal{R}e(\eta - \sqrt{-\nu_*}) \right)$$

Remark 3.2. — $\nu_* \leq 0$ can always be assumed.

- For $\nu_* > 0$, ϕ presents damped oscillations with a period $\frac{2\pi}{\nu_*}$ and a magnitude of order $e^{-\eta t}$. The theorem shows that F is interlaced with ϕ : F can be above ϕ but only if it's already been below, and not for too long.
- The rate of convergence is independent from the function f , but this result does not give a bound for the operator norm of the semi-group in L^2 . As will be seen in the sequel, ϕ_0'' depends on $F''(0)$ and $\|\partial_x f_0\|^2$, which can be arbitrarily large with $F_0 = 1$ (we could obtain a bound by using estimates from the pseudodifferential calculus theory, but our aim was to avoid resorting to this powerful tool and to stay very elementary). The result in [38] does the job with no derivative - but not exactly with the L^2 norm ; it could be possible to do the same in the present work.

Démonstration. The Gronwall lemma gives

$$\left[(\partial_t + \eta)^2 + \nu_* \right] F_t \leq \left[(\partial_t + \eta)^2 + \nu_t \right] F_t \leq \left(\left[(\partial_t + \eta)^2 + \nu_t \right] F_t \right)_{t=0} e^{-\lambda t} := C_0 e^{-\lambda t}.$$

Let ϕ be the solution of

$$\begin{cases} \left[(\partial_t + \eta)^2 + \nu_* \right] \phi = C_0 e^{-\lambda t} \\ \phi_0 = F_0, \phi'_0 = F'_0. \end{cases}$$

Thus

$$\left[(\partial_t + \eta)^2 + \nu_* \right] (F_t - \phi_t) \leq 0.$$

In the case where $\nu_* \leq 0$, using twice the Gronwall lemma gives

$$\begin{aligned} (\partial_t + \eta - \sqrt{-\nu_*})(\partial_t + \eta + \sqrt{-\nu_*})(F_t - \phi_t) &\leq 0 \\ \Rightarrow (\partial_t + \eta + \sqrt{-\nu_*})(F_t - \phi_t) &\leq 0 \\ \Rightarrow F_t - \phi_t &\leq 0. \end{aligned}$$

So now assume $\nu_* > 0$ and define $h_t = e^{\eta t} (F_t - \phi_t)$, so that $h_t'' + \nu_* h_t \leq 0$, $h_0 = h'_0 = 0$. Define $M_t = \sup \{-h_s, 0 \leq s \leq t\}$ so that $h_t \geq -M_t$. M_t is nondecreasing and it is constant when h is increasing. Let us show that $h_t \leq M_t$ at every time. Assume it is false and consider $s = \inf\{t > 0, h_t > M_t\}$. M_t is constant for t in a neighborhood of s , $h_{s-\epsilon} < M_s$ and $h_{s+\epsilon} > M_s$ for $\epsilon > 0$ small enough. So, as $h_s'' \leq -\nu_* h_s < 0$, necessarily $h'_s > 0$, which leads to $(h'_s)^2 + \nu_* h_s^2 > \nu_* M_s^2$. Now consider $u = \sup\{0 \leq t \leq s, h'_t \leq 0\}$ (which exists if s exists, as $h'_0 \leq 0$) and note that $M_u = M_s$. We get $(h'_u)^2 + \nu_* h_u^2 = \nu_* h_u^2 \leq \nu_* M_u^2$. Yet for $t \in (u, s]$, $h'_t > 0$ so

$$h'_t (h_t'' + \nu_* h_t) = \frac{1}{2} \frac{d}{dt} \left((h'_t)^2 + \nu_* h_t^2 \right) \leq 0$$

and we've reached a contradiction.

Concerning the length of an interval where $F > \phi$, in other words where $h > 0$, define on this interval $\delta_t = -\frac{1}{h_t} (h_t'' + \nu_* h_t) \geq 0$. Thus h_t is solution of

$$\psi_t'' + (\nu_* + \delta_t) \psi_t = 0$$

and so vanishes, according to the Sturm-Liouville comparison theorem (cf. [37] for instance), between two successive zeros of $\cos(\nu_* t + \theta)$ for any θ . \square

In Section 3.2, the kFP and RTP models are proved to satisfy an inequality of the form (3.4). Section 3.3 is devoted to numerical studies, whose conclusion is that the method can give the good order of magnitude for the exponential rate of convergence, but shouldn't be trusted to compute parameters which accurately give the asymptotically fastest convergence. Finally an appendix gathers the proof of the technical lemmas used throughout this work.

3.2 Third order inequality

We start with considerations applying to both models. To compute the derivatives of F_t , we'll split L in its symmetric and anti-symmetric part. More precisely, if A and B are operators on $L^2(\mu)$, we denote by A^* the dual operator of A and by $[A, B]$ the Lie Brackets $AB - BA$, and \langle, \rangle stands for the scalar product on $L^2(\mu)$.

Lemma 3.3. *Assume*

$$L = K + R - R^*$$

with $K^* = K$. Then

$$\begin{aligned} F_t' &= \langle (2K)f_t, f_t \rangle \\ F_t'' &= \langle (2K)^2 f_t, f_t \rangle + 4 \langle [K, R]f_t, f_t \rangle \\ F_t''' &= \langle (2K)^3 f_t, f_t \rangle + 12 \langle [K^2, R]f_t, f_t \rangle + 4 \langle [[K, R], R - R^*] f_t, f_t \rangle. \end{aligned}$$

The proof is given in the appendix. The successive derivation of F_t could also be obtained with iterated Γ -calculus (see [65]), in particular for models where the invariant measure is not so easy to handle.

As in kinetic models the coercive part K of L only acts on the velocity variable, one cannot find any $\lambda > 0$ such that, for all f_t , $F_t' \leq -\lambda F_t$. We call μ_1 (resp. μ_2) the first (resp. second) marginal of μ , namely the position (resp. velocity) distribution at equilibrium. In our specific models we'll have $\mu = \mu_1 \otimes \mu_2$. We call $V = Ker(\mu_2 - 1)$ the set of functions which do not depend on y . The orthogonal projection to V and V^\perp will be respectively denoted by π_V and π_\perp :

$$\pi_V f(x, y) = \mathbb{E}(f(x, Z) | Z \sim \mu_2) = \int f(x, z) \mu_2(dz), \quad \pi_\perp = 1 - \pi_V.$$

We will note $f_V = \pi_V f_t$ and $f_\perp = \pi_\perp f_t$; as f_V only depends on x we will sometimes consider f_V as a one-parameter function in $L^2(\mu_1)$. Finally let $G_t = \|\partial_x f_t\|^2$, and recall that a measure ν is said to satisfy a Poincaré (or spectral gap) inequality with constant c if

$$\int |\partial_z g(z)|^2 d\nu(z) \geq c \int |g(z)|^2 d\nu(z) \tag{3.5}$$

whenever $\nu g = 0$.

Lemma 3.4. *We have $\mu_1 f_V = 0$. In particular, if μ_1 satisfies a Poincaré inequality with constant c , we have*

$$G_t \geq \|\partial_x f_V\|^2 \geq c \|f_V\|^2$$

If furthermore $F_t' \leq -d \|f_\perp\|^2$ then

$$\frac{1}{F_t} \left(G_t - \frac{c}{d} F_t' \right) \geq c$$

Démonstration. For the first assertion,

$$\int f_V(x) d\mu_1(x) = \int \left(\int f_t(x, y) d\mu_2(y) \right) d\mu_1(x) = \mu f_t = 0$$

Furthermore $\partial_x^* \partial_x$ is self-adjoint and stabilizes V , so it stabilizes V^\perp and

$$G_t = \|\partial_x(f_V + f_\perp)\|^2 = \|\partial_x f_V\|^2 + \|\partial_x f_\perp\|^2 \geq \|\partial_x f_V\|^2.$$

Then $G_t - \frac{c}{d} F_t' \geq c F_t'$ is clear. \square

Now we will show that in both models, the inequality (3.4) holds for some parameters.

3.2.1 The kinetic Fokker-Planck process

In this section (from Lemma 3.6 to Theorem 3.9) the generator is

$$L = y \partial_x - U'(x) \partial_y - y \partial_y + \partial_y^2 \quad (3.2)$$

The invariant measure is $\mu = e^{-U(x)} dx \otimes e^{-\frac{y^2}{2}} dy$ so that

$$\partial_x^* = U' - \partial_x, \quad \partial_y^* = y - \partial_y.$$

From now on we will make some assumptions on the potential U :

Assumption 3.5. *The potential U is smooth (C^2), U'' is bounded and $\mu_1 = e^{-U(x)} dx$ satisfies a Poincaré inequality with constant c_U*

The smoothness and the Poincaré inequality conditions are usual assumptions (for instance in [86], [38]) ; however the boundedness of U'' is quite restrictive, and could be an artefact due to the lack of subtlety of some of our computations.

We can decompose $L = K + R - R^*$ with

$$K = -\partial_y^* \partial_y \quad R = -\partial_x^* \partial_y \quad R^* = -\partial_y^* \partial_x.$$

We compute in appendix the brackets appearing in Lemma 3.3 :

Lemma 3.6.

$$\begin{aligned} [K, R] &= R \\ [K^2, R] &= R(1 + 2K) \\ [R, R^*] &= U'' K + \partial_x^* \partial_x. \end{aligned}$$

As expected the operator $-\partial_x^* \partial_x$ appears in the third derivative :

$$F_t''' = \langle (2K)^3 f_t, f_t \rangle + 12 \langle R(1 + 2K) f_t, f_t \rangle - 4 \langle \partial_x f_t, \partial_x f_t \rangle + 4 \langle U'''(x) \partial_y f_t, \partial_y f_t \rangle .$$

It brings the coercivity in position, which is missing in F_t' . However it is known (cf. [47], [75]) that no linear combination of F_t , F_t' , F_t'' and F_t''' can be non-positive for every $f \in L^2(\mu)$. In the particular cases treated in [47] and [75], G_t the norm of the gradient in space appears

naturally, thanks to Lemma 3.4. Indeed the smaller eigenvalue of $\partial_y^* \partial_y$ on $V^\perp = (\text{Ker} \partial_y^* \partial_y)^\perp$ is 1 (Poincaré inequality for the gaussian distribution) and thus

$$F'_t = 2 \langle K f_t, f_t \rangle = -2 \|\partial_y f_\perp\|^2 \leq -2 \|f_\perp\|^2.$$

On the other hand, Assumption 3.5 and Lemma 3.4 ensure $G_t \geq c_U \|f_V\|^2$ and lead to

$$\frac{1}{F_t} \left(G_t - \frac{c_U}{2} F'_t \right) \geq c_U. \quad (3.6)$$

Finally in order to close the differential inequality we need the first derivative of G_t (see Appendix for the proof) :

Lemma 3.7.

$$G'_t = \langle (-2RR^* + 2RU'') f_t, f_t \rangle + 2G_t.$$

We can now find a linear combination of F_t , G_t and their derivatives which is always non-positive. The terms in $\|f_V\|^2$ will be controlled by F_t'''' , the ones in $\|f_\perp\|^2$ by F'_t and G'_t .

Lemma 3.8. *Let $A \in \mathbb{R}$ and $\beta, k > 0$. Under Assumption 3.5, there exists $\tau_* \in \mathbb{R}$ such that for all $\tau \geq \tau_*$*

$$Q_3(\partial_t)F_t + Q_1(\partial_t)G_t \leq 0$$

with

$$Q_1(X) = 2\beta \left(X + 2\left(\frac{1}{\beta} - 1\right) - \frac{k}{2c_U\beta} \right)$$

$$Q_3(X) = X^3 + AX^2 + \tau X + k.$$

Démonstration. The above computations (Lemma 3.3 and 3.6) allow to write, for any $A \in \mathbb{R}$,

$$\begin{aligned} F_t'''' + AF_t'' &= \langle \left((2K)^3 + A(2K)^2 + 4[[K, R], R - R^*] + 12[K^2, R] + 4A[K, R] \right) f_t, f_t \rangle \\ &= \langle \left((2K)^3 + A(2K)^2 - 4[R, R^*] + 12R(1 + 2K) + 4AR \right) f_t, f_t \rangle \\ &= \langle \left((2K)^3 + A(2K)^2 - 4U''K - 4\partial_x^* \partial_x + 4R(3 + A + 6K) \right) f_t, f_t \rangle \\ &= \langle \left((2K)^3 + A(2K)^2 - 4U''K + 4R(3 + A + 6K) \right) f_t, f_t \rangle - 4G_t \end{aligned}$$

The operator $R(6 + 6(2K) + 2A)$ is annoying because, as a quadratic form on $L^2(\mu)$, it is neither positive nor non-positive; we'll give for it a not so subtle upper bound by the Cauchy-Schwarz and $2ab \leq a^2 + b^2$ inequalities with the sum of a term RR^* to be controlled by G'_t and of a term only acting on V^\perp , controled by F'_t .

More precisely, remark that $R = R\pi_\perp$ and furthermore that π_\perp commutes with the self-ajoint operators Id , K and $U''(x)$ which stabilize V (and so V^\perp too). Thus, for any $\beta > 0$,

$$\begin{aligned} &\langle (4R\pi_\perp(3 + A + \beta U'' + 6K)) f_t, f_t \rangle \\ &= 2 \langle (3 + A + \beta U'' + 6K) f_\perp, (2R^*) f_t \rangle \\ &\leq 4\beta \langle RR^* f_t, f_t \rangle + \frac{1}{\beta} \langle (3 + A + \beta U'' + 6K)^2 f_\perp, f_\perp \rangle. \end{aligned}$$

We obtain, taking into account lemma 3.7,

$$\begin{aligned}
F_t''' + AF_t'' + 2\beta G_t' + 4(1 - \beta)G_t &= \\
&< ((2K)^3 + A(2K)^2 - 4U''K + 4R(3 + A + 6K) - 4\beta RR^* + 4\beta RU'') f_t, f_t > \\
&= < ((2K)^3 + A(2K)^2 - 4U''K - 4\beta RR^*) f_t, f_t > + < 4R(3 + A + \beta U'' + 6K) f_t, f_t > \\
&\leq < ((2K)^3 + A(2K)^2 - 4U''K + \frac{1}{\beta}(3 + A + \beta U'' + 6K)^2) f_\perp, f_\perp > \\
&= < \left((2K)^3 + \left(A + \frac{9}{\beta} \right) (2K)^2 + \left(\frac{18 + 6A}{\beta} + 4U'' \right) (2K) + \frac{1}{\beta} (3 + A + \beta U'')^2 \right) f_\perp, f_\perp >.
\end{aligned}$$

Now we want to replace the terms with U'' by something that does not depend on x (under Assumption 3.5).

$$\begin{aligned}
< U''K f_t, f_t > &= - \iint U''(x) (\partial_y f_t(x, y))^2 \mu(dx, dy) \\
&\leq < (\min U'')K f_t, f_t >
\end{aligned}$$

and

$$< (3 + A + \beta U'')^2 f_\perp, f_\perp > \leq \|3 + A + \beta U''\|_\infty^2 \|f_\perp\|^2.$$

So by denoting

$$P(X) = X^3 + \left(A + \frac{9}{\beta} \right) X^2 + \left(\frac{18 + 6A}{\beta} + 4 \min U'' \right) X + \frac{1}{\beta} \|3 + A + \beta U''\|_\infty^2,$$

the previous computation leads to

$$F_t''' + AF_t'' + 2\beta G_t' + 4(1 - \beta)G_t \leq < P(2K) f_\perp, f_\perp >.$$

The eigenvalues of $2K$ on V^\perp being the $-2n$ for $n \in \mathbb{Z}_+$ (the eigenvectors are the so-called Hermite polynomials), consider any $k \geq 0$ and

$$\tau_k = \max_{n \geq 1} \frac{P(-2n) + k}{2n} \quad (< \infty)$$

so that $P(2K) + 2\tau K + k$ gets to be a non-positive bilinear form on V^\perp for all $\tau \geq \tau_k$, in other words

$$< P(2K) f_\perp, f_\perp > \leq -\tau F_t' - k \|f_\perp\|^2.$$

On the other hand $G_t \geq c_U \|f_V\|^2$ (cf. Lemma 3.4) so that

$$F_t''' + AF_t'' + 2\beta G_t' + \left(4(1 - \beta) - \frac{k}{c_U} \right) G_t + \tau F_t' + k F_t \leq 0.$$

□

Now it remains to get rid of G_t thanks to (3.6), and to find a common root for Q_1 and Q_3 in order for inequation (3.4) to hold.

Theorem 3.9. *Under assumption 3.5, there exist $\lambda, \eta > 0$, and $t \mapsto \nu_t \geq \nu_* \in \mathbb{R}$ with $\mathcal{R}e(\eta - \sqrt{-\nu_*}) > 0$ such that*

$$(\partial_t + \lambda) \left[(\partial_t + \eta)^2 + \nu_t \right] F_t \leq 0.$$

Démonstration. Let $A \in \mathbb{R}$ and $\beta \in (0, 1]$. Let $k \in [0, 4c_U(1 - \beta)]$, so that the root of

$$Q_1(X) = 2\beta \left(X + 2\left(\frac{1}{\beta} - 1\right) - \frac{k}{2c_U\beta} \right)$$

is zero for $k = 4c_U(1 - \beta)$ and negative otherwise. Let τ_k be given by Lemma 3.8; we choose $\tau \geq \tau_{4c_U(1-\beta)}$ large enough such that for all $k \geq 0$

$$Q_3(X) = X^3 + AX^2 + \tau X + k$$

has only one non-positive root, which is continuous with respect to k . This root is zero for $k = 0$, negative otherwise. Thus by continuity there exists a $k \in [0, 4c_U(1 - \beta)]$ such that Q_1 and Q_3 have a common root. We call this root $-\lambda$. Now Lemma 3.8 can be rewritten, with some constant $u, v, w \in \mathbb{R}$,

$$\begin{aligned} 0 &\geq (\partial_t + \lambda) (F_t'' + uF_t' + vF_t + wG_t) \\ &= (\partial_t + \lambda) \left(F_t'' + \left(u + w\frac{c_U}{2} \right) F_t' + \left(v + w \left(G_t - \frac{c_U}{2} F_t' \right) \frac{1}{F_t} \right) F_t \right). \end{aligned}$$

Let $\eta = \frac{2u+wc_U}{4}$, $\nu_* = v + wc_U - \eta^2$ and

$$\nu_t = \left(v + w \left(G_t - \frac{c_U}{2} F_t' \right) \frac{1}{F_t} \right) - \eta^2.$$

Inequation (3.6) exactly means $\nu_t \geq \nu_*$. It remains to show that, for some parameters, $-\lambda$ and $\mathcal{R}e(-\eta \pm \sqrt{-\nu_*})$ are negative. These are the real parts of the roots of

$$Q_3(X) + \frac{c_U}{2} X Q_1(X) + c_U Q_1(X) = X^3 + (A + \beta c_U) X^2 + \left(\tau + 2c_U - \frac{k}{2} \right) X + 2c_U(1 - \beta).$$

Take $\beta = 1$, $A > -c_U$, τ large enough and $k = 0$. Then zero is a common root and

$$Q_3(X) + \frac{c_U}{2} X Q_1(X) + c_U Q_1(X) = X \left(X^2 + (A + \beta c_U) X + (\tau + 2c_U) \right),$$

so that $\lambda = 0$. $X^2 + (A + \beta c_U) X + (\tau + 2c_U)$ has positive coefficients so if it has real roots, they are negative. Otherwise $\mathcal{R}e(\eta - \sqrt{-\nu_*}) = \eta = \frac{1}{2}(A + \beta c_U) > 0$. Now if β is chosen slightly less than 1 and k is such that Q_1 and Q_3 still have a common root, relying again on continuity, we still have $\mathcal{R}e(\eta - \sqrt{-\nu_*}) > 0$ but $-\lambda$ becomes a real root of a polynomial with positive coefficients and thus is negative. \square

3.2.2 The RTP

This section is a replica of the previous one. From Lemma 3.10 to Theorem 3.15, the generator is

$$Lf(x, y) = y\partial_x f(x, y) + a(x, y) (f(x, -y) - f(x, y)). \quad (3.3)$$

As in the kFP case we compute the derivatives of F_t and G_t , proceed with a differential equation and conclude with a particular choice of the parameters which are in parties to the above approach. First, the invariant measure has to be explicitied :

Lemma 3.10. *The unique (up to a constant) invariant measure of the telegraph model is*

$$\mu = e^{-U(x)} dx \otimes \frac{1}{2} (\delta_1 + \delta_{-1}) (dy),$$

where

$$U'(x) = a(x, 1) - a(x, -1).$$

Démonstration. Note that $yU'(x) = a(x, y) - a(x, -y)$. We check

$$\begin{aligned} \mu Lf &= \iint y\partial_x f(x, y) e^{-U(x)} dx dy + \iint a(x, y) (f(x, -y) - f(x, y)) e^{-U(x)} dx dy \\ &= \iint yf(x, y) U'(x) e^{-U(x)} dx dy + \iint (a(x, -y) - a(x, y)) f(x, y) e^{-U(x)} dx dy \\ &= 0 \end{aligned}$$

for all smooth $f \in L^2(\mu)$, so that μ is invariant. In the other hand, the process is clearly irreducible (from any point X_0 it can reach any ball in finite time with positive probability) and aperiodic (X_t can go back to X_0 at an arbitrarily small time s with positive probability) and uniqueness of its invariant probability follows. \square

We note $f_-(x, y) = f(x, -y)$ and remark that

$$V = \{f \in L^2(\mu), f_- = f\} \quad V^\perp = \{f \in L^2(\mu), f_- = -f\}$$

and, keeping the previous notation π_V and π_\perp (or f_V and f_\perp) for the orthogonal projections on V and V^\perp ,

$$\pi_\perp f = \frac{f - f_-}{2} \quad \pi_V f = \frac{f + f_-}{2} = f_V.$$

Thus $yV = V^\perp$ and $yV^\perp = V$, and more precisely

$$\pi_\perp y = y\pi_V \quad \pi_V y = y\pi_\perp.$$

Now recall that $\partial_x^* = \partial_x - U'$ and define

$$\begin{aligned} Kf &= -(a + a_-)\pi_\perp f \\ Rf &= y(\partial_x - U')\pi_\perp f \\ &= -\pi_V y\partial_x^* \pi_\perp f. \\ R^*f &= -\pi_\perp y\partial_x \pi_V f \\ &= -y\partial_x \left(\frac{f + f_-}{2} \right). \end{aligned}$$

Then

$$\begin{aligned}
K + R - R^* &= -(a + a_-)\pi_\perp + y(\partial_x - U')\pi_\perp + y\partial_x\pi_V \\
&= (-a - a_- - yU')\pi_\perp + y\partial_x(\pi_V + \pi_\perp) \\
&= -2a\pi_\perp + y\partial_x \\
&= L.
\end{aligned}$$

Note that $a + a_-$ and U' do not depend on y and so, seen as self-adjoint operators on $L^2(\mu)$, they commute with π_\perp and π_V . In particular this gives $K^* = K$.

Now thanks to these considerations we can compute the following brackets (see the appendix for details).

Lemma 3.11.

$$\begin{aligned}
[K, R] &= R(a + a_-) \\
[K^2, R] &= -R(a + a_-)^2 \\
[[K, R], R - R^*] &= R^*R(a + a_-) - R(a + a_-)R^* \\
RR^* &= \partial_x^*\partial_x\pi_V \\
R^*R &= (\partial_x^*\partial_x + U'')\pi_\perp.
\end{aligned}$$

From now on we make the following assumptions :

Assumption 3.12. *The rates $a(\cdot, 1)$ and $a(\cdot, -1)$ are positive, smooth, bounded and with bounded derivatives, and*

$$a_* := \min_{\mathbb{R}} (a(\cdot, 1) + a(\cdot, -1)) > 0.$$

Furthermore $U(x) = \int_0^x (a(z, 1) - a(z, -1)) dz$ satisfies Assumption 3.5.

Then, again F_t is controlled by G_t and F'_t . Indeed, Under Assumption 3.12,

$$F'_t = 2 \langle Kf_t, f_t \rangle = -2 \langle (a + a_-)f_\perp, f_\perp \rangle \leq -2a_* \|f_\perp\|^2$$

and (Lemma 3.4),

$$G_t = \|\partial_x f\|^2 \geq \|\partial_x f_V\|^2 \geq c_U \|f_V\|^2$$

so that

$$\frac{1}{F_t} \left(G_t - \frac{c_U}{2a_*} F'_t \right) \geq c_U. \quad (3.7)$$

We will also need the derivative of G_t , computed in the appendix :

Lemma 3.13.

$$G'_t = \langle 2(RU'' - U''K - R^*R(a + a_-)) f_t, f_t \rangle.$$

We are now ready to prove a result similar to Lemma 3.8

Lemma 3.14. *Under Assumption 3.12, there exist polynomials \tilde{Q}_1 and \tilde{Q}_3 respectively of first and third order such that*

$$\tilde{Q}_3(\partial_t)F_t + \tilde{Q}_1(\partial_t)G_t \leq 0$$

Démonstration. Lemma 3.3 and 3.11 give, for any $A \in \mathbb{R}$,

$$\begin{aligned} F_t''' + AF_t'' &= \\ &< \left[(2K)^3 + A(2K)^2 + 4R(-3(a+a_-)^2 + A(a+a_-)) \right. \\ &\quad \left. + 4R^*R(a+a_-) - 4R(a+a_-)R^* \right] f_t, f_t >. \end{aligned}$$

We consider any $h \geq 0$ and write

$$\begin{aligned} F_t''' + AF_t'' + 2(1+h)G_t' &= \\ &< \left[(2K)^3 + A(2K)^2 - 4(1+h)U''K - 4hR^*R(a+a_-) - 4R(a+a_-)R^* \right. \\ &\quad \left. + 4R(-3(a+a_-)^2 + A(a+a_-) + (1+h)U'') \right] f_t, f_t > \end{aligned}$$

Now for the extra $-4hR^*R(a+a_-)$, for any $\alpha \in (0, 1]$, *via* the Cauchy-Schwarz inequality,

$$\begin{aligned} -4h &< R^*R(a+a_-)f_t, f_t > \\ &= -4h < \partial_x(a+a_-)f_\perp, \partial_x f_\perp > \\ &= -4h < (a+a_-)\partial_x f_\perp, \partial_x f_\perp > -4h < (\partial_x(a+a_-))f_\perp, \partial_x f_\perp > \\ &= -4h < (a+a_-)\partial_x f_\perp, \partial_x f_\perp > -4h < (a+a_-)^{-\frac{1}{2}}(\partial_x(a+a_-))f_\perp, (a+a_-)^{\frac{1}{2}}\partial_x f_\perp > \\ &\leq -4ha_*(1-\alpha)\|\partial_x f_\perp\|^2 + \frac{h}{\alpha}\|(a+a_-)^{-\frac{1}{2}}(\partial_x(a+a_-))f_\perp\|^2. \end{aligned}$$

Then following again the steps of Lemma 3.8, for any $\beta > 0$, we bound

$$\begin{aligned} &< 4R(-3(a+a_-)^2 + A(a+a_-) + (1+h)U'') f_t, f_t > \\ &= 2 < (-3(a+a_-)^2 + A(a+a_-) + (1+h)U'') (a+a_-)^{-\frac{1}{2}}f_\perp, 2(a+a_-)^{\frac{1}{2}}R^*f_t > \\ &\leq \frac{1}{\beta}\|(-3(a+a_-)^2 + A(a+a_-) + (1+h)U'') (a+a_-)^{-\frac{1}{2}}f_\perp\|^2 + 4\beta < R(a+a_-)R^*f_t, f_t > \end{aligned}$$

Gathering all this, and recalling $K = -(a+a_-)\pi_\perp$ we get

$$\begin{aligned} F_t''' + AF_t'' + 2(1+h)G_t' &\leq \\ &< \left[-8(a+a_-)^3 + 4A(a+a_-)^2 + \frac{(-3(a+a_-)^2 + A(a+a_-) + (1+h)U'')^2}{\beta(a+a_-)} \right. \\ &\quad \left. + 4(1+h)U''(a+a_-) \right] f_\perp, f_\perp > -4ha_*(1-\alpha)\|\partial_x f_\perp\|^2 \\ &\quad + \frac{h}{\alpha}\left\| \frac{\partial_x(a+a_-)}{\sqrt{a+a_-}} f_\perp \right\|^2 - 4 < R(a+a_-)(1-\beta)R^*f_t, f_t >. \end{aligned}$$

For the last term, as long as $\beta \leq 1$,

$$\begin{aligned} -4 < R(a+a_-)(1-\beta)R^*f_t, f_t > &= -4 < (a+a_-)(1-\beta)\partial_x f_V, \partial_x f_V > \\ &\leq -4a_*(1-\beta)\|\partial_x f_V\|^2. \end{aligned}$$

Choose $\alpha < 1$, let $k \in [0, 4c_U(1-\beta)]$ and define h_k such that

$$-4h_k(1-\alpha) = -4(1-\beta) + \frac{k}{a_*c_U}.$$

We will note

$$\lambda_k = a_* \frac{2h_k(1-\alpha)}{(1+h_k)}$$

and define the function

$$\begin{aligned} H = & -8(a+a_-)^3 + 4A(a+a_-)^2 + \frac{(-3(a+a_-)^2 + A(a+a_-) + (1+h_k)U'')^2}{\beta(a+a_-)} \\ & + \frac{h_k(\partial_x(a+a_-))^2}{\alpha(a+a_-)} + 4(1+h_k)U''(a+a_-), \end{aligned}$$

so that everything comes down to

$$F_t''' + AF_t'' + 2(1+h_k)(G_t' + \lambda_k G_t) \leq \langle Hf_\perp, f_\perp \rangle - \frac{k}{c_U} \|\partial_x f_V\|^2.$$

Under Assumption 3.12, in one hand Lemma 3.4 gives $\|\partial_x f_V\|^2 \geq c_U \|f_V\|^2$ and, in the other hand H is bounded; so there exists τ_k such that

$$H - \tau_k(a+a_-) + k \leq 0.$$

Thus, for all $\tau \geq \tau_k$,

$$\langle Hf_\perp, f_\perp \rangle \leq -\tau F_t' - k \|f_\perp\|^2.$$

Finally we get

$$F_t''' + AF_t'' + \tau F_t' + k F_t + 2(1+h_k)(G_t' + \lambda_k G_t) \leq 0.$$

□

Here ends the proof that (3.4) is satisfied for the telegraph model :

Theorem 3.15. *Under assumption 3.5, there exist $\lambda, \eta > 0$, and $t \mapsto \nu_t \geq \nu_* \in \mathbb{R}$ with $\operatorname{Re}(\eta - \sqrt{-\nu_*}) > 0$ such that*

$$(\partial_t + \lambda) \left[(\partial_t + \eta)^2 + \nu_t \right] F_t \leq 0.$$

Démonstration. We keep the notations used in the proof of Lemma 3.14; our purpose is to find some parameters for wich

$$\tilde{Q}_1(X) = 2(1+h_k)(X + \lambda_k)$$

and

$$\tilde{Q}_3(X) = X^3 + AX^2 + \tau X + k$$

have a common root. Let $\beta \in (0, 1]$, $\alpha \in (0, 1)$, $A \in \mathbb{R}$ be fixed, we let k evolve in $[0, 4c_U(1-\beta)]$. The root of \tilde{Q}_1 ,

$$-\lambda_k = -a_* \frac{2(1-\beta) - \frac{k}{2c_U}}{(1+h_k)}$$

is zero for $k = 4c_U(1 - \beta)$, else negative. We take $\tau \geq \tau_{4c_U(1-\beta)}$ large enough so that, for any $k \in [0, 4c_U(1 - \beta)]$, \tilde{Q}_3 has a unique non-positive real root, which is continuous with respect to k . This real root is zero for $k = 0$ and negative otherwise, thus by continuity there exists a $k \geq 0$ such that \tilde{Q}_3 and \tilde{Q}_1 have a common root. We call $-\lambda$ this root and consider $u, v, w \in \mathbb{R}$ such that

$$\begin{aligned}\tilde{Q}_3(\partial_t)F_t + \tilde{Q}_1(\partial_t)G_t &= (\partial_t + \lambda)(F_t'' + uF_t' + vF_t + wG_t) \\ &= (\partial_t + \lambda)\left(F_t'' + \left(u + w\frac{c_U}{2a_*}\right)F_t' + \left(v + w\left(G_t - \frac{c_U}{2a_*}F_t'\right)\frac{1}{F_t}\right)F_t\right).\end{aligned}$$

Let $\eta = \frac{2u+a_*^{-1}c_U}{4}$, $\nu_* = v + wc_U - \eta^2$ and

$$\nu_t = \left(v + w\left(G_t - \frac{c_U}{2a_*}F_t'\right)\frac{1}{F_t}\right) - \eta^2.$$

(3.7) exactly gives $\nu_t \geq \nu_*$. It remains to find some parameters for which $-\lambda$ and $\mathcal{R}e(-\eta \pm \sqrt{-\nu_*})$ are negative. These are the real parts of the roots of

$$\tilde{Q}_3(X) + \frac{c_U}{2a_*}X\tilde{Q}_1(X) + c_U\tilde{Q}_1(X).$$

Take $A > -c_Ua_*^{-1}$, $\alpha \in (0, 1)$, $\beta = 1$, τ large enough and $k = 0$, then $h_k = \lambda_k = 0$ and zero is a common root. Thus $\tilde{Q}_1(X) = 2X$ and

$$\tilde{Q}_3(X) + \frac{c_U}{2a_*}X\tilde{Q}_1(X) + c_U\tilde{Q}_1(X) = X\left(X^2 + \left(A + \frac{c_U}{a_*}\right)X + \tau + 2c_U\right).$$

If $X^2 + (A + c_Ua_*^{-1})X + \tau + 2c_U$, polynomial with positive coefficients, has real roots, they are negative, and else $\mathcal{R}e(\eta \pm \sqrt{-\nu_*}) = \eta = \frac{1}{2}(A + c_Ua_*^{-1}) > 0$. Now if β is slightly less than 1 this is still the case by continuity, but then $-\lambda$ is a real root of a polynomial with positive coefficient so $\lambda > 0$. □

3.3 Numerical studies

Although Theorem 3.1 gives “explicit” bounds for the rate of convergence to equilibrium, it is not very easy to handle, as we get a set of polynomials Π depending on several parameters in some set \mathcal{C} , such that the best rate obtained *via* Theorem 3.1 is

$$r = \max_{A, \beta, k, \tau \in \mathcal{C}} \min \{-\mathcal{R}e(\alpha), \alpha \text{ root of } \Pi(X)\}$$

Using the notations of Lemma 3.8 and 3.14,

$$\Pi(X) = Q_3(X) + c_U\left(\frac{1}{2}X + 1\right)Q_1(X)$$

for the kFP process, and

$$\Pi(X) = \tilde{Q}_3(X) + c_U\left(\frac{1}{2a_*}X + 1\right)\tilde{Q}_1(X)$$

for the telegraph one. \mathcal{C} is the set of parameters for which Q_1 and Q_3 have a common root and the inequality is proven, for instance in the kFP model one need

$$0 < \beta < 1, \quad A > c_U, \quad k > 0, \quad \tau \geq \max_{n \geq 1} \frac{P(-2n) + k}{2n}$$

(P defined in lemma 3.8). Nevertheless we can numerically deal with this and compare the obtained results with the theoretical rates when they are known, namely in the case of a quadratic potential for the kFP process (see [47]) and for the constant jump rate of the telegraph on the torus (see [75]). Obviously, such examples may just be considered as some benchmarks and are not really interesting processes for MCMC algorithm. As a consequence, once we will have seen the numerical rates can be of the right order of magnitude for the kFP model, we won't push this analysis deeper.

First of all, we adapt Lemma 3.8 in order to allow some changes in the parameters. The same computations lead to

Lemma 3.16. *Consider the generator*

$$L_{v,b,U} = v(by\partial_x - U'(x)\partial_y) + (\partial_y^2 - by\partial_y),$$

with invariant measure $e^{-U(x)}dx \otimes e^{-b\frac{y^2}{2}}$. Under Assumption 3.5, for any $A, k \in \mathbb{R}$ and $\beta > 0$ there exists $\tau \in \mathbb{R}$ such that

$$Q_3(\partial_t)F_t + Q_1(\partial_t)G_t \leq 0$$

with

$$Q_1(X) = v^2\beta \left(X + 2b\left(\frac{1}{\beta} - 1\right) - \frac{k}{2v^2c_U\beta} \right)$$

$$Q_3(X) = X^3 + AX^2 + \tau X + k$$

$$\tau \geq \max_{n \geq 1} \frac{P(-2bn) + k}{2bn}$$

$$P(X) = X^3 + \left(A + \frac{9b^2}{\beta} \right) X^2 + \left(6b^2 \frac{3b + A}{\beta} + v^2 \min((6b - 4)U'') \right) X + \frac{1}{\beta} (3b^2 + Ab + v^2\beta \|U''\|_\infty)^2.$$

On the other hand,

$$G_t - \frac{c_U}{2b} F'_t \geq c_U F_t$$

In a MCMC algorithm, U would be given while b^{-1} the variance of the invariant speed and v the ratio between the antisymmetric and symmetric parts of the dynamics should be chosen to get the fastest convergence to equilibrium (given the instantaneous randomness injected in the system).

The real exponential rate of convergence for $L_{v,b,U}$ with U'' constant is (see [47])

$$r_{theor} = 1 - \sqrt{\left(1 - \frac{4v^2U''}{b} \right)_+}.$$

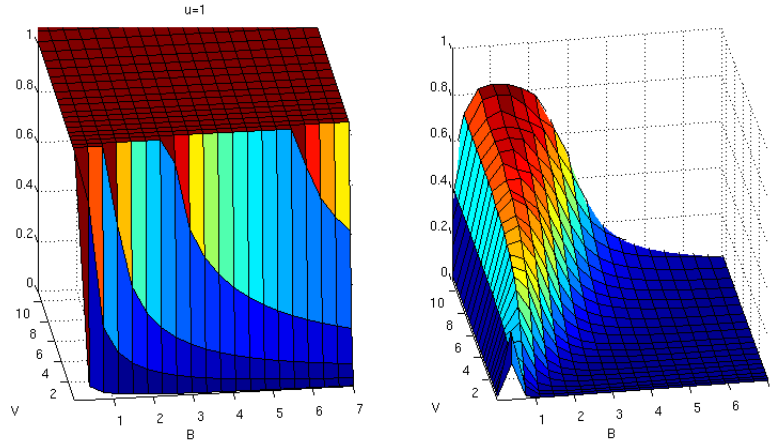


FIGURE 3.3 – Left : theoretical rate computed in [47]. Right : numerical rate given by Theorem 3.9. When v is small (*i.e.* the antisymmetric part of L is in a sense weak), the numerical rate is not very accurate and can miss the values for which the non reversible process is faster (asymptotically) than the reversible one. It becomes better with big values of v and for some b we get the right order of magnitude

Here are some numerical optimal rates given by Lemma 3.16 (to be compared to the theoretical one, in brackets if it is not 1) for $U'' = 1$ and different values of v and b (see also figure 3.3) :

$v \setminus b$	0.2	0.5	1	3	5
0.5	0.11	0.21	0.02	03.10^{-4} (0.18)	07.10^{-5} (0.10)
1	0.20	0.21	0.07	01.10^{-3}	02.10^{-4} (0.55)
2	0.24	0.26	0.23	06.10^{-3}	01.10^{-3}
10	0.26	0.32	0.56	0.23	0.03

3.4 Quelques remarques conclusives

Cette étude numérique, à $\|U''\|_\infty$ constante, ne souligne cependant pas le fait qu'à basse température (c'est-à-dire lorsque $\|U_\varepsilon''\|_\infty = \varepsilon^{-1}\|U''\|_\infty$ part à l'infini), nos estimations deviennent très mauvaises. En effet il n'y a aucune raison de penser que $\|U''\|_\infty$ soit un élément déterminant du taux de convergence : un potentiel super-quadratique à l'infini, plus coercif, devrait au contraire améliorer la convergence. D'ailleurs bien que l'hypothèse $\|U''\|_\infty < \infty$ se retrouve dans un certain nombre de travaux, elle n'est pas nécessaire pour obtenir un taux explicite de convergence (cf. [86, 38]). Une piste pour se débarrasser de cette hypothèse serait d'adapter la méthode des multiplicateurs proposée par Villani, p. 58 de [86].

Néanmoins, avant de poursuivre tête baissée dans cette piste de l'inégalité différentielle d'ordre 3, essayons d'en mesurer, sur ce premier essai, les intérêts et les défauts. Certes, la méthode aborde la description quantitative des oscillations telles que celles observées dans [45], ce qui était l'une des questions adressées par Villani dans l'acte de naissance de l'hypocoercivité [85]. Cependant, tenir compte des inhomogénéités temporelles comme nous l'avons fait n'a pas amélioré la précision de la méthode initiale de ce dernier basée sur une fonction de Lyapunov modifiée. On peut d'ailleurs considérer que notre méthode d'ordre 3 est une version compliquée

de la méthode de Villani, au sens où l'on a trouvé une équation différentielle $H_t' \leq -\lambda H_t$ avec $H_t = F_t'' + bF_t' + cF_t + dG_t$ qui est en fait la fonctionnelle de Villani avec quelques termes additionnels (simplement, au lieu d'ajouter "artificiellement" le terme croisé $\langle \partial_x f, \partial_y f \rangle$, on le récupère dans F_t'' ; mais on ajoute tout de même tout aussi artificiellement le terme G_t , quand bien même on s'en débarrasse ensuite). Les calculs sont plus lourds, et le résultat moins explicite (sous la forme de racines d'un polynôme de degré 3).

Plutôt que de persister dans l'étude des dérivées d'ordre supérieur, on va réinterpréter l'inégalité

$$(\partial_t + \lambda)(F_t'' + aF_t + bG_t)$$

comme une relation entre la dérivée seconde de $e^{\lambda t} \|\partial_y f_t\|^2$ et $e^{\lambda t} \|\partial_x f_t\|^2$. Distinguer ces deux normes permet de mesurer si l'on est plutôt à l'équilibre en position ou en vitesse. En ajoutant la variance, de dérivée $-2\|\partial_y f_t\|^2$, et l'inégalité de Poincaré, on a donc fermé un système différentiel entre ces quantités. On se rapproche alors des méthodes basés sur des estimations de gradient, les commutations gradient/semi-groupe telles que dans le critère Γ_2 ou les entrelacements de Bonnefont et Joulin ([18]), le récent travail de Kontis, Ottobre et Zegarliński ([60, 61]), ou encore les liens entre ces estimations de gradient et la contraction en distance de Wasserstein ([62, 3]).

Cette vision, ainsi que la volonté de continuer l'investigation sur d'éventuelles méthodes analytiques d'étude de la convergence de PDMP, ont été à l'origine du chapitre suivant.

3.5 Appendix

Proof of Lemma 3.3 :

Démonstration. From $\partial_t f_t = Lf_t$ comes

$$\begin{aligned} F_t' &= 2 \langle Lf_t, f_t \rangle \\ &= \langle (2K)f_t, f_t \rangle \\ F_t'' &= 2 \langle L^2 f_t, f_t \rangle + 2 \langle Lf_t, Lf_t \rangle \end{aligned}$$

From $(L + L^*)L = 2KL$ we compute

$$\begin{aligned} F_t'' &= \langle (2K)^2 f_t, f_t \rangle + 4 \langle KRf_t, f_t \rangle - 4 \langle KR^* f_t, f_t \rangle \\ &= \langle (2K)^2 f_t, f_t \rangle + 4 \langle [K, R]f_t, f_t \rangle \\ F_t''' &= \langle (2K)^2 Lf_t, f_t \rangle + \langle L^*(2K)^2 f_t, f_t \rangle + 4 \langle [K, R]Lf_t, f_t \rangle + 4 \langle L^*[K, R]f_t, f_t \rangle . \end{aligned}$$

Concerning the first two terms,

$$\begin{aligned} (2K)^2 L + L^*(2K)^2 &= (2K)^3 + (2K)^2(R - R^*) - (R - R^*)(2K)^2 \\ &= (2K)^3 + 4 \left([K^2, R] - [K^2, R^*] \right) \end{aligned}$$

and $[K^2, R^*]^* = -[K^2, R]$. For the last two terms, $[K, R]K + K[K, R] = [K^2, R]$ thus

$$4 \langle [K, R]Lf_t, f_t \rangle + 4 \langle L^*[K, R]f_t, f_t \rangle = 4 \langle [K^2, R]f_t, f_t \rangle + 4 \langle [[K, R], R - R^*] f_t, f_t \rangle .$$

□

Proof of Lemma 3.6 :

Démonstration. First ∂_x and ∂_x^* commute with ∂_y and ∂_y^* , and

$$\partial_x \partial_x^* = \partial_x^* \partial_x + U'', \quad \partial_y \partial_y^* = \partial_y^* \partial_y + 1.$$

Now

$$[K, R] = \partial_y^* \partial_y \partial_x^* \partial_y - \partial_x^* \partial_y \partial_y^* \partial_y = (\partial_y^* \partial_y - \partial_y \partial_y^*) \partial_x^* \partial_y = R$$

As well,

$$\begin{aligned} [K^2, R] &= -(\partial_y^* \partial_y)^2 \partial_x^* \partial_y + \partial_x^* \partial_y (\partial_y^* \partial_y)^2 \\ &= -\partial_x^* (\partial_y \partial_y^* - 1)^2 \partial_y + \partial_x (\partial_y \partial_y^*)^2 \partial_y \\ &= 2\partial_x^* \partial_y \partial_y^* \partial_y - \partial_x^* \partial_y \\ &= R(2K + 1). \end{aligned}$$

Finally

$$\begin{aligned} [R, R^*] &= \partial_x^* \partial_y \partial_y^* \partial_x - \partial_x \partial_y^* \partial_x^* \partial_y \\ &= \partial_x^* \partial_x (\partial_y^* \partial_y + 1) - (\partial_x^* \partial_x + U'') \partial_y^* \partial_y \\ &= -U'' \partial_y^* \partial_y + \partial_x^* \partial_x \\ &= U'' K + \partial_x^* \partial_x. \end{aligned}$$

□

Proof of Lemma 3.7 :

Démonstration.

$$\begin{aligned} G'_t &= 2 \langle \partial_x K f_t, \partial_x f_t \rangle + 2 \langle \partial_x (R - R^*) f_t, \partial_x f_t \rangle \\ &= -2 \langle \partial_x^* \partial_x \partial_y^* \partial_y f, f_t \rangle + 2 \langle [\partial_x^* \partial_x, R] f_t, f_t \rangle. \end{aligned}$$

For the first term

$$\partial_x^* \partial_x \partial_y^* \partial_y = \partial_x^* (\partial_y \partial_y^* - 1) \partial_x = RR^* - \partial_x^* \partial_x,$$

and for the second one

$$\begin{aligned} \partial_x^* \partial_x R &= -\partial_x^* \partial_x \partial_x^* \partial_y \\ &= -\partial_x^* (\partial_x^* \partial_x + U'') \partial_y \\ &= R \partial_x^* \partial_x + RU''. \end{aligned}$$

Finally,

$$\begin{aligned} G'_t &= \langle (-2RR^* + 2\partial_x^* \partial_x + 2RU'') f_t, f_t \rangle \\ &= \langle (-2RR^* + 2RU'') f_t, f_t \rangle + 2 \langle \partial_x f_t, \partial_x f_t \rangle. \end{aligned}$$

□

Proof of Lemma 3.11 :

Démonstration.

$$[K, R] = (a + a_-)\pi_\perp \pi_V y \partial_x^* \pi_\perp - \pi_V y \partial_x^* \pi_\perp (a + a_-)\pi_\perp = 0 + R(a + a_-)$$

The same computation holds for the second one :

$$\begin{aligned} [K^2, R] &= -((a + a_-)\pi_\perp)^2 \pi_V y \partial_x^* \pi_\perp + \pi_V y \partial_x^* \pi_\perp (a + a_-)^2 \pi_\perp \\ &= 0 - R(a + a_-)^2 \end{aligned}$$

Since $(a + a_-)$ commutes with π_V , a term $\pi_V \pi_\perp$ appears in $R(a + a_-)R = R^2(a + a_-) = 0$ we have

$$\begin{aligned} [[K, R], R - R^*] &= [R(a + a_-), -R^*] \\ &= R^*R(a + a_-) - R(a + a_-)R^*. \end{aligned}$$

Finally, as $\pi_\perp y \pi_V = y \pi_V$ and $y^2 = 1$ we get

$$RR^* = (y \partial_x^* \pi_\perp) (\pi_\perp \partial_x y \pi_V) = \partial_x^* \partial_x \pi_V$$

and similarly, as $\pi_V y \pi_\perp = y \pi_\perp$,

$$R^*R = \partial_x \partial_x^* \pi_\perp = (\partial_x^* \partial_x + U'')\pi_\perp.$$

□

Proof of Lemma 3.13 :

Démonstration.

$$\begin{aligned} G'_t &= 2 \langle \partial_x^* \partial_x L f_t, f_t \rangle \\ &= 2 \langle \partial_x^* \partial_x R f_t, f_t \rangle - 2 \langle \partial_x^* \partial_x R^* f_t, f_t \rangle - 2 \langle \partial_x^* \partial_x (a + a_-)\pi_\perp f_t, f_t \rangle \\ &= 2 \langle [\partial_x^* \partial_x, R] f_t, f_t \rangle - 2 \langle \partial_x^* \partial_x (a + a_-)\pi_\perp f_t, f_t \rangle. \end{aligned}$$

For the first term,

$$[\partial_x^* \partial_x, R] = -\partial_x^* \partial_x \pi_V y \partial_x^* \pi_\perp + \pi_V y \partial_x^* \pi_\perp \partial_x^* \partial_x = R[\partial_x, \partial_x^*] = RU''$$

We conclude, according to Lemma 3.11, with $\partial_x^* \partial_x \pi_\perp = R^*R - U''\pi_\perp$.

□

Chapitre 4

PDMP contractifs

By default, a jump is a true jump.

Laurent Miclo

4.1 Contexte et stratégie générale

Ce chapitre étudie la convergence à l'équilibre de la loi de $(X_t)_{t \geq 0}$, un processus de Markov déterministe par morceaux (PDMP) sur \mathbb{R}^d . Ces processus hybrides, qui satisfont une équation différentielle déterministe entre des sauts aléatoires, ont récemment vu l'intérêt à leur égard grandir dans un certain nombre de domaines : on rappelle la référence [7] pour un survol de quelques considérations récentes sur le sujet. L'ergodicité de ces processus et, le cas échéant, la vitesse de convergence à l'équilibre sont particulièrement étudiés. Sur ce dernier point les méthodes de couplages se sont récemment avérées efficaces pour démontrer des taux explicites de convergence en distance de Wasserstein (voir par exemple [27, 10, 15, 46, 19], parmi bien d'autres). Au contraire l'usage de techniques analytiques d'inégalités fonctionnelles était plutôt restreint. Nous allons donc nous attacher à combler cette lacune, ce qui nous fournira l'occasion de confronter les idées développées en conclusion du chapitre précédent à un modèle plus simple que le RTP (qui est, outre un PDMP, un processus cinétique).

Soit Ω un ouvert de \mathbb{R}^d . La dynamique est définie à l'aide d'un champ de vecteurs $b : \Omega \rightarrow \mathbb{R}^d$, d'un taux de saut $\lambda : \Omega \rightarrow \mathbb{R}_+$, et d'un noyau de transition Q qui sera indistinctement considéré comme une fonction de Ω vers $\mathcal{P}(\Omega)$ l'ensemble des mesures de probabilités sur Ω , ou comme un opérateur sur $\mathcal{C}(\Omega)$. Pour $x \in \Omega$ et $t > 0$, on note $\varphi_x(t)$ le flot associé à b , c'est-à-dire la solution de

$$\partial_t \varphi_x(t) = b(\varphi_x(t)), \quad \varphi_x(0) = x.$$

Partant d'un point x , le processus $(X_t)_{t \geq 0}$ suit de façon complètement déterministe le flot jusqu'à son premier temps de saut T_x , de loi

$$\mathbb{P}(T_x < s) = \int_0^s \lambda(\varphi_x(u)) e^{-\int_0^u \lambda(\varphi_x(w)) dw} du = 1 - e^{-\int_0^s \lambda(\varphi_x(w)) dw}.$$

À l'instant T_x , le processus saute selon la loi $Q(\phi_x(T_x))$, et reprend sa dynamique déterministe à partir de sa nouvelle position. Le générateur infinitésimal du processus est

$$Lf(x) = b(x) \cdot \nabla f(x) + \lambda(x) (Qf(x) - f(x)), \quad (4.1)$$

défini pour $f \in \mathcal{C}(\Omega)$ régulière. On note toujours

$$P_t f(x) = \mathbb{E}(f(X_t) | X_0 = x)$$

le semi-groupe associé. Sauf mention contraire, on fait dans ce chapitre les hypothèses suivantes :

- L'ouvert Ω est préservé par le flot : si $x \in \Omega$ alors $\varphi_x(t) \in \Omega$ pour tout $t > 0$.
- Le processus est non-explosif : il ne peut pas y avoir une infinité de sauts en un temps fini, de sorte que le processus est défini pour tout temps.
- Les fonctions λ et b sont régulières; on note $J_b(x) = [\partial_i b_k(x)]_{1 \leq i, k \leq d}$ la jacobienne de $b = (b_k)_{1 \leq k \leq d}$.
- Le processus admet une unique mesure invariante μ , et P_t est ergodique au sens où il converge faiblement vers μ quand t tend vers l'infini. De plus, et comme on l'a toujours supposé dans cette thèse, il existe une algèbre \mathcal{A} de fonctions régulières, stable par L et $(P_t)_{t \geq 0}$, qui est dense dans $L^2(\mu)$.

En réalité on aura besoin de plus de propriétés sur \mathcal{A} pour autoriser les manipulations à venir sur ses éléments, mais nous allons garder l'étude à un niveau formel et supposer que nos manipulations sont licites.

Ces hypothèses nous permettent de nous concentrer sur la quantification de l'ergodicité. Notons que l'unicité de la mesure invariante et l'ergodicité du processus peuvent souvent, dans les applications concernées, être prouvées en montrant que le processus est irréductible et admet une fonction de Lyapunov (cf. [41]).

Rappelons qu'en introduction (section 1.3.4), nous avons défini le carré du champ $\Gamma(f) = \frac{1}{2}L(f^2) - fLf$, ainsi que l'opérateur

$$\Gamma_2 = \frac{1}{2}L(\Gamma f) - \Gamma(f, Lf).$$

Nous avons vu que le critère de courbure de Bakry-Emery $\Gamma_2 > \rho\Gamma$ était équivalent à la sous-commutation

$$\Gamma(P_t f) \leq e^{-2\rho t} P_t \Gamma f. \tag{4.2}$$

Par exemple, pour le processus d'Ornstein-Uhlenbeck de générateur

$$L_{OU} f(x) = \Delta f(x) - \rho x \cdot \nabla f(x),$$

cela se lit

$$|\nabla P_t f|^2 \leq e^{-2\rho t} P_t |\nabla f|^2, \tag{4.3}$$

où $|\cdot|$ est la norme euclidienne sur \mathbb{R}^d . Néanmoins le critère de Bakry-Emery ne sera en général pas satisfait dans notre cadre, et (4.2) n'aura pas lieu. Ceci étant dit, il ne sera pas difficile d'adapter les arguments de la section 1.3.4 de manière à obtenir une estimation de gradient semblable à (4.3). Dans la suite on notera A^* la transposé d'une matrice A et, en conséquence, u^*v le produit scalaire de deux vecteurs.

Théorème 4.1. *Supposons que λ est constant et que $|\nabla Qf(x)|^2 \leq M(x)Q|\nabla f|^2(x)$ pour une fonction M telle que*

$$\forall (x, u) \in \Omega \times \mathbb{R}^d, \quad 2u^* J_b(x)u + \lambda (M(x) - 1) |u|^2 \leq -\eta |u|^2 \quad (4.4)$$

pour une constante $\eta \in \mathbb{R}$. Alors pour tout $t > 0$ et $f \in \mathcal{A}$,

$$|\nabla P_t f|^2(x) \leq e^{-\eta t} P_t |\nabla f|^2(x). \quad (4.5)$$

L'inégalité (4.4) peut s'interpréter comme une condition d'équilibre entre la dérive déterministe et les sauts. Plus précisément, supposons que $|\nabla Qf(x)|^2 \leq M(x)Q|\nabla f|^2(x)$ pour quelque fonction M sur Ω . Si $M < 1$, Q est une contraction de la distance de Wasserstein (ceci sera détaillé en section 4.2); ce qui signifie que deux particules sautant simultanément peuvent être couplées pour se rapprocher. Plus généralement M mesure à quel point deux telles particules peuvent être couplées pour ne pas trop s'éloigner l'une de l'autre. D'autre part J_b mesure comment deux trajectoires du flot déterministe ont tendance à se rapprocher ou à s'éloigner. En effet

$$\begin{aligned} \varphi_x(t) - \varphi_y(t) &= x - y + tJ_b(x)(x - y) + t \underset{y \rightarrow x}{o} (x - y) + \underset{t \rightarrow 0}{o} (t) \\ \Rightarrow |\varphi_x(t) - \varphi_y(t)|^2 &= |x - y|^2 + 2t(x - y)^* J_b(x)(x - y) + t \underset{y \rightarrow x}{o} (|x - y|^2) + |x - y| \underset{t \rightarrow 0}{o} (t), \end{aligned}$$

On voit que la condition $u^* J_b(x)u < 0$ pour tout $(x, u) \in \Omega \times \mathbb{R}^d$ implique que le flot contracte l'espace au voisinage de tout point de Ω .

Remarquons qu'en intégrant l'inégalité (4.5) par rapport à μ et en notant

$$W_t = \int |\nabla P_t f|^2 d\mu,$$

le théorème 4.1 implique que $W_t \leq e^{-\eta t} W_0$ pour tous $t > 0$, $f \in \mathcal{A}$, ou de façon équivalente que $\partial_t W_t \leq -\eta W_t$ pour tous $t > 0$, $f \in \mathcal{A}$, ou encore que $(\partial_t W_t)_{t=0} \leq -\eta W_0$ pour tout $f \in \mathcal{A}$.

Dans le cas d'un taux de saut non constant, sous une condition similaire à l'inégalité (4.4), on va pouvoir démontrer qu'il existe des constantes $\beta > 0$ et $\eta \in \mathbb{R}$ telles que

$$\partial_t W_t \leq -\eta W_t + 2\beta \mathcal{E}_t \quad (4.6)$$

où \mathcal{E}_t est l'énergie usuelle de $P_t f$ par rapport à μ ,

$$\mathcal{E}_t = \int \Gamma(P_t f) d\mu.$$

Rappelons que cette quantité apparaissait en section 1.3.1 lorsque l'on s'intéressait à la variance de $P_t f$ par rapport à μ ,

$$V_t = \int (P_t f)^2 d\mu - \left(\int P_t f d\mu \right)^2.$$

Dans ce contexte, il était dit que μ satisfait une inégalité de Poincaré (ou de trou spectral) par rapport à Γ lorsqu'il existe une constante $c > 0$ telle que $V_0 \leq c\mathcal{E}_0$ pour tout $f \in \mathcal{A}$.

Dans le cas des PDMP, le carré du champ est un opérateur non local (autrement dit ce n'est pas le carré d'un opérateur différentiel d'ordre 1), et on n'a en général pas d'inégalité de

Poincaré par rapport à Γ . En revanche il est tout à fait possible que la mesure invariante du processus satisfasse une inégalité de Poincaré de type diffusive, c'est-à-dire de la forme

$$\forall f \in \mathcal{A} \quad \int f^2 d\mu - \left(\int f d\mu \right)^2 \leq c \int |\nabla f|^2 d\mu, \quad (4.7)$$

en d'autres termes $V_t \leq cW_t$. Si l'on suppose une telle inégalité, d'après l'inégalité (4.6), si $\eta > 0$,

$$\begin{aligned} \partial_t (W_t + \beta V_t) &\leq -\eta W_t \\ &\leq -\frac{\eta}{1 + \beta c} (W_t + \beta V_t). \end{aligned}$$

Ceci nous amène au résultat suivant :

Théorème 4.2. *Supposons l'inégalité (4.7), et que $|\nabla Qf(x)|^2 \leq M(x)Q|\nabla f|^2(x)$ pour une fonction M telle que pour μ -presque tout $x \in \Omega$ et pour tout $u \in \mathbb{R}^d$,*

$$u^* \left(2J_b(x) + \frac{\nabla \lambda(x)(\nabla \lambda(x))^*}{\beta \lambda(x)} \right) u + \lambda(x) (M(x) - 1) |u|^2 \leq -\eta |u|^2$$

pour des constantes $\eta, \beta > 0$. Alors

$$W_t + \beta V_t \leq (W_0 + \beta V_0) e^{-\frac{\eta t}{\beta c + 1}}.$$

Remarquons que

$$W_t + \beta V_t = \|\nabla P_t f\|_{L^2(\mu)}^2 + \beta \|P_t f - \mu f\|_{L^2(\mu)}^2$$

est équivalente au carré de la norme \mathcal{H}^1 usuelle de $P_t f - \mu f$. Ainsi le théorème 4.2 fournit une décroissance dans $\mathcal{H}^1(\mu)$ plutôt que dans $L^2(\mu)$. En ce sens, nos arguments peuvent être vus comme une méthode de fonctionnelle de Lyapunov modifiée (présenté en section 1.4.1; voir également [86, 39, 11], etc.), encore qu'elle soit très simple dans notre cas. Dans ce contexte, l'hypothèse d'une inégalité de Poincaré de la forme (4.7) est relativement standard; elle était par exemple présente au chapitre 3. On a rappelé dans l'introduction de ce manuscrit des critères classiques sur une fonction F de \mathbb{R}^d pour montrer que $e^{-F(x)} dx$ satisfait une telle inégalité, et différents moyens pour estimer la constante c . Cependant pour les PDMP les mesures invariantes sont en général assez peu connues, avec des formules explicites de densités peu commodes à manipuler, si même on en dispose. Le reste de ce chapitre sera donc consacré à l'obtention des telles inégalités, qui sont intéressantes en elles-mêmes puisque elles fournissent des inégalités de concentration pour la mesure μ . Comme nous le verrons, notre méthode sera basée sur une idée initiale de Malrieu et Talay ([71]) pour les chaînes de Markov à espace continu, elle-même étant l'analogue en temps discret de la Proposition 1.15 pour les processus à temps continu.

Pour donner une idée de ce que permettront *in fine* de démontrer les idées que nous allons développer, énonçons une application dans un cas particulier, le TCP à taux linéaire :

Proposition 4.3. *Considérons $\delta \in [0, 1)$ et le générateur sur \mathbb{R}_+*

$$Lf(x) = f'(x) + x(f(\delta x) - f(x)).$$

Il existe des constantes c et r explicites telles que pour tout $f \in \mathcal{A}$,

$$\mu \left((P_t f)^2 \ln (P_t f)^2 \right) - \mu (P_t f)^2 \ln \mu \left((P_t f)^2 \right) \leq c e^{-rt} \mu |f'|^2.$$

Ce chapitre est organisé comme suit. Des versions légèrement généralisées des théorèmes 4.1 et 4.2 sont démontrés en section 4.2. Les inégalités de Beckner (qui incluent l'inégalité de Poincaré) pour les PDMP sont étudiées à l'aide des chaînes incluses des processus en section 4.3. La section 4.4 regroupe plusieurs exemples illustratifs, et la proposition 4.3 en particulier y est démontrée. Enfin un résultat de perturbation pour les inégalités de Poincaré et de log-Sobolev, notamment utilisé dans la preuve de la proposition 4.3, est énoncé et prouvé en annexe.

4.2 Exponential decay

We keep the notations and assumptions of the introduction. In particular we study the semi-group $(P_t)_{t \geq 0}$ with generator L defined by (4.1).

When A is a linear operator on \mathcal{A} and ϕ is a bilinear symmetric one, for $f, g \in \mathcal{A}$ we define

$$\Gamma_{A,\phi}(f, g) = \frac{1}{2} (A\phi(f, g) - \phi(f, Ag) - \phi(Af, g)).$$

With respect to f , $\Gamma_{A,\phi}(f, f)$ is quadratic, and linear with respect to A and ϕ . We will always note $f \mapsto \phi(f)$ the quadratic form associated to a bilinear form $f, g \mapsto \phi(f, g)$ and similarly we will always note $f, g \mapsto q(f, g)$ the symmetric bilinear form associated by polarization to a quadratic form $f \mapsto q(f)$ on \mathcal{A} . Let

$$\psi(s) = P_s \phi(P_{t-s} f), \quad s \in [0, t]$$

which interpolates between $\phi(P_t f)$ and $P_t(\phi f)$. Then

$$\psi'(s) = 2P_s \Gamma_{L,\phi}(P_{t-s} f).$$

To prove Theorems 4.1 and 4.2 we should consider $\phi(f) = |\nabla f|^2$. In fact it will be convenient for the applications to work with a weighted gradient $\phi_a(f) = a|\nabla f|^2$ with $a > 0$ a scalar field on Ω .

Lemma 4.4. 1. For all $f \in \mathcal{A}$

$$\Gamma_{b^* \nabla, \phi_a}(f) = \frac{b^* \nabla a}{2a} \phi_a(f) - a(\nabla f)^* J_b \nabla f.$$

2. Suppose there exists a function M on Ω such that, for all $f \in \mathcal{A}$, $\phi_a(Qf) \leq MQ(\phi_a(f))$, and let I be the identity operator on \mathcal{A} .

$$\Gamma_{\lambda(Q-I), \phi_a}(f) \geq -a(\nabla f)^*(\nabla \lambda)(Qf - f) + \frac{\lambda}{2}(1 - M)\phi_a(f).$$

Démonstration. First we note that

$$\nabla(b^* \nabla f) = J_b \nabla f + H_f b$$

with $H_f(x) = [\partial_i \partial_k f(x)]_{1 \leq i, k \leq d}$ the Hessian of f , and

$$b^* \nabla (a|\nabla f|^2) = (b^* \nabla a)|\nabla f|^2 + 2ab^* H_f \nabla f$$

Thus

$$\begin{aligned}\Gamma_{b,\nabla,\phi}(f) &= \frac{1}{2}b^*\nabla\left(a|\nabla f|^2\right) - a(\nabla f)^*\nabla(b^*\nabla f) \\ &= \frac{1}{2}(b^*\nabla a)|\nabla f|^2 - a^2(\nabla f)^*J_b\nabla f.\end{aligned}$$

As far as the second point is concerned,

$$\begin{aligned}\Gamma_{\lambda(Q-I),\phi}(f) &= \frac{1}{2}\lambda(Q(\phi_a(f)) - \phi_a(f)) - a(\nabla f)^*(\nabla\lambda)(Qf - f) - \lambda a(\nabla f)^*(\nabla Qf - \nabla f) \\ &\geq \frac{\lambda}{2}\left(Q(\phi_a(f)) + \phi_a(f) - 2\sqrt{\phi_a(f)\phi_a(Qf)}\right) - (Df)^*(D\lambda)(Qf - f).\end{aligned}$$

We conclude by

$$2\sqrt{\phi_a(f)\phi_a(Qf)} \leq 2\sqrt{M\phi_a(f)Q\phi_a(f)} \leq M\phi_a(f) + Q(\phi_a(f)).$$

□

We can now state the following :

Theorem 4.5. *Assume λ is constant and there exist a function M on Ω and a constant $\eta \in \mathbb{R}$ such that, for all $f \in \mathcal{A}$, $\phi_a(Qf) \leq MQ(\phi_a(f))$ and*

$$\forall(x, u) \in \Omega \times \mathbb{R}^d, \quad 2u^*J_b(x)u + \left(\lambda(M(x) - 1) - \frac{b^*\nabla a(x)}{a(x)} + \eta\right)|u|^2 \leq 0.$$

Then

$$\phi_a(P_t f) \leq e^{-\eta t} P_t(\phi_a(f)).$$

In particular with $a = 1$ we retrieve Theorem 4.1.

Démonstration. From Lemma 4.4, since in the constant case $\nabla\lambda = 0$,

$$\begin{aligned}\Gamma_{L,\phi_a}(f) &\geq -a(\nabla f)^*J_b\nabla f + a\left(\frac{b^*\nabla a}{2a} + \frac{\lambda}{2}(1 - M)\right)|\nabla f|^2 \\ &\geq \frac{\eta}{2}\phi_a(f).\end{aligned}$$

Hence if $\psi(s) = P_s\phi_a(P_{t-s}f)$,

$$\psi'(s) = 2P_s\Gamma_{L,\phi_a}(P_{t-s}f) \geq \eta\psi(s)$$

and $\psi(t) \geq e^{\eta t}\psi(0)$, which concludes. □

Remark that we did not use the ergodicity of the process here, and that η can be negative.

This commutation between the semigroup and the gradient leads to a contraction in Wasserstein distance. More precisely, define on Ω the distance associated to the weighted gradient $D = \sqrt{a}\nabla$ by

$$d(x, y) = \inf \left\{ \int_0^1 \frac{|\gamma'(s)|}{\sqrt{a(\gamma(s))}} ds, \gamma : [0, 1] \rightarrow \Omega, \text{ smooth}, \gamma(0) = x, \gamma(1) = y \right\}$$

and the associated Wasserstein distance between two probability laws ν_1, ν_2 having a finite p^{th} moment (*i.e.* for which there exists a $x_0 \in \Omega$ with $\nu_i [d(\cdot, x_0)] < \infty$) by

$$\mathcal{W}_{d,p}(\nu_1, \nu_2) = \inf_{X \sim \nu_1, Y \sim \nu_2} (\mathbb{E}[d^p(X, Y)])^{\frac{1}{p}}.$$

A function f will be called κ -Lipschitz with respect to D if $\forall x, y \in \Omega$,

$$f(x) - f(y) \leq \kappa d(x, y).$$

This is equivalent for a smooth function to $\|Df\|_{\infty} \leq \kappa$. The function $y \mapsto d(x, \cdot)$ is 1-Lipschitz with respect to D . Since the metric space (Ω, d) is locally diffeomorphic to $(\mathbb{R}^d, |\cdot|)$, thanks to Rademacher's theorem, this means a κ -Lipschitz function g is differentiable almost everywhere, with $\|Dg\|_{\infty} \leq \kappa$. We have the Kantorovich-Rubinstein dual representation (see [87])

$$\mathcal{W}_{d,1}(\nu_1, \nu_2) = \sup \{ \nu_1 f - \nu_2 f, \|Df\|_{\infty} \leq 1 \},$$

where we use the operator notation $\nu f = \int f d\nu$.

Recall that by duality a Markov semi-group acts on the right on probability laws by

$$(\nu_1 P_t) f := \nu_1 (P_t f).$$

If P_t were absolutely continuous with respect to the Lebesgue measure for $t > 0$ - which is not the case for a PDMP since for all time t there is a non-zero probability that the process hasn't jumped yet - the gradient estimate of Theorem 4.5 would yield, from [62, Theorem 2.2], a contraction of the $\mathcal{W}_{d,2}$ distance :

$$\mathcal{W}_{d,2}(\nu_1 P_t, \nu_2 P_t) \leq e^{-\frac{\eta}{2}t} \mathcal{W}_{d,2}(\nu_1, \nu_2).$$

Instead of trying to adapt Kuwada's result, since our work is more concerned about variance than Wasserstein distance, we will only state the weaker result :

Corollary 4.6. *In the settings of Theorem 4.5, for all laws ν_1, ν_2 with finite first moment, if $P_t \nu_1$ and $P_t \nu_2$ still have finite first moment,*

$$\mathcal{W}_{d,1}(\nu_1 P_t, \nu_2 P_t) \leq e^{-\frac{\eta}{2}t} \mathcal{W}_{d,1}(\nu_1, \nu_2).$$

Démonstration. Theorem 4.5 yields the weaker gradient estimate

$$\|DP_t f\|_{\infty} \leq e^{-\frac{\eta}{2}t} P_t \|Df\|_{\infty} = e^{-\frac{\eta}{2}t} \|Df\|_{\infty}.$$

This implies the $\mathcal{W}_{d,1}$ decay, thanks to the Kantorovich-Rubinstein dual representation □

We won't push the analysis further concerning the Wasserstein distance, but refer to the study in [10] of the TCP process where an exponential decay is first obtained for a distance equivalent to $d(x, y) = \sqrt{|x - y|}$ and then is transposed to $d(x, y) = |x - y|^p$ via moments estimates and Hölder inequality. For further considerations on gradient-semigroup commutation, one shall consult [18, 3, 62].

We now turn to the non-constant jump rate case. Let a be a non-negative scalar field on Ω . Throughout this chapter we will say a probability measure ν satisfies a weighted Poincaré inequality with constant c and weight a if for all $f \in \mathcal{A}$

$$\nu f^2 - (\nu f)^2 \leq c\nu \left(a|\nabla f|^2 \right). \quad (4.8)$$

Let $V_t = \mu(P_t f)^2 - (\mu f)^2$ and $W_t = \mu\phi_a(P_t f)$.

Theorem 4.7. *Assume that μ satisfies the weighted Poincaré inequality (4.8), that μ -almost everywhere $\lambda > 0$ and that there exist a function M and constants $\eta, \beta > 0$ such that for μ -almost all $x \in \Omega$, for all $f \in \mathcal{A}$ and for all $u \in \mathbb{R}^d$, $\phi_a(Qf) \leq MQ(\phi_a(f))$ and*

$$u^* \left(2J_b(x) + \frac{a}{\beta\lambda(x)} \nabla\lambda(x)(\nabla\lambda(x))^* \right) u + \left(\lambda(x)(M(x) - 1) - \frac{b^*\nabla a(x)}{a(x)} + \eta \right) |u|^2 \leq \quad (4.9)$$

Then

$$W_t + \beta V_t \leq e^{-\frac{\eta t}{\beta c + 1}} (W_0 + \beta V_0),$$

and

$$W_t \leq (1 + \beta c) e^{-\frac{\eta t}{\beta c + 1}} W_0.$$

Démonstration. Since μ is the invariant measure of the process, $\mu Lg = 0$ for all $g \in \mathcal{A}$. In particular if ϕ is a quadratic form on \mathcal{A} , $\mu(L\phi(f)) = 0$ and

$$\begin{aligned} \partial_t(\mu(\phi(P_t f))) &= 2\mu(\phi(P_t f), LP_t f) \\ &= -2\mu\Gamma_{L,\phi}(P_t f). \end{aligned}$$

In particular

$$\partial_t W_t = -2\mu\Gamma_{L,\phi_a}(P_t f).$$

From Lemma 4.4,

$$\begin{aligned} \Gamma_{\lambda(Q-I),\phi_a}(f) &\geq -a(\nabla f)^*(\nabla\lambda)(Qf - f) + \frac{\lambda}{2}(1 - M)\phi_a(f) \\ &\geq -\frac{a^2}{2\beta\lambda}|(\nabla f)^*\nabla\lambda|^2 - \frac{\beta\lambda}{2}(Qf - f)^2 + \frac{\lambda}{2}(1 - M)\phi_a(f). \end{aligned}$$

Again from Lemma 4.4 and from Inequality (4.9),

$$\Gamma_{L,\phi_a}(f) \geq \frac{\eta}{2}\phi_a(f) - \frac{\beta\lambda}{2}(Qf - f)^2.$$

On the other hand, if $\phi_2(f) = f^2$ then Γ_{L,ϕ_2} is the usual *carré du champ* operator. From the Leibniz rule, $\Gamma_{b,\nabla,\phi_2}f = 0$, so that

$$\begin{aligned} \partial_t V_t &= -2\mu\Gamma_{\lambda(Q-I),\phi_2}(P_t f) \\ &= -\mu\lambda \left(Q(P_t f)^2 + (P_t f)^2 - 2(P_t f)(QP_t f) \right) \\ &\leq -\mu\lambda (QP_t f - P_t f)^2 \end{aligned}$$

the last inequality being a consequence of the Cauchy-Schwartz inequality for Q . At the end of the day, we get

$$\partial_t (W_t + \beta V_t) \leq -\eta W_t$$

and, thanks to the weighted Poincaré inequality (4.8),

$$\partial_t (W_t + \beta V_t) \leq -\frac{\eta}{1 + \beta c} (W_t + \beta V_t),$$

which yields the first assertion. Then

$$W_t \leq W_t + \beta V_t \leq e^{-\frac{\eta t}{\beta c + 1}} (W_0 + \beta V_0) \leq (1 + \beta c) e^{-\frac{\eta t}{\beta c + 1}} W_0.$$

□

Note that η could depend on x , so that the weight that intervenes in the Poincaré inequality may be different from a . For instance for the TCP with linear rate on \mathbb{R}_+ (Example 4.4.4), one could consider $a(x) = x$ and $\eta(x) = -\kappa - \alpha x$ for some $\kappa, \alpha > 0$. Then it would be sufficient to prove an inequality with weight $\tilde{a} = 1 + x$, which is weaker than both the classical inequality with constant weight and the inequality with weight a .

4.3 Functional inequalities for PDMP

4.3.1 Confining operators

The variance is a way among others to quantify the distance to equilibrium. In this section we suppose that for all $f \in \mathcal{A}$ the so-called p -entropies

$$\begin{aligned} \text{Ent}_p f &= \frac{\mu f^2 - \left(\mu f^{\frac{2}{p}}\right)^p}{p-1} \quad \text{for } p \in (1, 2], \\ \text{Ent}_1 f &= \mu \left(f^2 \log f^2\right) - \left(\mu f^2\right) \log \left(\mu f^2\right) \end{aligned}$$

are well-defined. In this chapter we say that μ satisfies a Beckner's inequality $\mathcal{B}(p, c)$ if

$$\forall f \in \mathcal{A}, \quad \text{Ent}_p f \leq \frac{2c}{p} \mu |\nabla f|^2. \quad (4.10)$$

Note that this is a diffusion-like Beckner's inequality, involving an energy different from $\mu(\Gamma f)$. For $p = 2$ this is the Poincaré inequality, for $p = 1$ this is the Gross log Sobolev one, so that (4.10) is an interpolation between these two inequalities : indeed, $\text{Ent}_p f$ is non increasing with $p \in [1, 2]$ (see [63]; note that we took the definitions of [17]). In particular $\mathcal{B}(p, c)$ implies $\mathcal{B}\left(q, \frac{cq}{p}\right)$ whenever $q \geq p$. For the global study of this inequalities and of more general Φ -entropies, we refer to [26] and [17].

In this section, for the sake of simplicity, we won't consider weighted inequalities such as the weighted Poincaré inequality (4.8). Everything would work the same, and, at least in dimension one, a weighted inequality can be seen as a non-weighted one through a change of variable (see Section 4.4.4).

Remark that if μ satisfies $\mathcal{B}(p, c)$ for $p \in [1, 2]$, then it satisfies a Poincaré inequality. In this case, providing the inequality (4.9) of Theorem 4.7 holds, W_t decays exponentially fast, and

$$\text{Ent}_p P_t f \leq \frac{2c}{p} W_t \leq \frac{2c(1 + \beta c)}{p} e^{-\frac{\eta t}{\beta c + 1}} W_0.$$

Let $\psi : \Omega \rightarrow \Omega$ be a smooth function with Jacobian matrix J_ψ , and let $|J_\psi|$ be the euclidian operator norm of J_ψ , namely

$$|J_\psi| = \sup \left\{ |J_\psi u|, u \in \mathbb{R}^d, |u| = 1 \right\}.$$

We say ψ is γ -Lipschitz (where $\gamma \in \mathbb{R}_+$) if for all $x \in \Omega$, $|J_\psi(x)| \leq \gamma$. It is clear that in this case when the law of a random variable Z satisfies $\mathcal{B}(p, c)$ then the law of $\psi(Z)$ satisfies $\mathcal{B}(p, \gamma^2 c)$. In order to get Beckner's inequalities for the invariant law of a PDMP we will prove a generalization of this fact, based on an initial idea of Malrieu and Talay [71].

Let H be a Markov kernel on Ω .

Definition 4.8. *Let $c, \gamma > 0$, $p \in [1, 2]$. We say that H is (c, γ, p) -confining if both the following conditions are satisfied :*

- sub-commutation : $\forall f \in \mathcal{A}, \forall x \in \Omega$,

$$\left| \nabla \left(H f^{\frac{2}{p}} \right)^{\frac{p}{2}} \right|^2 (x) \leq \gamma H |\nabla f|^2 (x). \quad (4.11)$$

- Local Beckner's inequality : $\forall f \in \mathcal{A}, \forall x \in \Omega$,

$$\frac{H f^2 (x) - \left(H f^{\frac{2}{p}} \right)^p (x)}{p - 1} \leq \frac{2c}{p} H |\nabla f|^2 (x). \quad (4.12)$$

If $\gamma < 1$ we say H is (c, γ, p) -contractive. When there is no ambiguity for p , H will simply be called *confining* (or *contractive*) if there exist $c, \gamma > 0$ satisfying both conditions.

Examples :

- Let ψ be a γ -Lipschitz function and $H f(x) = f(\psi(x))$. The sub-commutation (4.11) is clear, and the local inequality (4.12) holds with $c = 0$, since $H(x)$ is a Dirac mass.
- The sub-commutation is always satisfied with $\gamma = 0$ if $H(x) = \nu$ is a constant kernel, namely is a probability on Ω , so that ν is confining iff it satisfies a Beckner's inequality.
- if N is a standard Gaussian vector on \mathbb{R}^d and $(B_t)_{t \geq 0}$ a Brownian motion on \mathbb{R}^d then

$$K_t f(x) = \mathbb{E} (f(x + B_t)) = \mathbb{E} (f(x + \sqrt{t}N))$$

is $(t, 1)$ -confined for the usual gradient and $p = 1$ (see [5]). If the Brownian motion is replaced by an elliptic diffusion, a sub-commutation is given by its Bakry-Emery curvature.

- Remark this definition could be extended to a Markov kernel $H : \Omega_1 \rightarrow \mathcal{P}(\Omega_2)$ with $\Omega_1 \subset \mathbb{R}^d$ and $\Omega_2 \subset \mathbb{R}^n$. For instance if φ is the flow associated to a vector fields b on Ω_1 then $H f(t) = f(\varphi_x(t))$ is a Markov kernel from \mathbb{R}_+ to $\mathcal{P}(\Omega_1)$, and $\partial_t H f = H(b^* \nabla f)$.

Here is maybe the most important, although very simple, result of this chapter :

Lemma 4.9. For $i = 1, 2$, let H_i be a (c_i, γ_i, p) -confining Markov kernel on Ω .

1. Then $H_1 H_2$ is a $(c_2 + \gamma_2 c_1, \gamma_1 \gamma_2, p)$ -confining Markov kernel.
2. If $\nu \in \mathcal{P}(\Omega)$ satisfies $\mathcal{B}(p, c)$ then νH_2 satisfies $\mathcal{B}(p, c_2 + \gamma_2 c)$.
3. If H is (c, γ, p) -contractive and if the Markov chain generated by H is ergodic, meaning that H^n converges weakly to some $\nu \in \mathcal{P}(\Omega)$ when n goes to infinity, then the invariant law ν satisfies $\mathcal{B}(p, c(1 - \gamma)^{-1})$.

Démonstration. Let $p \in (1, 2]$ (the case $p = 1$ is similar and already treated in [27]). First,

$$\left| \nabla \left(H_1 H_2 f^{\frac{2}{p}} \right)^{\frac{p}{2}} \right|^2 \leq \gamma_1 H_1 \left(\left| \nabla \left(H_2 f^{\frac{2}{p}} \right)^{\frac{p}{2}} \right|^2 \right) \leq \gamma_1 \gamma_2 H_1 H_2 |\nabla f|^2$$

and

$$\begin{aligned} \frac{H_1 H_2 f^2 - \left(H_1 H_2 f^{\frac{2}{p}} \right)^p}{p-1} &= \frac{1}{p-1} \left(H_1 \left[H_2 f^2 - \left(H_2 f^{\frac{2}{p}} \right)^p \right] + H_1 \left(H_2 f^{\frac{2}{p}} \right)^p - \left(H_1 H_2 f^{\frac{2}{p}} \right)^p \right) \\ &\leq \frac{2c_2}{p} H_1 H_2 |\nabla f|^2 + \frac{1}{p-1} \left(H_1 g^2 - \left(H_1 g^{\frac{2}{p}} \right)^p \right) \quad \text{with } g = \left(H_2 f^{\frac{2}{p}} \right)^{\frac{p}{2}} \\ &\leq \frac{2c_2}{p} H_1 H_2 |\nabla f|^2 + \frac{2c_1}{p} H_1 |\nabla g|^2 \\ &\leq \frac{2(c_2 + \gamma_2 c_1)}{p} H_1 H_2 |\nabla f|^2. \end{aligned}$$

The second point is obtained from the first one by considering $H_1 = \nu$. Concerning the third assertion, by induction from the first one we get for all $n \in \mathbb{N}$

$$\frac{H^n f^2 - \left(H^n f^{\frac{2}{p}} \right)^p}{p-1} \leq \frac{2c}{p} \left(\sum_{k=0}^n \gamma^k \right) H^n |\nabla f|^2.$$

The weak convergence of H^n to ν concludes. □

Example : Let $(E_k)_{k \geq 0}$ be an i.i.d. sequence of standard exponential variables, and $(X_k)_{k \geq 0}$ be the Markov chain on \mathbb{R}_+ defined by $X_{k+1} = \frac{X_k + E_k}{2}$. Its transition operator is

$$Pf(x) = \mathbb{E} \left(f \left(\frac{x + E}{2} \right) \right).$$

Clearly $(Pf)'(x) = \frac{1}{2} P(f')(x)$, so that $|(Pf)'|^2 \leq \frac{1}{4} P|f'|^2$. On the other hand $P(x)$, the law of $\frac{x+E}{2}$, is the image by a $\frac{1}{2}$ -Lipschitz transformation of the exponential law $\mathcal{E}(1)$, which satisfies a Poincaré inequality $\mathcal{B}(2, 4)$ (cf. Theorem 1.12). Thus P is $(2, \delta^2, 2)$ -contractive. On the other hand it is clear the chain is irreducible, it admits $C = [0, 3]$ as a small set and $V(x) = x + 1$ as a Lyapunov function (since $PV(x) \leq \frac{3}{4}V(x) + 1_{x < 3}$) so that it is ergodic (see [41] for definitions and proof). According to Lemma 4.9, the invariant measure satisfies a Poincaré inequality $\mathcal{B} \left(2, \frac{2}{1-\delta^2} \right)$.

4.3.2 The embedded chain

Recall $X = (X_t)_{t \geq 0}$ is a process on Ω with generator given by (4.1). Let $(S_k)_{k \geq 0}$ be the jump times of X and let $Z_k = X_{S_k}$. The Markov chain $(Z_k)_{k \geq 0}$ is called the embedded chain associated to X .

For $s \in [S_k, S_{k+1})$, $X_s = \varphi_{Z_k}(s - S_k)$ where we recall $\varphi_x(t)$ is the flow associated to the vector field b . Since

$$\frac{d}{dt} (f(\varphi_x(t))) = (b^* \nabla f)(\varphi_x(t)),$$

we shall say that a function f is non-decreasing (resp. constant, concave, etc.) along the flow if $t \mapsto f(\varphi_x(t))$ is non-decreasing (resp. constant, etc.) for all $x \in \Omega$; in other word if $b^* \nabla f \geq 0$ (resp. $= 0$, etc.).

Conditionally to the event $Z_k = x$, the inter-jump time $T_k = S_{k+1} - S_k$ has a density

$$p_x(t) = \lambda(\varphi_x(t)) e^{-\int_0^t \lambda(\varphi_x(s)) ds}$$

on \mathbb{R}^+ . We assume the inter-jump times are *a.s.* finite (which is clear if $\liminf_{t \rightarrow \infty} \lambda(\varphi_x(t)) > 0$ for all x), and define

$$Kf(x) = \int_0^{+\infty} f(\varphi_x(t)) p_x(t) dt = \mathbb{E}(f(\varphi_x(T_k)) | Z_k = x).$$

Then $P = KQ$ is the transition operator for the chain Z .

Transferring properties from X to Z , or the converse, is far from obvious. In fact it is quite easy to find counter-examples for which one is ergodic and not the other (see examples 34.28 and 34.33 of [31]). In [30] this problem is solved with the definition of another embedded chain by adding observation points at constant rate. That being said, in the following we won't delve into this issue, and simply assume Z has a unique invariant law μ_e (which can often be proved under conditions of irreducibility, aperiodicity and existence of a Lyapunov function). In this case we can express μ from μ_e :

Lemma 4.10 (from [31]). *Assume $C = \mu_e K \left(\frac{1}{\lambda} \right) = \mu_e \left[\int_0^\infty e^{-\int_0^t \lambda(\varphi_x(s)) ds} dt \right] < \infty$. Then*

$$\mu f = C^{-1} \mu_e K \left(\frac{f}{\lambda} \right).$$

In other words, $\mu = \nu_e \widetilde{K}$ where

$$\begin{aligned} \widetilde{K}f(x) &= \frac{1}{K\left(\frac{1}{\lambda}\right)(x)} K \left(\frac{f}{\lambda} \right) (x) \\ \nu_e f &= \frac{1}{C} \mu_e \left[f K \left(\frac{1}{\lambda} \right) \right]. \end{aligned}$$

Démonstration. As we assumed the unicity of μ , it is characterized by $\mu L = 0$, and

$$\begin{aligned} \mu_e \left[\int_0^\infty Lf(\phi_x(t)) e^{-\int_0^t \lambda(\phi_x(s)) ds} dt \right] &= \mu_e \left[\int_0^\infty \frac{d}{dt} \left(f(\phi_x(t)) e^{-\int_0^t \lambda(\phi_x(s)) ds} \right) dt \right] \\ &\quad + \mu_e \left[\int_0^\infty \lambda(\phi_x(t)) Qf(\lambda(\phi_x(t))) e^{-\int_0^t \lambda(\phi_x(s)) ds} dt \right] \\ &= -\mu_e f(x) + \mu_e P f(x) \\ &= 0. \end{aligned}$$

□

In the following we will always assume the condition $C < \infty$ holds, so that ν_e and \widetilde{K} are well defined.

Here is our plan : from Lemma 4.9, we may establish a Beckner's inequality for μ_e by proving the operator P is contractive. By perturbative results on functional inequalities (see Theorem 1.13 and 1.14, [26], or Appendix in this chapter) this may give an inequality for ν_e . Finally, again from Lemma 4.9, we may transfer the inequality from ν_e to μ by proving the operator \widetilde{K} is confining.

The rest of this section will thus enlight some general facts which will later help us (mostly in dimension 1) prove K and \widetilde{K} are confining. It is strongly inspired by the work of Chafaï, Malrieu and Paroux [27], in which a log-Sobolev inequality is proved for the invariant measure of the embedded chain of a particular PDMP, the TCP with linear rate (see example 4.4.4).

Until the end of this section we suppose $\lambda > 0$ almost everywhere. Then

$$\Lambda_x(t) = \int_0^t \lambda(\varphi_x(u)) du$$

is invertible. Remark that

$$\Lambda_{\varphi_x(s)}(t) = \Lambda_x(t+s) - \Lambda_x(s) \quad (4.13)$$

which yields both

$$b(x)^* \nabla_x (\Lambda_x(t)) = \frac{d}{ds} \Big|_{s=0} (\Lambda_{\varphi_x(s)}(t)) = \lambda(\varphi_x(t)) - \lambda(x) \quad (4.14)$$

and, taking $u = \Lambda_{\varphi_x(s)}(t)$ in $t+s = \Lambda_x^{-1}(\Lambda_{\varphi_x(s)}(t) + \Lambda_x(s))$,

$$\Lambda_{\varphi_x(s)}^{-1}(u) = \Lambda_x^{-1}(u + \Lambda_x(s)) - s. \quad (4.15)$$

If $X_0 = x$ and if T_x is the next time of jump then

$$E = \int_0^{T_x} \lambda(\phi_x(u)) du$$

is independant from X_0 , and has a standard exponential law. In other words $T_x \stackrel{dist}{=} \Lambda_x^{-1}(E)$, and $T_{\varphi_x(t)} \stackrel{dist}{=} \Lambda_{\varphi_x(t)}^{-1}(\Lambda_x(T_x))$.

Lemma 4.11. *If λ is non-decreasing along the flow, then for all $x \in \Omega$ and $t > 0$, the law of $T_{\varphi_x(t)}$ is the image of the law of T_x by a 1-Lipschitz function.*

Démonstration. Let $x \in \Omega$ and $t > 0$. For $s > 0$ we note $G(s) = \Lambda_{\varphi_x(t)}^{-1}(\Lambda_x(s))$, so that $T_{\varphi_x(t)} \stackrel{dist}{=} G(T_x)$. From $\frac{d}{du}(\Lambda_x(u)) = \lambda(\varphi_x(u))$, we get

$$\begin{aligned} G'(s) &= \frac{\lambda(\varphi_x(s))}{\lambda(\varphi_{\varphi_x(t)}(\Lambda_{\varphi_x(t)}^{-1}(\Lambda_x(s)))} \\ &= \frac{\lambda(\varphi_x(s))}{\lambda(\varphi_x(t + G(s)))} \end{aligned}$$

From the relation (4.15) and the fact that Λ_x (hence Λ_x^{-1}) is non-decreasing,

$$t + G(s) = \Lambda_x^{-1}(\Lambda_x(s) + \Lambda_x(t)) \geq \Lambda_x^{-1}(\Lambda_x(s)) = s.$$

Thus $\lambda(\varphi_x(s)) \leq \lambda(\varphi_x(t + G(s)))$ and $|G'(s)| \leq 1$. □

The assumption that the jump rate is non-decreasing along the flow is natural in several applications where the role of the jump mechanism is to counteract a deterministic trend (growth/fragmentation models for cells [19], TCP dynamics [27], etc.). In this context, the more the system is driven away by the flow, the more it is likely to jump. From a mathematical point of view, thanks to Lemma 4.11, a Beckner's inequality for the law $K(x)$ may be transferred to $K(\varphi_x(t))$ for all $t > 0$.

In fact this is also true for \widetilde{K} . Let \widetilde{T}_x be a random variable on \mathbb{R}_+ with density $\frac{e^{-\Lambda_x(t)}}{\int_0^\infty e^{-\Lambda_x(w)} dw}$, so that

$$\widetilde{K}f(x) = \mathbb{E} \left[f \left(\varphi_x \left(\widetilde{T}_x \right) \right) \right].$$

Lemma 4.12. *If λ is non-decreasing along the flow, then for all $x \in \Omega$ and $t > 0$, the law of $\widetilde{T}_{\varphi_x(t)}$ is the image of the law of \widetilde{T}_x by a 1-Lipschitz function.*

Démonstration. We will prove Lemma 4.11 applies here. Indeed the law of $\widetilde{T}_{\varphi_x(t)}$ is the law of $\widetilde{T}_x - t$ conditionnaly to the event $\widetilde{T}_x > t$, exactly as the law of $T_{\varphi_x(t)}$ is the law of $T_x - t$ conditionnaly to the event $T_x > t$. We need to find a jump rate which define \widetilde{T}_x as the jump time of a Markov process.

Let $e^{-V(s)} ds$ be a positive density on \mathbb{R}_+ , assume V is convex and let

$$r(t) = \frac{e^{-V(t)}}{\int_t^\infty e^{-V(s)} ds}.$$

Note that $r(t) = \frac{d}{dt} \left(-\ln \int_t^\infty e^{-V(s)} ds \right)$, so that

$$e^{-\int_0^t r(s) ds} = \int_t^\infty e^{-V(s)} ds.$$

Differentiating this equality yields

$$r(t) e^{-\int_0^t r(s) ds} = e^{-V(t)}.$$

We want to prove r is non-decreasing. From the convexity of V ,

$$r(t) = \frac{e^{-V(t)}}{\int_t^\infty e^{-V(s)} ds} = \frac{\int_t^\infty V'(s) e^{-V(s)} ds}{\int_t^\infty e^{-V(s)} ds} \geq V'(t).$$

As a consequence,

$$r'(t) = r(t) (r(t) - V'(t)) \geq 0.$$

In the case of \widetilde{T}_x , if λ is non-decreasing along the flow then $V(t) = \Lambda_x(t) - \ln \int_0^\infty e^{-\Lambda_x(w)} dw$ is convex. \square

Lemma 4.13. *For all $f \in \mathcal{A}$, $x \in \Omega$,*

$$b(x)^* \nabla (Kf)(x) = \lambda(x) K \left(\frac{b^* \nabla f}{\lambda} \right).$$

In particular if λ is non-decreasing along the flow, $|b^ \nabla (Kf)| \leq K |b^* \nabla f|$.*

Démonstration. From the representation

$$Kf(x) = \mathbb{E} (f(\varphi_x(T_x))) = \mathbb{E} (f(\varphi_x(\Lambda_x^{-1}(E)))) ,$$

we compute

$$\begin{aligned} b(x)^* \nabla (Kf)(x) &= \left. \frac{d}{ds} \right|_{s=0} (Kf(\varphi_x(s))) \\ &= \mathbb{E} \left(\left. \frac{d}{ds} \right|_{s=0} f[\varphi_{\varphi_x(s)}(\Lambda_{\varphi_x(s)}^{-1}(E))] \right) \\ &= \mathbb{E} \left(\left. \frac{d}{ds} \right|_{s=0} f[\varphi_x(s + \Lambda_{\varphi_x(s)}^{-1}(E))] \right) \\ &= \mathbb{E} \left(\left. \frac{d}{ds} \right|_{s=0} f[\varphi_x(\Lambda_x^{-1}(E + \Lambda_x(s)))] \right) \quad (\text{from Relation (4.15)}) \\ &= \mathbb{E} \left(\Lambda'_x(0) (\Lambda_x^{-1})'(E) (b^* \nabla f)[\varphi_x(\Lambda_x^{-1}(E))] \right) \\ &= \mathbb{E} \left(\frac{\lambda(x)}{\lambda(\varphi_x(\Lambda_x^{-1}(E)))} (b^* \nabla f)[\varphi_x(\Lambda_x^{-1}(E))] \right). \end{aligned}$$

If λ is non-decreasing along the flow, $\lambda(\varphi_x(t)) \geq \lambda(x)$ for all $t \geq 0$. □

Lemma 4.14. Let $h(x) = \int_0^\infty e^{-\Lambda_x(u)} du$. Then for all $f \in \mathcal{A}$, $x \in \Omega$,

$$b(x)^* \nabla (\widetilde{K}f)(x) = \frac{\widetilde{K}(hb^* \nabla f)(x)}{h(x)}.$$

In particular if λ is non-decreasing along the flow $|b^* \nabla \widetilde{K}f|(x) \leq \widetilde{K}|b^* \nabla f|(x)$.

Démonstration.

$$\widetilde{K}f(x) = \mathbb{E} [f(\varphi_x(\widetilde{T}_x))].$$

Note that $F_x(t) = \int_0^t \frac{e^{-\Lambda_x(s)}}{\int_0^\infty e^{-\Lambda_x(w)} dw} ds$ the cumulative function of \widetilde{T}_x is invertible. Let U be uniform on $[0, 1]$. Then

$$\begin{aligned} \widetilde{K}f(x) &= \mathbb{E} [f(\varphi_x(F_x^{-1}(U)))] \\ \Rightarrow b(x)^* \nabla \widetilde{K}f(x) &= \mathbb{E} \left[\left. \frac{d}{ds} \right|_{s=0} f(\varphi_x(s + F_{\varphi_x(s)}^{-1}(U))) \right] \\ &= \mathbb{E} \left[(1 + b(x)^* \nabla_x (F_x^{-1}(U))) (b^* \nabla f)(\varphi_x(F_x^{-1}(U))) \right], \end{aligned}$$

If $u \in [0, 1]$, from $\nabla_x (F_x(F_x^{-1}(u))) = \nabla_x(u) = 0$ we get

$$b(x)^* \nabla_x (F_x^{-1}(u)) = \frac{-b(x)^* \nabla_x (F_x)(F_x^{-1}(u))}{F'_x(F_x^{-1}(u))}. \quad (4.16)$$

On the first hand $F'_x(t) = \frac{e^{-\Lambda_x(t)}}{\int_0^\infty e^{-\Lambda_x(w)} dw}$. On the other hand from the equality (4.14) we compute

$$\begin{aligned} b(x)^* \nabla_x (F_x)(t) &= \frac{\int_0^t (\lambda(x) - \lambda(\phi_x(s))) e^{-\Lambda_x(s)} ds}{\int_0^\infty e^{-\Lambda_x(w)} dw} + F_x(t) \frac{\int_0^\infty (\lambda(\phi_x(w)) - \lambda(x)) e^{-\Lambda_x(w)} dw}{\int_0^\infty e^{-\Lambda_x(w)} dw} \\ &= \lambda(x) F_x(t) + \frac{[e^{-\Lambda_x(s)}]_0^t}{\int_0^\infty e^{-\Lambda_x(w)} dw} - \lambda(x) F_x(t) - F_x(t) \frac{[e^{-\Lambda_x(s)}]_0^\infty}{\int_0^\infty e^{-\Lambda_x(w)} dw} \\ &= \frac{-1 + e^{-\Lambda_x(t)} + F_x(t)}{\int_0^\infty e^{-\Lambda_x(w)} dw}. \end{aligned}$$

Relation (4.16) yields

$$\begin{aligned} 1 + b(x)^* \nabla_x (F_x^{-1}(u)) &= 1 - \frac{-1 + e^{-\Lambda_x(F_x^{-1}(u))} + F_x(F_x^{-1}(u))}{e^{-\Lambda_x(F_x^{-1}(u))}} \\ &= e^{\Lambda_x(t)} (1 - F_x(t)). \end{aligned}$$

with $t = F_x^{-1}(u)$. Thanks to Equation (4.13),

$$\begin{aligned} e^{\Lambda_x(t)} (1 - F_x(t)) &= e^{\Lambda_x(t)} \int_t^\infty \frac{e^{-\Lambda_x(s)}}{\int_0^\infty e^{-\Lambda_x(u)} du} ds \\ &= \frac{\int_0^\infty e^{-\Lambda_x(w+t) + \Lambda_x(t)} dw}{\int_0^\infty e^{-\Lambda_x(u)} du} \\ &= \frac{\int_0^\infty e^{-\Lambda_{\varphi_x(t)}(w)} dw}{\int_0^\infty e^{-\Lambda_x(u)} du}. \end{aligned}$$

Bringing the pieces together, we have proved

$$\begin{aligned} b(x)^* \nabla (\widetilde{K}f)(x) &= \mathbb{E} \left[\frac{h(\varphi_x(F_x^{-1}(U)))}{h(x)} (b^* \nabla f)(\varphi_x(F_x^{-1}(U))) \right] \\ &= \frac{\widetilde{K}(hb^* \nabla f)(x)}{h(x)} \end{aligned}$$

When λ is non-decreasing along the flow, from (4.14), $x \mapsto \Lambda_x(t)$ is non-decreasing along the flow for all $t \geq 0$, and $h(\varphi_x(t)) \leq h(x)$. \square

4.4 Examples

4.4.1 The TCP with constant rate

A simple yet instructive example on \mathbb{R}_+ is the TCP with constant rate of jump with generator

$$Lf(x) = f'(x) + \lambda (\mathbb{E}(f(Rx)) - f(x))$$

where R is a random variable on $[0, 1)$ and $\lambda > 0$ is constant. In [69], ergodicity is proved and it is shown the moments of the invariant measure μ are all finite; so \mathcal{A} may be chosen as the set of smooth functions for which all derivatives grow at most polynomially at infinity.

Applying Theorem 4.5 with $J_b = 0$, $M = \mathbb{E}(R^2)$ and $a = 1$, we get

Proposition 4.15. *for all $f \in \mathcal{A}$,*

$$|(P_t f)'|^2 \leq e^{-\lambda(1-\mathbb{E}(R^2))t} P_t |f'|^2.$$

Corrolary 4.6 then yields a contraction at rate $\lambda(1 - \mathbb{E}(R^2))$ of the Wasserstein distance $\mathcal{W}_1(\nu_1 P_t, \nu_2 P_t)$. In fact by coupling two processes starting at different points to have the same jump times and the same factor R at each jump, one get that for any $p \geq 1$, the \mathcal{W}_p distance decays at rate $\lambda p^{-1}(1 - \mathbb{E}(R^p))$ (see [27]), and those rates are optimal (see [70]). In particular $\lambda(1 - \mathbb{E}(R^2))$ is the rate of decay of \mathcal{W}_2^2 .

The moments of μ are such that μ has an exponential tail. Hence we can't expect a Beckner's inequality $\mathcal{B}(p, c)$ to hold for $p < 2$. Let

$$Kf = \int_0^\infty f(x+s)\lambda e^{-\lambda s} ds.$$

Obviously $(Kf)' = K(f')$. Moreover the exponential law $\mathcal{E}(1)$ satisfies a Poincaré inequality $\mathcal{B}(2, 4)$, so that by the change of variable $z \mapsto z/\lambda$, $\mathcal{E}(\lambda)$ satisfies $\mathcal{B}(2, 4\lambda^{-2})$. Finally, $K(x)$ is the image of $\mathcal{E}(\lambda)$ by the translation $u \mapsto u + x$, which is a 1-Lipschitz transformation. As a conclusion,

Lemma 4.16. *The operator K is $(4\lambda^{-2}, 1, 2)$ -confining.*

As far as the jump operator $Qf(x) = \mathbb{E}(f(Rx))$ is concerned, we have already used the sub-commutation

$$((Qf)')^2 \leq \mathbb{E}(R^2) Q(f')^2.$$

However a local Poincaré inequality (4.12) for $Q(x)$ would mean $\forall f \in \mathcal{A}, x > 0$,

$$\begin{aligned} \mathbb{E}(f^2(Rx)) - (\mathbb{E}(f(Rx)))^2 &\leq c\mathbb{E}\left[\left((f(Rx))'\right)^2\right] \\ \Leftrightarrow \mathbb{E}(g_x^2(R)) - (\mathbb{E}(g_x(R)))^2 &\leq \frac{c}{x^2}\mathbb{E}\left[(g_x'(R))^2\right] \end{aligned}$$

with $g_x(r) = f(rx)$. This implies the law of R satisfies $\mathcal{B}(2, cx^{-2})$ for all $x > 0$, hence $\mathcal{B}(2, 0)$, which means R is deterministic. Indeed, when R is deterministic, the local inequality always holds :

Lemma 4.17. *If $R = \delta$ a.s. with a constant $\delta \in [0, 1)$ then Q is $(0, \delta^2, p)$ -contractive.*

When R is random, what prevent to straightforwardly use our argument is the possibility of arbitrarily little concentrated jump, for instance with uniform law on $(0, x)$ for any x . It's a shame because if, say, R is uniform on $(0, \frac{1}{2})$, it means when the process jumps it is at least divided by 2 but can be even much more contracted. In particular its invariant measure should be more concentrated near zero than the process with $R = \frac{1}{2}$ a.s. which, as we will see, satisfies a Poincaré inequality. This illustrates a limit of our procedure.

Proposition 4.18. *If $R = \delta$ is deterministic then μ satisfies the Poincaré inequality*

$$\forall f \in \mathcal{A}, \quad \mu f^2 - (\mu f)^2 \leq \frac{4}{\lambda^2(1-\delta^2)} \mu(f')^2.$$

As a consequence,

$$\forall f \in \mathcal{A}, \quad \mu(P_t f - \mu f)^2 \leq \frac{4e^{-\lambda(1-\delta^2)t}}{\lambda(1-\delta^2)} \mu(f')^2.$$

Démonstration. Since K and Q are confining, from Lemma 4.9, $P = KQ$ is $(4\lambda^{-2}\delta^2, \delta^2, 2)$ -confining and μ_e the invariant measure of the embedded chain satisfies $\mathcal{B}\left(2, \frac{4\delta^2}{\lambda^2(1-\delta^2)}\right)$. From Lemma 4.10, $\mu = \mu_e K$ and so by Lemma 4.9 again μ satisfies $\mathcal{B}\left(2, \frac{4}{\lambda^2(1-\delta^2)}\right)$. The second inequality is a consequence of this Poincaré inequality and of Proposition 4.15. \square

In fact in this example the spectrum of the generator in $L^2(\mu)$ is explicit : there are polynomial eigenfunctions, and since the tail of μ is exponential, polynomials are dense in $L^2(\mu)$ and these eigenfunctions are the only one in $L^2(\mu)$. The eigenvalues are $l_k = \lambda(\mathbb{E}[R^k] - 1)$ with $k \in \mathbb{Z}^+$. The convergence rate of the L^2 -norm obtained in Proposition 4.18 for a deterministic R appears to be $\frac{1}{2}|l_2|$ and not the spectral gap $|l_1|$, and of course

$$\frac{1}{2}|l_2| = \lambda \mathbb{E} \left[(1-R) \frac{1+R}{2} \right] \leq \lambda \mathbb{E}(1-R) = |l_1|.$$

Nevertheless $\frac{1}{2}|l_1| \leq |l_2|/2$ so we get the right rate up to a factor 1/2.

Note that, contrary to the Ornstein-Uhlenbeck case (see Proposition 1.15), a pointwise Poincaré inequality of the form

$$\forall x > 0, f \in \mathcal{A}, \quad P_t f^2(x) - (P_t f)^2(x) \leq c(t) P_t (f')^2(x). \quad (4.17)$$

cannot hold. Indeed $P_t(x)$ is the mix of a Dirac mass at $x+t$ and of a smooth density with support included in $[0, \delta x+t]$. If one consider $f \in \mathcal{A}$ which is constant equal to 0 on $[0, \delta x+t]$ and constant equal to 1 in the neighborhood of $x+t$ then the left part of (4.17) is non-zero while the right part vanishes. More precisely if there have been n jumps during the time t , then $X_t \in [\delta^n(x+t), \delta^n x+t] = I_n$, so that the support of $P_t(x)$ is included in $\bigcup_{n \geq 0} I_n$. To control the variance of $P_t f$, one need to control the variations of f inside each interval I_n , which is done with $P_t (f')^2$, but also the variations between two different intervals, which may be done with $P_t (Qf - f)^2$.

Proposition 4.19. *If $R = \delta$ is deterministic then P_t satisfies the local inequality*

$$\forall x > 0, f \in \mathcal{A}, \quad P_t f^2(x) - (P_t f)^2(x) \leq \lambda(t+t^2) P_t (Qf - f)^2(x) + \lambda t^2 (1+t)^2 P_t (f')^2(\delta x).$$

Démonstration. Let $\phi(f)(x) = (f(\delta x) - f(x))^2$. We compute

$$\begin{aligned} \Gamma_{\partial_x, \phi} f(x) &= (f(\delta x) - f(x)) (\delta f'(\delta x) - f'(x)) - (f(\delta x) - f(x)) (f'(\delta x) - f'(x)) \\ &= -(1-\delta) (f(\delta x) - f(x)) f'(\delta x) \end{aligned}$$

and

$$\begin{aligned} \Gamma_{\lambda(Q-I), \phi} f(x) &= \frac{\lambda}{2} \left((f(\delta x) - f(x))^2 + (f(\delta^2 x) - f(\delta x))^2 - 2(f(\delta x) - f(x)) (f(\delta^2 x) - f(\delta x)) \right) \\ &= \frac{\lambda}{2} (f(\delta^2 x) + f(x) - 2f(\delta x))^2. \end{aligned}$$

Hence for all $\beta > 0$,

$$\begin{aligned} \Gamma_{L, \phi} f(x) &\geq -(1-\delta) (f(\delta x) - f(x)) f'(\delta x) \\ &\geq -\frac{1-\delta}{2} \left(\frac{1}{\beta} \phi(f)(x) + \beta (f'(\delta x))^2 \right). \end{aligned}$$

With $\psi(s) = P_s\phi(P_{t-s}f)(x)$ and $\beta = 1 + s$ we get

$$\begin{aligned}\psi'(s) &\geq -\left(\frac{1}{1+s}\psi(s) + (1+s)P_s((P_{t-s}f)'(\delta x))^2\right) \\ \Rightarrow \frac{d}{ds}\left((1+s)\psi(s)\right) &\geq -(1+s)^2P_s((P_{t-s}f)'(\delta x))^2.\end{aligned}$$

From Proposition 4.15, this yields

$$\begin{aligned}\psi(0) &\leq (1+t)\psi(t) + \left(\int_0^t (1+s)^2 e^{-\lambda(1-\delta^2)(t-s)} ds\right) P_t(f')^2(\delta x) \\ \Rightarrow \phi(P_t f)(x) &\leq (1+t)P_t\phi(f)(x) + t(1+t)^2 P_t(f')^2(\delta x).\end{aligned}$$

With $\phi_2(f) = f^2$ we now consider the usual *carré-du-champs* operator

$$\begin{aligned}\Gamma_{L,\phi_2}f(x) &= \frac{\lambda}{2}\left(f^2(\delta x) - f^2(x)\right) - \lambda f(x)(f(\delta x) - f(x)) \\ &= \frac{\lambda}{2}\phi(f)(x),\end{aligned}$$

so that

$$\begin{aligned}P_t f^2(x) - (P_t f)^2(x) &= \lambda \int_0^t P_s \phi(P_{t-s}f) ds \\ &\leq \lambda(t+t^2)P_t\phi(f)(x) + \lambda t^2(1+t)^2 P_t(f')^2(\delta x).\end{aligned}$$

□

It would be natural to expect an inequality of the form

$$P_t f^2(x) - (P_t f)^2(x) \leq c(t)P_t\left((Qf - f)^2 + (f')^2\right)$$

with a bounded $t \mapsto c(t)$. We could prove such an inequality in the same way as Lemma 4.9 if we had a function γ such that

$$(QP_t f - P_t f)^2 \leq \gamma(t)P_t(Qf - f)^2 \quad (4.18)$$

and a time t for which $\gamma(t) < 1$. Note that in the present case the Bakry-Emery criterion is not satisfied, which implies that an inequality of the form (4.18) cannot hold with $\gamma'(0) < 0$ or, equivalently, with $\gamma(t) = e^{-rt}$ for some $r > 0$.

4.4.2 The storage model

Consider the generator on \mathbb{R}_+

$$Lf(x) = -xf'(x) + \lambda(x)(\mathbb{E}[f(x+U)] - f(x)).$$

This is, in a sense, the converse of the TCP : the jumps send the process away from 0 and the flow brings it back. Applying Theorem 4.5 with $M = 1$, $a = 1$ and $J_b = -1$, we get

$$|\nabla P_t f|^2 \leq e^{-2t}P_t|\nabla f|^2. \quad (4.19)$$

Besides in this case it is easy to obtain Wasserstein decay, as the distance s between two processes starting at different point and coupled to have the same jump times and the same U at each jump satisfies $s' = -s$, and such a decay implies (4.19) (see [62]; the converse is not clear, since P_t is a mix of a Dirac mass and a smooth density).

4.4.3 A switching process

Our framework is not so well fitted to switched linear ODE, studied in [16, 29, 64] among others. Nevertheless consider the constant rate case, namely a Markov process (X_t, I_t) with generator

$$Lf(x, i) = x^* A_i \nabla_x f(x, i) + \lambda (f(x, -i) - f(x, i))$$

where $i = \pm$, $x \in \mathbb{R}^d$ and $A_i \in \mathcal{M}_d(\mathbb{R})$. It is not necessarily ergodic - even when the matrices are Hurwitz, *i.e.* matrices whose eigenvalues have negative real part - and highly depends on λ , as is enlightened in [15, 64]. We won't try here to give conditions to ensure stability. Ergodicity is quantified in [16] under a condition of strong coercivity for the flows and no condition on the rate of jump. Our approach can illustrate the interaction between those.

Let ϕ^- and ϕ^+ be the respective flows of A_+ and A_- and

$$Qf(x) = \mathbb{E}_{\mathcal{E}(\lambda)} \left[f \left(\phi_x^-(T) \right) \right].$$

Rather than the initial one we will investigate the generator

$$\tilde{L}f(x) = x^* A_+ \nabla f(x) + \lambda (Qf(x) - f(x)).$$

In other word we will pretend moving through the flow associated to A_- is instantaneous. With the notations of Section 4.2, $b(x) = A_+^* x$ so that $J_b(x) = A_+^*$. In the other hand, since $\phi_x^-(s) = e^{sA_-} x$,

$$J_{\phi_x^-(s)}(x) = e^{sA_-},$$

which yields, when $\nabla = a \nabla$,

$$|\nabla Qf|^2(x) \leq \mathbb{E}_{\mathcal{E}(\lambda)} \left[\left(\frac{a(x) \|e^{TA_-}\|}{a(\phi_x^-(T))} \right)^2 \right] Q|\nabla f|^2(x).$$

It is crucial here that the operator norm $\|\cdot\|$ is associated to the euclidian scalar product,

$$\|B\| = \sup_{x \neq 0} \frac{x^* B^* B x}{x^* x},$$

so that even if the eigenvalues of A_- all have negative real parts, $\|e^{tA_-}\|$ is not necessarily less than 1 for small t . For the sake of simplicity take $a = 1$.

Let σ_+ be the greatest eigenvalue of $\frac{1}{2} (A_+ + A_+^T)$, so that

$$\max_{|x|=1} (x^T A_+ x) = \sigma_+.$$

Now Theorems 4.5 and Corollary 4.6 give

Proposition 4.20. *Let \tilde{P}_t be the semigroup generated by \tilde{L} and*

$$\eta = \lambda \left(1 - \mathbb{E}_{\mathcal{E}(\lambda)} \left[|e^{TA_-}|^2 \right] \right) - 2\sigma_+.$$

Then

$$|\nabla_d \tilde{P}_t f|^2 \leq e^{-\eta t} \tilde{P}_t |\nabla_d f|^2.$$

If \mathcal{W}_1 is the usual Wasserstein distance associated to the euclidian distance,

$$\mathcal{W}_1 \left(\delta_x \tilde{P}_t, \delta_y \tilde{P}_t \right) \leq e^{-\frac{\eta}{2} t} |x - y|.$$

It is known that when A_+ and A_- are both Hurwitz, if the rate λ is small enough, then the process is stable : it spends enough time on each flow to be brought to zero. Indeed suppose A_- is Hurwitz. Then

$$\mathbb{E}_{\mathcal{E}(\lambda)} \left[\|e^{TA_-}\|^2 \right] \xrightarrow{\lambda \rightarrow 0} 0$$

and if $\sigma_+ < 0$ we obtain a real contraction since $\eta > 0$.

For example, consider

$$A_+ = \begin{pmatrix} -\alpha & 1 \\ -1 & -\alpha \end{pmatrix} \quad A_- = \begin{pmatrix} -\alpha & \frac{1}{h} \\ -h & -\alpha \end{pmatrix}.$$

Thus $\sigma_+ = -\alpha$ and the eigenvalues of A_- are $-\alpha \pm i$. If u and v are associated unitary eigenvectors and if $h < 1$, we compute $|u^*v| = \frac{1-h^2}{1+h^2}$. Hence for all $x \in \mathbb{R}^2$,

$$\frac{|e^{tA_-}x|^2}{|x|^2} \leq e^{-2\alpha t} \frac{1 + \frac{1-h^2}{1+h^2}}{1 - \frac{1-h^2}{1+h^2}} = \frac{e^{-2\alpha t}}{h^2},$$

which gives

$$\mathbb{E}_{\mathcal{E}(\lambda)} \left[\|e^{TA_-}\|^2 \right] \leq \frac{\lambda}{(\lambda + 2\alpha)h^2}.$$

Thus

$$\eta \geq \lambda \left(1 + \frac{2\alpha}{\lambda} - \frac{1}{\left(1 + \frac{2\alpha}{\lambda}\right)h^2} \right)$$

which is positive as soon as $\left(1 + \frac{2\alpha}{\lambda}\right)h > 1$, or in other words

$$\lambda < \frac{2\alpha}{\frac{1}{h} - 1}.$$

4.4.4 The TCP with linear rate

The main ideas of this work have been adapted from the study in [27] of the process with generator

$$Lf(x) = f'(x) + x(f(\delta x) - f(x)), \quad (4.20)$$

were $\delta \in [0, 1)$. We keep the general notations for $(P_t)_{t \geq 0}$, Q , λ and μ (for the proof of ergodicity, see [42]), and write $\text{Ent} f = \mu(f^2 \ln f^2) - \mu(f^2) \ln \mu(f^2)$. In this section, we will prove Proposition 4.3. To cope with the rate of jump that vanishes at the origin, we will apply Theorem 4.7 with a weight a that behaves linearly near 0. More precisely, let

$$\begin{aligned} a(x) &= 1 - e^{-(1-\delta)x}, \\ \phi_a(f) &= a|f'|^2, \\ W_t &= \mu(\phi_a(P_t f)). \end{aligned}$$

Lemma 4.21. *Suppose μ satisfies the weighted Poincaré inequality*

$$\forall f \in \mathcal{A}, \quad \mu(f - \mu f)^2 \leq c\mu(\phi_a(f))$$

for some $c > 0$, and let

$$\eta = \left(\frac{3 + \sqrt{5}}{2} - 1 \right)^{-1} + \ln \left(\frac{3 + \sqrt{5}}{2} \right) \simeq 1.58.$$

Then for all $\beta > \eta^{-1}$, $t > 0$ and $f \in \mathcal{A}$,

$$W_t \leq e^{-(1-\delta)\frac{\eta-\frac{1}{\beta}}{1+\beta c}t} (1 + \beta c) W_0.$$

Démonstration. Note that a is a concave function, so that

$$a(\delta x) = a(\delta x + (1-\delta)0) \geq \delta a(x) + (1-\delta)a(0) = \delta a(x).$$

Thus

$$\phi_a(Qf)(x) = a(x)\delta^2|f'(\delta x)| \leq \delta a(\delta x)|f'(\delta x)|^2 = \delta Q(\phi_a(f))(x).$$

Hence, from Lemma 4.4, for any $\beta > 0$,

$$\begin{aligned} 2\Gamma_{L,\phi_a}(f) &\geq \frac{a'}{a}\phi_a(f) - 2af'(Qf - f) + \lambda(1-\delta)\phi_a(f) \\ &\geq \left(\frac{a'}{a} + \lambda(1-\delta) - \frac{a}{\lambda\beta} \right) \phi_a(f) - \beta\lambda(Qf - f)^2. \end{aligned}$$

Note that $a(x) \leq (1-\delta)x = (1-\delta)\lambda(x)$, so that

$$\begin{aligned} \frac{a'(x)}{a(x)} + \lambda(x)(1-\delta) - \frac{a(x)}{\lambda(x)\beta} &\geq (1-\delta) \left(\frac{1}{e^{(1-\delta)x} - 1} + x - \frac{1}{\beta} \right) \\ &\geq (1-\delta) \left(\frac{1}{e^x - 1} + x - \frac{1}{\beta} \right). \end{aligned}$$

The function $g(x) = \frac{1}{e^x - 1} + x$ goes to $+\infty$ at 0 and $+\infty$ and admits a unique positive critical point for which

$$\begin{aligned} e^x &= (e^x - 1)^2 \\ \Rightarrow x &= \ln \left(\frac{3 + \sqrt{5}}{2} \right). \end{aligned}$$

Hence for all $x > 0$, $g(x) \geq g\left(\ln\left(\frac{3+\sqrt{5}}{2}\right)\right) = \eta$

$$2\Gamma_{L,\phi_a}(f) \geq (1-\delta) \left(\eta - \frac{1}{\beta} \right) \phi_a(f) - \beta\lambda(Qf - f)^2.$$

With $V_t = \mu(P_t f - \mu f)$, this yields

$$W'_t + \beta V'_t \leq -(1-\delta) \left(\eta - \frac{1}{\beta} \right) W_t,$$

and, thanks to the weighted Poincaré inequality, if $\beta\eta > 1$,

$$\begin{aligned} W'_t + \beta V'_t &\leq -(1-\delta) \frac{\eta - \frac{1}{\beta}}{1 + \beta c} (W_t + \beta V_t) \\ \Rightarrow W_t + \beta V_t &\leq e^{-(1-\delta)\frac{\eta-\frac{1}{\beta}}{1+\beta c}t} (W_0 + \beta V_0). \end{aligned}$$

Finally,

$$W_t \leq W_t + \beta V_t \leq e^{-(1-\delta)\frac{\eta-\frac{1}{\beta}}{1+\beta c}t} (W_0 + \beta V_0) \leq e^{-(1-\delta)\frac{\eta-\frac{1}{\beta}}{1+\beta c}t} (1 + \beta c) W_0.$$

□

Remark that $h(\beta) = \frac{\beta\eta-1}{\beta+\beta^2c}$ goes to 0 when β either goes to η^{-1} or to $+\infty$, and admits a unique positive critical point for which

$$\begin{aligned} \beta^2 c \eta - 2\beta c - 1 &= 0 \\ \Rightarrow \beta &= c \left(1 + \sqrt{1 + \frac{\eta}{c}} \right). \end{aligned}$$

Corollary 4.22. *Suppose μ satisfies the weighted inequalities, for all $f \in \mathcal{A}$,*

$$\mu(f - \mu f)^2 \leq c_1 \mu(\phi_a(f)),$$

$$\text{Ent} f \leq c_2 \mu(\phi_a(f)) \tag{4.21}$$

for some $c_1, c_2 > 0$, and let η be such as defined in Lemma 4.21. Then for all $\beta > \eta^{-1}$, $t > 0$ and $f \in \mathcal{A}$,

$$\text{Ent} P_t f \leq c_2 e^{-(1-\delta)\frac{\eta-\frac{1}{\beta}}{1+\beta c_1}t} (1 + \beta c_1) \mu(f')^2.$$

Démonstration. From Lemma 4.21 and the fact $a \leq 1$,

$$\text{Ent} P_t f \leq c W_t \leq c e^{-(1-\delta)\frac{\eta-\frac{1}{\beta}}{1+\beta c}t} (1 + \beta c) W_0 \leq c e^{-(1-\delta)\frac{\eta-\frac{1}{\beta}}{1+\beta c}t} (1 + \beta c) \mu(f')^2.$$

□

Thus, to prove Proposition 4.3, it only remains to prove a weighted log-Sobolev inequality holds. In order to simplify some upcoming computations, we consider an intermediate weight $\alpha(x) = 1 - e^{-\sqrt{2}x}$. Then, by concavity, $a(x) = \alpha\left(\frac{1-\delta}{\sqrt{2}}x\right) \geq \frac{1-\delta}{\sqrt{2}}\alpha(x)$. If we prove a log-Sobolev inequality (4.21) holds with weight α , it implies such an inequality with weight a .

Let

$$\psi(x) = \int_0^\infty \frac{1}{\sqrt{\alpha(y)}} dy.$$

It is a concave, non-decreasing, one-to-one function. If Z is a random variable with law μ and $Y = \psi(Z)$, then

$$\begin{aligned} \mathbb{E}\left(f^2(Z) \ln f^2(Z)\right) - \mathbb{E}\left(f^2(Z)\right) \ln \mathbb{E}\left(f^2(Z)\right) &\leq c \mathbb{E}\left(\alpha(Z)(f')^2(Z)\right) \\ \Leftrightarrow \mathbb{E}\left(g^2(Y) \ln g^2(Y)\right) - \mathbb{E}\left(g^2(Y)\right) \ln \mathbb{E}\left(g^2(Y)\right) &\leq c \mathbb{E}\left((g')^2(Y)\right) \end{aligned}$$

with $g(y) = f(\psi^{-1}(y))$. As a consequence we will study the Markov process $\psi(X) = (\psi(X_t))_{t \geq 0}$, where $X = (X_t)_{t \geq 0}$ has generator (4.20), and prove a classical log-Sobolev for the invariant measure of this twisted process, which will imply the weighted log-Sobolev assumed in Corollary 4.22.

We still denote by Q the jump kernel of X , so that

$$Q_\alpha g(z) = g\left(\psi\left(\delta\psi^{-1}(z)\right)\right)$$

is the jump kernel of $\psi(X)$. Let K_α and \widetilde{K}_α be such as defined in Section 4.3.2, but corresponding to the process $\psi(X)$.

Lemma 4.23. For all $g \in \mathcal{A}$,

$$\begin{aligned} |(Q_\alpha g)'| &\leq \sqrt{\delta} Q_\alpha |g'| \\ |(K_\alpha g)'| &\leq K_\alpha |g'| \\ |(\widetilde{K}_\alpha g)'| &\leq \widetilde{K}_\alpha |g'|. \end{aligned}$$

Démonstration. By concavity, $\alpha(\delta x) \geq \delta \alpha(x)$ for all $x \geq 0$. Thus

$$\begin{aligned} (Q_\alpha g)'(z) &= \delta (\psi^{-1})'(z) \psi'(\delta \psi^{-1}(z)) Q_\alpha g'(z) \\ &= \frac{\delta (\psi^{-1})'(z)}{\sqrt{\alpha(\delta \psi^{-1}(z))}} Q_\alpha g'(z) \\ &\leq \frac{\sqrt{\delta} (\psi^{-1})'(z)}{\sqrt{\alpha(\psi^{-1}(z))}} Q_\alpha |g'| (z) \\ &= \sqrt{\delta} Q_\alpha |g'| (z). \end{aligned}$$

On the other hand the vector field associated to $\psi(X)$ is $b_\alpha(z) = \frac{1}{\sqrt{\alpha(\psi^{-1}(z))}}$, and the rate of jump is non-decreasing along the flow. Hence, according to Lemma 4.13,

$$b_\alpha |(K_\alpha g)'| \leq K_\alpha (b_\alpha |g'|)$$

(and according to Lemma 4.14, the same goes for \widetilde{K}_α). Note that the support of both probability measure $K_\alpha(z)$ and $\widetilde{K}_\alpha(z)$ is $[z, \infty]$, and that b_α is non-increasing along the flow, so that

$$|(K_\alpha g)'|(z) \leq \frac{K_\alpha (b_\alpha |g'|)(z)}{b_\alpha(z)} \leq K_\alpha (|g'|)(z)$$

(and the same goes for \widetilde{K}_α). □

Lemma 4.24. The law $K_\alpha(z)$ (resp. $\widetilde{K}_\alpha(z)$) can be obtained from $K_\alpha(0)$ (resp. $\widetilde{K}_\alpha(0)$) through a 1-Lipschitz transformation.

Démonstration. Let T_x be the first time of jump of X starting from x . According to Lemma 4.11, there exists a 1-Lipschitz functions G such that $T_x \stackrel{dist}{=} G(T_0)$. Note that $K_\alpha(\psi(x))$ is the law of $\psi(x + T_x)$. Let $H(z) = \psi(x + G(\psi^{-1}(z)))$, so that $\psi(x + T_x) \stackrel{dist}{=} H(\psi(T_0))$. We compute

$$\begin{aligned} |H'(z)| &= |G'(\psi^{-1}(z)) (\psi^{-1})'(z) \psi'(x + G(\psi^{-1}(z)))| \\ &\leq \frac{\psi'(x + G(\psi^{-1}(z)))}{\psi'(\psi^{-1}(z))}. \end{aligned}$$

Now ψ is concave, and in the proof of Lemma 4.11 we have seen that $x + G(s) \geq s$ for all $s \geq 0$; hence $|H'(z)| \leq 1$ for all $z \geq 0$.

Similarly, let \widetilde{T}_x be a random variable on \mathbb{R}_+ with density $\frac{e^{-\int_0^t (x+s)ds}}{\int_0^\infty e^{-\int_0^u (x+s)ds} du}$, so that $\widetilde{K}_\alpha(\psi(x))$

is the law of $\psi(x + \widetilde{T}_x)$. From Lemma 4.12 there exists a 1-Lipschitz functions \widetilde{G} such that $\widetilde{T}_x \stackrel{dist}{=} \widetilde{G}(\widetilde{T}_0)$, and the previous argument concludes. □

Lemma 4.25. Both $K_\alpha(0)$ and $\widetilde{K}_\alpha(0)$ satisfies $\mathcal{B}(1, 1)$.

Démonstration. If T_0 is the first time of jump starting from 0 then $K_\alpha(0)$ is the law of $\psi(T_0)$. For any $f \in \mathcal{A}$,

$$\begin{aligned} K_\alpha f(0) &= \int_0^\infty f(\psi(u)) u e^{-\frac{u^2}{2}} du \\ &= \int_0^\infty f(z) e^{-\frac{(\psi^{-1}(z))^2}{2} + \ln \psi^{-1}(z) + \frac{1}{2} \ln(\alpha(\psi^{-1}(z)))} du. \end{aligned}$$

On the other hand, if N is a standard gaussian variablen then $\widetilde{K}_\alpha(0)$ is the law of $\psi(|N|)$, and for all $f \in \mathcal{A}$

$$\begin{aligned} \widetilde{K}_\alpha f(0) &= \int_0^\infty f(\psi(u)) \left(\frac{\pi}{2}\right)^{-\frac{1}{2}} e^{-\frac{u^2}{2}} du \\ &= \int_0^\infty f(z) e^{-\frac{(\psi^{-1}(z))^2}{2} + \frac{1}{2} \ln(\alpha(\psi^{-1}(z))) - \frac{1}{2} \ln\left(\frac{\pi}{2}\right)} du. \end{aligned}$$

For $\varepsilon \in \{0, 1\}$, let $V_\varepsilon(z) = \frac{1}{2}(\psi^{-1}(z))^2 - \varepsilon \ln \psi^{-1}(z) - \frac{1}{2} \ln(\alpha(\psi^{-1}(z)))$; we want to prove V_ε is strictly convex. Writing $x = \psi^{-1}(z)$, we compute

$$\begin{aligned} V'_\varepsilon(z) &= \sqrt{\alpha(x)} \left(x - \frac{\varepsilon}{x} - \frac{\alpha'(x)}{2\alpha(x)} \right) \\ V''_\varepsilon(z) &= \frac{\alpha'(x)}{2} \left(x - \frac{\varepsilon}{x} - \frac{\alpha'(x)}{2\alpha(x)} \right) + \alpha(x) \left(1 + \frac{\varepsilon}{x^2} - \frac{\alpha''(x)}{2\alpha(x)} + \frac{1}{2} \left(\frac{\alpha'(x)}{\alpha(x)} \right)^2 \right) \\ &= \varepsilon \left(\frac{\alpha(x)}{x^2} - \frac{\alpha'(x)}{2x} \right) + \frac{\alpha'(x)x}{2} + \frac{(\alpha'(x))^2}{4\alpha(x)} + \alpha(x) - \frac{1}{2}\alpha''(x). \end{aligned}$$

As a first step, note that $V_1''(z) \geq V_0''(z)$: indeed, $V_1''(z) - V_0''(z) = \frac{h(x)}{x^2}$ with

$$\begin{aligned} h(x) &= \alpha(x) - \frac{x}{2}\alpha'(x) \\ \Rightarrow h'(x) &= \frac{1}{2}\alpha'(x) - \frac{x}{2}\alpha''(x) > 0 \end{aligned}$$

(since α is non-decreasing and concave). Since $h(0) = 0$, it implies $h(x) \geq 0$ for all $x \geq 0$, in other words $V_1''(z) \geq V_0''(z)$. On the other hand,

$$\begin{aligned} V_0''(z) &\geq \alpha(x) - \frac{1}{2}\alpha''(x) \\ &= 1, \end{aligned}$$

the last ligne being the very reason we decided to work with α rather than with a . As a consquence, both $K_\alpha(0)$ and $\widetilde{K}_\alpha(0)$ satisfies $\mathcal{B}(1, 1)$ (see for instance [5, Theorem 5.4.7], applied to the diffusion with generator $\partial_x^2 - V_\varepsilon \partial_x$). \square

To sum up the consequences of the previous result,

Corollary 4.26.

1. The operators K_α and \widetilde{K}_α are $(1, 1, 1)$ -confining and the operator Q_α is $(0, \sqrt{\delta}, 1)$ -contractive.
2. The invariant measure ν_α of the embedded chain associated to $\psi(X)$ satisfies $\mathcal{B}\left(1, \frac{\sqrt{\delta}}{1-\sqrt{\delta}}\right)$.

Démonstration. The sub-commutation property has been showed in Lemma 4.23, and the local inequality is clear for Q_α which is deterministic, and is a consequence of Lemma 4.24 and 4.25 for K_α and \widetilde{K}_α .

From Lemma 4.9, the transition operator of the embedded chain associated to $\psi(X)$, $P_\alpha = K_\alpha Q_\alpha$, is $(\sqrt{\delta}, \sqrt{\delta}, 1)$ -confining, conclusion follows again from Lemma 4.9. \square

The last step of our procedure is the study of a perturbation of ν_α . Since the rate of jump of $Z = \psi(X)$ at point z is $\lambda_\alpha(z) = \psi^{-1}(z)$ and the operator K_α is such that $K_\alpha f(\psi(x)) = \mathbb{E}(f(\psi(x + T_x)))$, according to Lemma 4.10, we have to study the function g defined by

$$\begin{aligned} g(\psi(x)) &= K_\alpha\left(\frac{1}{\lambda_\alpha}\right)(\psi(x)) \\ &= \mathbb{E}\left(\frac{1}{x + T_x}\right) \end{aligned}$$

Lemma 4.27. *The function g is decreasing, and $\ln g$ is $\frac{2}{\sqrt{\pi}}$ -Lipschitz.*

Démonstration. Let

$$h(x) = \mathbb{E}\left(\frac{1}{x + T_x}\right) = \int_0^\infty e^{-\int_0^t (x+u) du} dt = \int_0^\infty e^{-t^2 - 2xt} dt,$$

so that $g(z) = h(\psi^{-1}(z))$. Since h is decreasing and ψ^{-1} is increasing, g is decreasing. Moreover, as $|(\ln g)'(z)| = \sqrt{\alpha}(\psi^{-1}(s))|(\ln h)'(\psi^{-1}(z))|$ and $\alpha \leq 1$, it is sufficient to prove $\ln h$ is $\frac{2}{\sqrt{\pi}}$ -Lipschitz. From $h(x) = e^{x^2} \int_x^\infty e^{-s^2} ds$, one gets

$$\begin{aligned} h'(x) &= -1 + 2xh(x) \\ \Rightarrow (\ln h)'(x) &= -\frac{1}{h(x)} + 2x \\ \Rightarrow (\ln h)''(x) &= \frac{(\ln h)'(x)}{h(x)} + 2 \\ \Rightarrow \left(\frac{1}{h}(\ln h)'\right)'(x) &= \frac{2}{h(x)} \geq 0 \\ \Rightarrow (\ln h)'(x) &\geq \frac{h(x)}{h(0)}(\ln h)'(0) \\ &\geq \frac{h'(0)}{h(0)} = -\frac{1}{h(0)}. \end{aligned}$$

On the other hand $(\ln h)' \leq 0$; thus $\ln g$ is $\frac{1}{h(0)}$ -Lipschitz, and

$$h(0) = \int_0^\infty e^{-t^2} dt = \frac{\sqrt{\pi}}{2}.$$

\square

We have in mind to apply to ν_α and g the perturbation Lemma 4.29 of the Annex. In that aim we need to bound $g(m_\alpha)$, where m_α is the median of ν_α , and $\nu_\alpha(g^{-1})$. In fact, note that ν_α , which is the invariant measure of the embedded chain associated to the process $\psi(X)$, is also the image through the function ψ of the invariant measure of μ_e , the invariant measure of the embedded chain associated to the initial process X . In particular if m_e is the median of μ_e then $m_\alpha = \psi(m_e)$. Keeping the notation $h(x) = g(\psi(x))$, we have $g(m_\alpha) = h(m_e)$ and $\nu_\alpha(g^{-1}) = \mu_e(h^{-1})$.

Lemma 4.28. *We have*

$$\frac{h(0)}{h(m_e)} \leq \frac{2\pi}{5} \left(1 + \frac{\delta}{\sqrt{1-\delta^2}} \right)$$

and

$$\mu_e(h^{-1}) \leq \frac{6}{5} \left(2 + \frac{\delta\sqrt{2}}{\sqrt{1-\delta^2}} \right).$$

Démonstration. Recall that, keeping the notations of Section 4.3.2, if T_x is the first time of jump of the process starting from x and E is a standard exponential variable, then $T_x \stackrel{dist}{=} \Lambda_x^{-1}(E)$. In the present case $\Lambda_x(t) = \int_0^t (x+u)du$, so that $T_x \stackrel{dist}{=} \sqrt{x^2 + 2E} - x$. In particular if Y is a random variable with measure μ_e , $Y \stackrel{dist}{=} \delta\sqrt{Y^2 + 2E}$, or in other words $Y^2 \stackrel{dist}{=} \delta^2(Y^2 + 2E)$, so that

$$(1 - \delta^2)\mathbb{E}(Y^2) = 2\delta^2\mathbb{E}(E) = 2\delta^2.$$

From this,

$$\mathbb{P}(Y \geq t) \leq \frac{\delta^2}{(1 - \delta^2)t^2},$$

which implies

$$m_e \leq \frac{2\delta}{\sqrt{1-\delta^2}}.$$

Moreover

$$\begin{aligned} h(x) &= \mathbb{E}\left(\frac{1}{x+T_x}\right) \\ &\geq \frac{1}{x+2}\mathbb{P}(T_x \leq 2) \\ &\geq \frac{1}{x+2}\mathbb{P}(T_0 \leq 2) \\ &= \frac{1}{x+2}\mathbb{P}(\sqrt{2E} \leq 2) \\ &= \frac{1}{x+2}(1 - e^{-2}) \geq \frac{5}{6(x+2)}. \end{aligned}$$

Hence

$$\begin{aligned} \frac{h(0)}{h(m_e)} &\leq \frac{\sqrt{\pi}}{2} \times \frac{6}{5}(m_e + 2) \\ &\leq \frac{2\pi}{5} \left(1 + \frac{\delta}{\sqrt{1-\delta^2}} \right). \end{aligned}$$

Finally, if Y is a random variable with measure μ_e ,

$$\begin{aligned}\mu_e(h^{-1}) &= \mathbb{E}\left(\frac{1}{h(Y)}\right) \\ &\leq \frac{6}{5}\mathbb{E}(Y+2) \\ &\leq \frac{6}{5}\left(2 + \sqrt{\mathbb{E}(Y^2)}\right).\end{aligned}$$

□

We can now bring the pieces together.

proof of Proposition 4.3. We have proved in Corollary 4.26 that ν_α satisfies a log-Sobolev inequality. From Lemma 4.27, 4.28 and 4.29, the perturbation ν_g of ν_α defined by $\nu_g f = \frac{1}{\nu_\alpha(g)}\nu_\alpha(gf)$ also satisfies such an inequality. From Lemma 4.10, the invariant measure of $\psi(X)$ is $\tilde{\nu}_g K_\alpha$, and it also satisfies a log-Sobolev inequality since \tilde{K}_α is confined (Corollary 4.26). It means μ , the invariant measure of X , satisfies a weighted log-Sobolev inequality

$$\mu(f^2 \ln f^2) - \mu(f^2) \ln \mu(f^2) \leq c\mu(\alpha|f'|^2).$$

As we have already noticed, $\alpha \leq \frac{\sqrt{2}}{1-\delta}a$, so that the conditions of Corollary 4.22 are fulfilled, and Proposition 4.3 is proved. □

Annex

Monotonous perturbation on the half-line

Let ν be a probability measure on \mathbb{R}_+ with a positive smooth density (still denoted by ν), and g be a positive smooth function on \mathbb{R}_+ such that $\nu(g) = 1$. We define ν_g , the perturbation of ν by g , by $\nu_g(f) = \nu(fg)$ for all bounded f . Let m be the median of ν , defined by $\nu([0, m]) = \frac{1}{2}$.

The aim of this section is to prove the following :

Lemma 4.29. *Suppose g is non-increasing and $g(0) := \lim_{x \rightarrow 0} g(x) \neq \infty$.*

1. *If ν satisfies the Poincaré inequality $\mathcal{B}(2, c_1)$, then ν_g satisfies $\mathcal{B}(2, c_2)$ with*

$$c_2 = 8 \frac{g(0)}{g(m)} c_1.$$

2. *If $\ln g$ is κ -Lipschitz and ν satisfies the log-Sobolev inequality $\mathcal{B}(1, c_1)$ then ν_g satisfies $\mathcal{B}(1, c_2)$ with for all $\varepsilon \in (0, 1)$*

$$c_2 \leq \left(\frac{2}{1-\varepsilon} + 8 \frac{g(0)}{g(m)} \left(2 + \frac{c_1 \kappa^2 + \varepsilon \ln \nu(g^{1-\frac{1}{\varepsilon}})}{1-\varepsilon} \right) \right) c_1.$$

Remark : actually as far as point 2 is concerned the monotonicity of g is only needed to get the explicit estimate of c_2 : as soon as ν satisfies a log-Sobolev inequality and $\ln g$ is Lipschitz, ν_g satisfies a log-Sobolev inequality (see [2]).

Moreover when ν satisfies a log-Sobolev inequality and $\ln g$ is Lipschitz, $\nu(g^\alpha)$ is finite for all $\alpha \in \mathbb{R}$ (see [1]), so that c_2 is finite.

proof of point 1. According to Muckenhoupt work (see [5, Theorem 6.2.2 p. 99]), a probability with density $h > 0$ satisfies $\mathcal{B}(2, c)$ iff $B_{m_h}(h)$ is finite when m_h is the median of $h(t)dt$ and

$$B_\alpha(h) = \max \left(\sup_{x \in (\alpha, \infty)} \left(\int_x^\infty h(t)dt \int_\alpha^x \frac{1}{h(t)}dt \right), \sup_{x \in (0, \alpha)} \left(\int_0^x h(t)dt \int_x^\alpha \frac{1}{h(t)}dt \right) \right).$$

Furthermore, in that case, the optimal c (namely the smallest c such that $\mathcal{B}(2, c)$ holds) is such that

$$\frac{1}{2} \inf_{\alpha > 0} B_\alpha(h) \leq \frac{1}{2} B_{m_h}(h) \leq c \leq 4 \inf_{\alpha > 0} B_\alpha(h) \leq 4 B_{m_h}(h).$$

In the present case, for all $x \geq m$,

$$\begin{aligned} \int_x^\infty g(t)\nu(t)dt \int_m^x \frac{1}{g(t)\nu(t)}dt &\leq \int_x^\infty g(x)\nu(t)dt \int_m^x \frac{1}{g(x)\nu(t)}dt \\ &\leq 2c_1. \end{aligned}$$

and for all $x \leq m$

$$\begin{aligned} \int_0^x g(t)\nu(t)dt \int_x^m \frac{1}{g(t)\nu(t)}dt &\leq \int_0^x g(0)\nu(t)dt \int_x^m \frac{1}{g(m)\nu(t)}dt \\ &\leq 2 \frac{g(0)}{g(m)} c_1. \end{aligned}$$

Hence ν_g satisfies $\mathcal{B}(2, c_2)$ with

$$c_2 \leq 4 \inf_{\alpha > 0} B_\alpha(\nu g) \leq 4 B_m(\nu g) \leq 8 \frac{g(0)}{g(m)} c_1.$$

□

proof of point 2. Following a computation of Aida and Shigekawa ([2]), we apply the inequality $\mathcal{B}(1, c_1)$, namely

$$\forall f \in \mathcal{A}, \quad \nu(f^2 \ln f^2) \leq 2c_1 \nu(f')^2 + (\nu f^2) \ln(\nu f^2),$$

to the function $f\sqrt{g}$, which reads

$$\forall f \in \mathcal{A}, \quad \nu_g(f^2 \ln f^2) + \nu_g(f^2 \ln g) \leq 2c_1 \nu_g \left(f' + \frac{g'}{2g} \right)^2 + (\nu_g f^2) \ln(\nu_g f^2). \quad (4.22)$$

From the inequality $(a+b)^2 \leq 2a^2 + 2b^2$ and the assumption on $\ln g$,

$$\nu_g \left(f' + \frac{g'}{2g} f \right)^2 \leq 2\nu_g(f')^2 + \frac{1}{2} \kappa^2 \nu_g(f^2).$$

On the other hand, from the Young inequality $st \leq s \ln s - s + e^t$ applied with $s = \varepsilon f^2$ and $t = -\varepsilon^{-1} \ln \left(\frac{g}{g(0)} \right)$ for any $\varepsilon > 0$,

$$\begin{aligned} -\nu_g(f^2 \ln g) &= -\nu_g \left(f^2 \ln \left(\frac{g}{g(0)} \right) \right) - \ln g(0) \nu_g(f^2) \\ &\leq \varepsilon \nu_g(f^2 \ln f^2) - (\varepsilon(1 - \ln \varepsilon) + \ln g(0)) \nu_g(f^2) + \nu_g \left(\left(\frac{g(0)}{g} \right)^{\frac{1}{\varepsilon}} \right). \end{aligned}$$

Thus Inequality (4.22) yields

$$(1 - \varepsilon)\nu_g(f^2 \ln f^2) \leq 4c_1\nu_g(f')^2 + (c_1\kappa^2 - \varepsilon(1 - \ln \varepsilon) - \ln g(0))\nu_g(f^2) + \nu_g\left(\left(\frac{g(0)}{g}\right)^{\frac{1}{\varepsilon}}\right) + \nu_g(f^2) \ln \nu_g(f^2).$$

Thanks to Gross' Lemma (2.2 of [51]), this implies (for $\varepsilon < 1$)

$$\nu_g(f^2 \ln f^2) - \nu_g(f^2) \ln \nu_g(f^2) \leq \frac{4c_1}{1 - \varepsilon}\nu_g(f')^2 + \gamma\nu_g(f^2) \quad (4.23)$$

with

$$\begin{aligned} \gamma &= \frac{c_1\kappa^2 - \varepsilon(1 - \ln \varepsilon) - \ln g(0)}{1 - \varepsilon} + \frac{\varepsilon}{1 - \varepsilon} \left(1 + \ln \left(\frac{\nu_g\left(\left(\frac{g(0)}{g}\right)^{\frac{1}{\varepsilon}}\right)}{\varepsilon} \right) \right) \\ &= \frac{c_1\kappa^2 + \varepsilon \ln \nu_g\left(g^{-\frac{1}{\varepsilon}}\right)}{1 - \varepsilon}. \end{aligned}$$

It is classical to retrieve a log-Sobolev inequality from Inequality (4.23) and a Poincaré inequality, thanks to the following inequality (see [34], p.146) : if $h = f - \nu_g f$,

$$\nu_g(f^2 \ln f^2) - \nu_g(f^2) \ln \nu_g(f^2) \leq \nu_g(h^2 \ln h^2) - \nu_g(h^2) \ln \nu_g(h^2) + 2\nu_g(h^2).$$

Together with Inequality (4.23) applied to h , and since $h' = f'$,

$$\nu_g(f^2 \ln f^2) - \nu_g(f^2) \ln \nu_g(f^2) \leq \frac{4c_1}{1 - \varepsilon}\nu_g(f')^2 + (\gamma + 2)\nu_g((f - \nu_g f)^2)$$

Since ν satisfies $\mathcal{B}(1, c_1)$ it also satisfies $\mathcal{B}(2, 2c_1)$. Thus, according to point 1 of Lemma 4.29, ν_g satisfies $\mathcal{B}\left(2, 16\frac{g(0)}{g(m)}c_1\right)$, which means

$$\nu_g(f^2 \ln f^2) - \nu_g(f^2) \ln \nu_g(f^2) \leq 2\left(\frac{2}{1 - \varepsilon} + 8\frac{g(0)}{g(m)}(2 + \gamma)\right)c_1\nu_g(f')^2.$$

□

Chapitre 5

Processus RTP complètement dégénéré

De cette façon, toute connaissance dégénère en probabilité et cette probabilité est plus ou moins grande selon notre expérience de la véracité ou de la fausseté de notre entendement, et selon la simplicité ou la complexité de la question.

David Hume

Les trois premiers chapitres s'attachaient à comprendre des phénomènes de convergence hypocoercive. Le bilan en est mitigé. En particulier aucune des différentes approches ne semble convenir pour le processus RTP complètement dégénéré (sur \mathbb{R}), que nous allons maintenant étudier et dont nous rappelons la définition, via son générateur

$$L_\varepsilon f(x, y) = yf'(x, y) + \frac{1}{\varepsilon} (yU'(x))_+ (f(x, -y) - f(x, y)), \quad x \in \mathbb{R}, y \in \{-1, 1\}. \quad (5.1)$$

En y réfléchissant un peu, les difficultés rencontrées à son égard étaient prévisibles. Supposons que $(X_0, Y_0) = (x_1, 1)$ et que le potentiel U décroît sur un intervalle $[x_1 - \eta, x_2 + \eta]$. Dans ce cas aucun saut n'est possible tant que le processus n'atteint pas $x_2 + \eta$, l'évolution est déterministe et pour tout $x \in [x_1, x_2]$ et $t \leq \eta$,

$$\delta_{(x,1)} P_t = \delta_{(x+t,1)}.$$

En particulier il n'y a aucune propriété de régularisation (la loi transportée le long du semi-groupe est toujours étrangère à la mesure de Lebesgue). Il n'y a pas non plus de contraction du gradient, au sens où, toujours pour $x \in [x_1, x_2]$ et $t \leq \eta$,

$$\partial_x (P_t f)(x, 1) = P_t \partial_x f(x, 1).$$

Le carré du champs est nul, et

$$\begin{aligned} \|P_t f - \mu f\|^2 &= \|f - \mu f\|^2 - 2 \int_0^t \mu \Gamma P_s f ds \\ &= \|f - \mu f\|^2 \end{aligned}$$

si le support de f est inclus dans $[x_1, x_2] \times \{1\}$ et $t \leq \eta$. Bref, difficile d'envisager étudier l'ergodicité de ce processus par des arguments infinitésimaux en temps.

D'un autre côté, en dimension un au moins, la dynamique est plutôt simple. Rappelons que lorsque le processus commence à remonter le long du potentiel, le prochain temps de saut T est donné par une v.a. $E \sim \mathcal{E}(1)$ et la relation

$$\varepsilon E = \int_0^T y U'(x + ys) ds = U(x + T) - U(x),$$

(sauf si entre temps le processus a atteint un point critique du potentiel et a recommencé à descendre). En particulier si $(X_0, Y_0) = (x_0, 1)$ et que le potentiel est croissant sur $[x_0, x_1]$, la probabilité que le processus atteigne x_1 du premier coup, c'est-à-dire avant d'être redescendu à $(x_0, -1)$, est simplement

$$\mathbb{P}(\varepsilon E \geq U(x_1) - U(x_0)) = e^{-\frac{U(x_1) - U(x_0)}{\varepsilon}}.$$

Ceci va nous permettre d'estimer, quand la température ε tend vers 0, le temps de sortie du bassin d'attraction d'un minimum local du potentiel. En considérant ensuite la dynamique discrète à l'échelle des régions métastables, c'est-à-dire la façon dont le processus saute d'un bassin local à un autre, on pourra montrer une condition nécessaire et suffisante sur un schéma de température pour que le processus inhomogène de générateur $t \mapsto L_{\varepsilon_t}$ converge en probabilité vers tout voisinage des minima globaux du potentiel. Nous verrons qu'elle exige les mêmes schémas logarithmiques que les théorèmes 1.17 et 1.20 pour la diffusion réversible.

Bien que le recuit simulé ne soit pas un algorithme réellement pertinent en dimension un, nous restreignons notre étude à ce cadre. Comme discuté au chapitre 3, un candidat ergodique en dimension quelconque pour généraliser le générateur (5.1) serait

$$L_\varepsilon f(x, y) = y \cdot \nabla_x f(x, y) + \frac{(y \cdot \nabla_x U(x))_+}{\varepsilon} (f(x, -y) - f(x, y)) + r \left(\int f(x, z) dz - f(x, y) \right).$$

La difficulté supplémentaire, par rapport au cas de la dimension un, est dès lors de comprendre par quels chemins le RTP, dont les trajectoires sont des lignes brisées, s'échappe des régions métastables; ou, pour reprendre les termes de Lelièvre ([66]), comment le RTP se comporte face à des barrières entropiques plutôt qu'énergétiques. En particulier cela pourrait dépendre de la façon dont le paramètre r varie avec la température. Pour l'heure nous laisserons de côté cette étude des "chemins de réaction" (*reactive paths* en anglais; voir [25] sur la question dans le cas d'une diffusion).

5.1 Main results

As a first step we will consider the process $(X_t, Y_t)_{t \geq 0}$ on $\mathbb{R} \times \{\pm 1\}$ at low (but fixed) temperature, namely with generator (5.1). We want to understand how long it takes for the RTP to escape from a local minimum. Let U be a Morse potential with two local minima, denoted by $x_0 \leq x_2$, a local maximum $x_1 \in (x_0, x_2)$, and which is convex outside (x_0, x_2) . Recall that U is said to be a Morse function if all its critical points are non-degenerate; in other words if $(U'(x) = 0) \Rightarrow (U''(x) \neq 0)$. Suppose $(X_0, Y_0) = (x_0, -1)$, and let

$$\tau = \inf \{t, X_t = x_1\}$$

be the exit time of $[x_0, x_1]$ (note that, contrary to a diffusion which may fall back, when the RTP reaches x_1 it deterministically leaves $[x_0, x_1]$ and falls down to x_2). Then the energy barrier to overcome in order to leave $[x_0, x_1]$ is $U(x_1) - U(x_0)$. We will prove what is usually called an Eyring–Kramers formula (or an Arrhenius law)

Theorem 5.1.

$$\begin{aligned}\mathbb{E}[\tau] &= \sqrt{\frac{8\pi\varepsilon}{U''(x_0)}} e^{\frac{U(x_1)-U(x_0)}{\varepsilon}} \left(1 + o_{\varepsilon \rightarrow 0}(1)\right) \\ \mathbb{P}(\tau \geq t\mathbb{E}[\tau]) &\xrightarrow{\varepsilon \rightarrow 0} e^{-t}\end{aligned}$$

This can be compared to the case of a Fokker-Planck diffusion Z_t with generator

$$L_{di}f(x) = -U'(x)f'(x) + \varepsilon f''(x),$$

which has been studied in much more general settings. Let $\eta > 0$ be small and

$$\tau_{di} = \inf \{t, Z_t = x_1 + \eta\}.$$

The work of Bovier & al applies here and yields :

Theorem 5.2 (from [20, 21]).

$$\begin{aligned}\mathbb{E}[\tau_{di}] &= \frac{2\pi e^{\frac{U(x_1)-U(x_0)}{\varepsilon}}}{\sqrt{|U''(x_1)|U''(x_0)}} \left(1 + o_{\varepsilon \rightarrow 0}(1)\right) \\ \mathbb{P}(\tau_{di} \geq t\mathbb{E}[\tau_{di}]) &\xrightarrow{\varepsilon \rightarrow 0} e^{-t}.\end{aligned}$$

Remarks :

- Both processes sample the same Gibbs law. To simulate the diffusion, one needs to sample a Brownian motion, while the RTP can be constructed from a sequence of independant exponential variables $(E_k)_{k \geq 0}$, since T the next time of jump starting from (x, y) can be defined, as long as $Y_t U'(X_t) > 0$, as

$$T = \inf \left\{ t > 0, \int_0^t \frac{1}{\varepsilon} (yU'(x + ys))_+ ds = E \right\}. \quad (5.2)$$

- The RTP moves at constant speed. In particular it takes a constant time (constant in the sense it does not depends on ε) to pass through the interval $(x_1 - \eta, x_1 + \eta)$, whose probability under the law μ_ε is of order

$$\exp\left(-\frac{1}{\varepsilon} (\min(U(x_1 - \eta), U(x_1 + \eta)) - \min(U(x_0), U(x_2)))\right).$$

Since the ratio between the average times spent in this interval and outside of it should be of this order, it means the time between two crossing of $(x_1 - \eta, x_1 + \eta)$ needs to be of order the inverse of this probability, which explains the exponential factor of $\mathbb{E}[\tau]$ had to be expected.

- The fact $U''(x_1)$ does not appear in the RTP case is also natural. Indeed the probability that the process starting at $(x_0, 1)$ reaches x_1 in one shot (*i.e.* before coming back to $(x_0, -1)$) depends only on $U(x_1) - U(x_0)$, and not on the local geometry of the potential near x_1 . On the contrary the process stays mainly in the neighborhood of x_0 , so that $U''(x_0)$ does intervene. If U were flat, for instance $U = d(1 - 1_{[x_0-1, x_0+1]})$, an adaptation of the proof of Theorem 5.1 would yields $\mathbb{E}[\tau] \simeq 2e^{\frac{d}{\varepsilon}}$.
- Both $\mathbb{E}[\tau]$ and $\mathbb{E}[\tau_{di}]$ have the same exponential order.

With Theorem 5.1 in mind we then turn to the study of the inhomogeneous process with generator

$$L_t f(x, y) = yf'(x, y) + \frac{1}{\varepsilon_t} (yU'(x))_+ (f(x, -y) - f(x, y)), \quad (5.3)$$

where the cooling schedule $t \mapsto \varepsilon_t > 0$ is non-increasing and goes to 0. We consider a smooth Morse potential U on \mathbb{R} with a finite number of local extrema (one of which at least is a non-global minimum), unbounded and convex at infinity. We recall that z is said to be reachable from x at height V if $\max\{U(x + t(z - x)), t \in [0, 1]\} \leq V$ and we call the depth of a local minimum x the smallest V such that there exist a z with $U(z) < U(x)$ which is reachable from x at height $U(x) + V$ (the depth of a global minimum is set to $+\infty$). The critical depth of U , denoted by E^* , is then defined as the maximal among the depths of all local minima of U which are not global minima (recall Fig. 1.1 in Section 1.3).

Adapting to our settings the work of Hajek ([53]) on simulated annealing on a discrete space, we will prove the following :

Theorem 5.3. 1. *If S is a neighborhood of all local minima of U , then*

$$\lim_{t \rightarrow \infty} \mathbb{P}(X_t \in S) = 1.$$

2. *If S is a neighborhood of all local minima of depth E , and its complementary S^c is a neighborhood of all other local minima,*

$$\lim_{t \rightarrow \infty} \mathbb{P}(X_t \in S) = 0 \quad \Leftrightarrow \quad \int_0^\infty (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{E}{\varepsilon_s}} ds = \infty$$

3. *As a consequence,*

$$\forall \delta > 0 \quad \lim_{t \rightarrow \infty} \mathbb{P}\left(U(X_t) < \min_{\mathbb{R}} U + \delta\right) = 1 \quad \Leftrightarrow \quad \int_0^\infty (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{E^*}{\varepsilon_s}} ds = \infty$$

In particular if the cooling schedule is of the parametric form

$$\varepsilon_t = \frac{c}{\ln(1+t)}$$

with $c > 0$, the algorithm succeed (*i.e.* the RTP converges in probability to any neighborhood of the global minima of U) if and only if $c \geq E^*$.

It is somehow a negative result : it means this process does not allow faster cooling schedules than the classical reversible diffusion one. In the other hand it is somehow a positive result, since it allows cooling schedule as fast as the diffusion *and* it is easier and faster to compute numerically. Finally, positive or negative, it is above all a theoretical result. In practice the

simulation is done in a finite time horizon ; a context where the theoretical logarithmic schedules are far from efficient. The next step of the study should thus be to give non-asymptotical results in the spirit of the work of Catoni ([23]).

In the rest of the chapter are proved the above results. Section 5.2 is concerned with the low-temperature regime for a double-well potential ; Theorem 5.1 is proved, which introduces a discussion on the timescale of total variation and Wasserstein convergence to equilibrium. The proof of Theorem 5.3, in inhomogeneous settings, is addressed in Section 5.3. Finally Section 5.4 investigate the non-minimal rate case.

5.2 Escape time at low temperature

In this section the temperature $\varepsilon > 0$ is fixed through time, (X_t, Y_t) is the Markov process with generator L defined by (5.1) ; the double-well potential U has three local extrema $x_0 < x_1 < x_2$ such as described in the previous section in the settings of Theorem 5.1.

5.2.1 Eyring–Kramers formula

Suppose $(X_0, Y_0) = (x_0, -1)$ and let

$$\tau = \inf \{t, X_t = x_1\}.$$

We start the proof of Theorem 5.1 with the following Lemma :

Lemma 5.4. *for $\delta > 0$ small enough,*

$$\int_0^\delta \frac{t}{\varepsilon} (-U'(x_0 - t)) e^{-\frac{U(x_0-t)-U(x_0)}{\varepsilon}} dt = \sqrt{\frac{\pi\varepsilon}{2U''(x_0)}} \left(1 + o_{\varepsilon \rightarrow 0}(1)\right).$$

Démonstration.

$$\begin{aligned} \int_0^\delta \frac{t}{\varepsilon} (-U'(x_0 - t)) e^{-\frac{U(x_0-t)-U(x_0)}{\varepsilon}} dt &= \int_0^{\frac{\delta}{\sqrt{\varepsilon}}} s (-U'(x_0 - \sqrt{\varepsilon}s)) e^{-\frac{U(x_0-\sqrt{\varepsilon}s)-U(x_0)}{\varepsilon}} ds \\ &= \sqrt{\varepsilon} \int_0^\infty f_\varepsilon(s) ds \end{aligned}$$

with

$$f_\varepsilon(s) = \frac{s (-U'(x_0 - \sqrt{\varepsilon}s))}{\sqrt{\varepsilon}} e^{-\frac{U(x_0-\sqrt{\varepsilon}s)-U(x_0)}{\varepsilon}} 1_{s \leq \delta/\sqrt{\varepsilon}}.$$

On one hand

$$f_\varepsilon(s) \xrightarrow{\varepsilon \rightarrow 0_+} s^2 U''(x_0) e^{-\frac{U''(x_0)}{2} s^2}$$

and on the other hand, writing $M = \sup_{t \in [0, \delta]} |U^{(3)}(x_0 + t)|$,

$$|f_\varepsilon(s)| \leq s^2 \left(U''(x_0) + \frac{\delta M}{2} \right) e^{-\frac{U''(x_0)}{2} s^2 + \frac{\delta s^2 M}{6}}.$$

As M decreases with δ , for δ small enough, $\delta M < 3U''(x_0)$. Thus by the dominated convergence theorem

$$\begin{aligned} \int_0^\infty f_\varepsilon(s) ds &\xrightarrow{\varepsilon \rightarrow 0} \int_0^\infty s^2 U''(x_0) e^{-\frac{U''(x_0)}{2} s^2} ds \\ &= \sqrt{\frac{\pi}{2U''(x_0)}} \end{aligned}$$

□

Proof of Theorem 5.1. The principle of this proof is quite simple : starting at $(x_0, -1)$, the RTP starts to climb to the left. After a time with bounded expectation (since the potential is convex there), it turns back, reaches $(x_0, 1)$ and starts to climb towards x_1 . If it reaches x_1 then this attempt to escape is a success. If not, it goes back to $(x_0, -1)$, and starts anew. Since the duration of the last (and successful) attempt is negligible with respect to the expected duration of a failed attempt (of order $\sqrt{\varepsilon}$ according to Lemma 5.4) times the expected number of failed attempts, conclusion follows with a law of small numbers.

More precisely let $\tau_0 = 0$, and for all $k \in \mathbb{N}$

$$\begin{aligned} \tilde{\tau}_k &= \min(\inf \{t > \tau_k, (X_t, Y_t) = (x_0, 1)\}, \tau) \\ \tau_{k+1} &= \min(\inf \{t > \tilde{\tau}_k, (X_t, Y_t) = (x_0, -1)\}, \tau). \end{aligned}$$

We also define $N = \max \{k, \tau_k < \tau\}$, so that

$$\tau = \tau - \tau_N + \sum_{k=1}^N (\tau_k - \tau_{k-1}).$$

Conditionally to $N \geq k$, $\tau_k - \tau_{k-1}$ is the length of a failed attempt to reach x_1 ; it has the same law as τ_1 , given $N \geq 1$, and is independant of N . In particular

$$\mathbb{E}[\tau] = \mathbb{E}[\tau - \tau_N] + \mathbb{E}[N] \mathbb{E}[\tau_1 \mid N \geq 1] \quad (5.4)$$

and N is a geometric variable with parameter

$$\begin{aligned} q_\varepsilon &= \mathbb{P}(\varepsilon E > U(x_1) - U(x_0)) \\ &= e^{-\frac{U(x_1) - U(x_0)}{\varepsilon}} \end{aligned}$$

(and so with expectation $\mathbb{E}[N] = q_\varepsilon^{-1}$). Now we decompose

$$\mathbb{E}[\tau_1 \mid N \geq 1] = 2\mathbb{E}\left[\frac{\tilde{\tau}_0}{2}\right] + 2\mathbb{E}\left[\frac{\tau_1 - \tilde{\tau}_0}{2} \mid N \geq 1\right]. \quad (5.5)$$

Remark $\frac{\tilde{\tau}_0}{2}$ is the jump time starting from $(x_0, -1)$, which is independant of N and whose law has density $\frac{1}{\varepsilon} (-U'(x_0 - t)) e^{-\frac{U(x_0-t) - U(x_0)}{\varepsilon}}$. For any $\delta > 0$, as soon as $\varepsilon < \frac{1}{2}$,

$$\begin{aligned} \mathbb{E}\left[\frac{\tilde{\tau}_0}{2} 1_{\frac{\tilde{\tau}_0}{2} > \delta}\right] &= \int_\delta^\infty \frac{t}{\varepsilon} (-U'(x_0 - t)) e^{-\frac{U(x_0-t) - U(x_0)}{\varepsilon}} dt \\ &\leq \int_\delta^\infty \frac{t}{\varepsilon} (-U'(x_0 - t)) e^{-\frac{U(x_0-t) - U(x_0)}{2\varepsilon} - \frac{U(x_0-\delta) - U(x_0)}{2\varepsilon}} dt \\ &\leq \frac{e^{-\frac{U(x_0-\delta) - U(x_0)}{2\varepsilon}}}{\varepsilon} \int_\delta^\infty t (-U'(x_0 - t)) e^{-(U(x_0-t) - U(x_0))} dt \\ &= o_{\varepsilon \rightarrow 0}(\sqrt{\varepsilon}). \end{aligned}$$

On the other hand thanks to Lemma 5.4, for δ small enough,

$$\mathbb{E} \left[\frac{\tilde{\tau}_0}{2} 1_{\frac{\tilde{\tau}_0}{2} < \delta} \right] = \sqrt{\frac{\pi\varepsilon}{2U''(x_0)}} \left(1 + o_{\varepsilon \rightarrow 0}(1) \right).$$

If $\delta < x_1 - x_0$, Lemma 5.4 also applies to $\frac{\tau_1 - \tilde{\tau}_0}{2}$, and since $\mathbb{P}(N \geq 1)$ goes to 1,

$$\mathbb{E} \left[\frac{\tau_1 - \tilde{\tau}_0}{2} 1_{\frac{\tau_1 - \tilde{\tau}_0}{2} < \delta} \mid N \geq 1 \right] = \sqrt{\frac{\pi\varepsilon}{2U''(x_0)}} \left(1 + o_{\varepsilon \rightarrow 0}(1) \right).$$

Furthermore

$$\begin{aligned} \mathbb{E} \left[\frac{\tau_1 - \tilde{\tau}_0}{2} 1_{\frac{\tau_1 - \tilde{\tau}_0}{2} \geq \delta} \mid N \geq 1 \right] &\leq (x_1 - x_0) \frac{\mathbb{P} \left(\frac{\tau_1 - \tilde{\tau}_0}{2} \geq \delta \right)}{\mathbb{P}(N \geq 1)} \\ &= o_{\varepsilon \rightarrow 0}(\sqrt{\varepsilon}). \end{aligned}$$

Gathering the previous estimates in Equality (5.5) we get

$$\mathbb{E}[\tau_1 \mid N \geq 1] = \sqrt{\frac{8\pi\varepsilon}{U''(x_0)}} \left(1 + o_{\varepsilon \rightarrow 0}(1) \right)$$

Besides, note that $\frac{\tilde{\tau}_N - \tau_N}{2}$ has the same law as $\frac{\tilde{\tau}_0}{2}$, so that

$$\mathbb{E}[\tau - \tau_N] = 2\mathbb{E} \left[\frac{\tilde{\tau}_N - \tau_N}{2} \right] + x_1 - x_0 = o_{\varepsilon \rightarrow 0}(1) = o_{\varepsilon \rightarrow 0}(\sqrt{\varepsilon}q_\varepsilon^{-1}),$$

and finally Equality (5.4) becomes

$$\mathbb{E}[\tau] = \sqrt{\frac{8\pi\varepsilon}{U''(x_0)}} e^{\frac{U(x_1) - U(x_0)}{\varepsilon}} \left(1 + o_{\varepsilon \rightarrow 0}(1) \right).$$

Now as far as the second assertion of the theorem is concerned,

$$\begin{aligned} \mathbb{P}(\tau \geq t\mathbb{E}[\tau]) &= \mathbb{P} \left(\tau - \tau_N + \sum_{k=1}^N (\tau_k - \tau_{k-1}) \geq t\mathbb{E}[\tau] \right) \\ &= \mathbb{P} \left(\frac{\mathbb{E}[N]\mathbb{E}[\tau_1 \mid N \geq 1]}{\mathbb{E}[\tau]} \frac{N}{\mathbb{E}[N]} \left(\frac{\tau - \tau_N}{N\mathbb{E}[\tau_1 \mid N \geq 1]} + \sum_{k=1}^N \frac{\tau_k - \tau_{k-1}}{N\mathbb{E}[\tau_1 \mid N \geq 1]} \right) \geq t \right) \end{aligned}$$

First $\frac{N}{\mathbb{E}[N]}$ converges in law to an exponential variable of parameter 1. Second, if the times of jump of the process are defined by the same sequence $(E_k)_{k \geq 0}$ of exponential variables (according to Equality (5.2)) for all ε , then the processes at different temperatures are all defined at once on the same probability space, and in this case

$$N = \min\{k \geq 0, \varepsilon E_{2k} > U(x_1) - U(x_0)\}$$

almost surely goes to infinity. Thanks to the law of large numbers,

$$\frac{\tau - \tau_N}{N\mathbb{E}[\tau_1 \mid N \geq 1]} + \sum_{k=1}^N \frac{\tau_k - \tau_{k-1}}{N\mathbb{E}[\tau_1 \mid N \geq 1]} \xrightarrow{\varepsilon \rightarrow 0} 1 \quad a.s.$$

At last $\frac{\mathbb{E}[N]\mathbb{E}[\tau_1 \mid N \geq 1]}{\mathbb{E}[\tau]} \xrightarrow{\varepsilon \rightarrow 0} 1$. At the end of the day $\frac{\tau}{\mathbb{E}[\tau]}$ converges to an exponential law, which concludes. \square

5.2.2 Metastability

This first result yields a lower bound of the total variation distance d_{tv} between

$$\nu_t = \mathcal{L}(X_t, Y_t \mid X_0 = x_0, Y_0 = -1)$$

and the invariant measure η_ε . We recall the total variation distance between two laws ν_1, ν_2 on a topological space W is defined by

$$d_{tv}(\nu_1, \nu_2) = \sup_{A \in \mathcal{B}(W)} |\nu_1(A) - \nu_2(A)|$$

where $\mathcal{B}(W)$ is the Borel σ -algebra on W . We write $t_\varepsilon = t\mathbb{E}[\tau]$.

Corollary 5.5. *Assume $U(x_2) < U(x_0)$. In that case,*

$$\liminf_{\varepsilon \rightarrow 0} d_{tv}(\nu_{t_\varepsilon}, \eta_\varepsilon) \geq e^{-t}.$$

Démonstration. The assumption $U(x_2) < U(x_0)$ implies that the first marginal of η_ε concentrates near the unique global minimum of U , x_2 . In particular

$$\eta_\varepsilon[(x_1, \infty) \times \{\pm 1\}] \xrightarrow{\varepsilon \rightarrow 0} 1.$$

However

$$\nu_t[(x_1, \infty) \times \{\pm 1\}] \leq \mathbb{P}(\tau \leq t).$$

As a consequence

$$\begin{aligned} d_{tv}(\nu_{t_\varepsilon}, \eta_\varepsilon) &\geq (\eta_\varepsilon - \nu_{t_\varepsilon})[(x_1, \infty) \times \{\pm 1\}] \\ &\geq \eta_\varepsilon[(x_1, \infty) \times \{\pm 1\}] - \mathbb{P}(\tau \leq t\mathbb{E}[\tau]) \\ &\xrightarrow{\varepsilon \rightarrow 0} e^{-t}. \end{aligned}$$

□

We will now give a more precise result concerning the first marginal and the Wasserstein distance

$$\mathcal{W}(\nu_1, \nu_2) = \sup_{\pi} \mathbb{E}_{\pi} [|X - Z|]$$

where the infimum is taken over all probability measure π on \mathbb{R}^2 whose first (resp. second) marginal is ν_1 (resp. ν_2). We start with the following lemma (recall that we fixed $(X_0, Y_0) = (x_0, -1)$):

Lemma 5.6. *Let $t > 0$. Then*

$$\limsup_{\varepsilon \rightarrow 0} \mathbb{E} [X_{t_\varepsilon}^2] < \infty.$$

Démonstration. Let $M > U(x_1)$, and for all $k \geq 0$ let

$$s_k = \inf\{s > 0, U(X_s) > M + k\}.$$

From Theorem 5.10 (stated and proved in the next section) we get that for all $k \geq 0$, there exist $\Gamma, \varepsilon_0 > 0$ such that for all $\varepsilon \leq \varepsilon_0$ and all $t > 0$,

$$\begin{aligned} \mathbb{P}(s_k \leq t_\varepsilon) &\leq 1 - \exp\left(-\Gamma\left(e^{-\frac{M+k}{\varepsilon}} + t_\varepsilon \frac{e^{-\frac{M+k-U(x_2)}{\varepsilon}}}{\sqrt{\varepsilon}}\right)\right) \\ &\leq 1 - \exp\left(-\tilde{\Gamma}e^{-\frac{M+k}{\varepsilon_0}}\right) \end{aligned}$$

where $\tilde{\Gamma}$ depends on t but not on ε . In fact since we consider here escape times from imbricated intervals, and from the convexity of U outside of a compact, Γ (and hence $\tilde{\Gamma}$) and ε_0 can be choosed uniformly over k (see the remark at the end of the proof of Theorem 5.10). Remarking that $U^{-1}(v) = \sup\{|x|, U(x) \leq v\}$ is sub-linear, we write

$$\begin{aligned} \mathbb{E}[X_{t_\varepsilon}^2] &\leq (U^{-1}(M))^2 + \sum_{k \geq 0} (U^{-1}(M+k+1))^2 \mathbb{P}(s_k \leq t_\varepsilon) \\ &\leq (U^{-1}(M))^2 + \sum_{k \geq 0} (U^{-1}(M+k+1))^2 \left(1 - \exp\left(-\tilde{\Gamma}e^{-\frac{M+k}{\varepsilon_0}}\right)\right) \\ &\leq (U^{-1}(M))^2 + \tilde{\Gamma} \sum_{k \geq 0} (U^{-1}(M+k+1))^2 e^{-\frac{M+k}{\varepsilon_0}} \end{aligned}$$

which is finite. □

Let $m_t = e^{-t}\delta_{x_0} + (1 - e^{-t})\delta_{x_2}$ and μ_t be the first marginal of ν_t , namely

$$\mu_t = \mathcal{L}(X_t \mid X_0 = x_0, Y_0 = -1).$$

Theorem 5.7. *Assume $U(x_2) < U(x_1)$. Then, t being fixed,*

$$\mathcal{W}(\mu_{t_\varepsilon}, m_t) \xrightarrow{\varepsilon \rightarrow 0} 0.$$

Démonstration. Consider $(X_t, Y_t)_{t \geq 0}$ a trajectory of the process, from which we will define a variable W of law m_t . In a first time assume $p_\varepsilon := \mathbb{P}(\tau \geq t_\varepsilon) \leq e^{-t}$. If $\tau \geq t_\varepsilon$, set $U = 1$. Else set $U = 1$ with probability $\frac{e^{-t}-p_\varepsilon}{1-p_\varepsilon}$, and else set $U = 0$. Similarly in the case where $p_\varepsilon \geq e^{-t}$, if $\tau \leq t_\varepsilon$, set $U = 0$, and else set $U = 0$ with probability $\frac{p_\varepsilon - e^{-t}}{p_\varepsilon}$, and else set $U = 1$.

Either ways, U is a Bernoulli variable with parameter e^{-t} such that

$$\mathbb{P}(U = 1_{\tau \geq t_\varepsilon}) = 1 - |e^{-t} - p_\varepsilon| \xrightarrow{\varepsilon \rightarrow 0} 1.$$

We naturally set $W = Ux_0 + (1 - U)x_2$. The Cauchy-Schwarz inequality

$$\mathbb{E}[|X_{t_\varepsilon} - W|1_{U \neq 1_{\tau \geq t_\varepsilon}}] \leq \sqrt{\mathbb{E}[|X_{t_\varepsilon} - W|^2] \mathbb{P}(U \neq 1_{\tau \geq t_\varepsilon})},$$

together with Lemma 5.6, yields

$$\mathbb{E}[|X_{t_\varepsilon} - W|] = \mathbb{E}[|X_{t_\varepsilon} - W|1_{U=1_{\tau \geq t_\varepsilon}}] + o_{\varepsilon \rightarrow 0}(1).$$

We write

$$\tau^{(2)} = \inf\{t > \tau, X_t < x_1\},$$

and remark that $\tau^{(2)} - \tau$, which is independant from τ , is the reaching time of $(x_1, -1)$ starting from $(x_1, 1)$: from Theorem 5.1, and since $U(x_2) < U(x_1)$, we see that $\mathbb{E}[\tau]$ is negligible with respect to $\mathbb{E}[\tau^{(2)} - \tau]$, that $\frac{\tau^{(2)} - \tau}{\mathbb{E}[\tau^{(2)} - \tau]}$ converges in law to an exponential variable and so that $\frac{\tau^{(2)} - \tau}{\mathbb{E}[\tau]}$ diverges in probability to infinity. In particular

$$\mathbb{P}(\tau^{(2)} \geq t_\varepsilon) \geq \mathbb{P}(\tau^{(2)} - \tau \geq t_\varepsilon) \xrightarrow{\varepsilon \rightarrow 0} 1.$$

Using again Lemma 5.6, we get

$$\begin{aligned} \mathbb{E}[|X_{t_\varepsilon} - W|] &= \mathbb{E}[|X_{t_\varepsilon} - W| \mathbf{1}_{U=1_{\tau \geq t_\varepsilon}, \tau^{(2)} > t_\varepsilon}}] + o_\varepsilon(1) \\ &= \mathbb{E}[|X_{t_\varepsilon} - x_0| \mathbf{1}_{t_\varepsilon < \tau}] + \mathbb{E}[|X_{t_\varepsilon} - x_2| \mathbf{1}_{\tau < t_\varepsilon < \tau^{(2)}}] + o_\varepsilon(1) \end{aligned}$$

Both expectations are treated in the same way ; let us focus on the first one. Let

$$t' = \sup\{s < t_\varepsilon, X_s = x_0\}, \quad t'' = \min(\inf\{s > t_\varepsilon, X_s = x_0\}, \tau)$$

and $I = t'' - t'$. Note that $|X_{t_\varepsilon} - x_0| < I$, and that, for $0 \leq a \leq b$, the following events are included :

$$\{a \leq I < b\} \subset A$$

where

$$A = \left\{ \exists s_1, s_2 \in [t_\varepsilon - b, t_\varepsilon + b], s_1 < s_2, X_{s_1} = x_0, |X_{s_2} - x_0| = \frac{a}{2} \right\}.$$

Since the process starts anew when it reaches x_0 ,

$$\begin{aligned} \mathbb{P}(A) &\leq \mathbb{P}_{X_0=x_0} \left(\exists s < 2b, |X_s - x_0| = \frac{a}{2} \right) \\ &\leq \mathbb{P}(\theta(a) \leq 2b), \end{aligned}$$

where

$$\begin{aligned} \theta(a) &= \inf\{s > 0, U(X_s) > U(x_0) + \delta(a)\} \\ \text{and } \delta(a) &= \begin{cases} \min\left(U\left(x_0 - \frac{a}{2}\right), U\left(x_0 + \frac{a}{2}\right)\right) - U(x_0) & \text{if } x_0 + \frac{a}{2} \leq x_1 \\ U\left(x_0 - \frac{a}{2}\right) - U(x_0) & \text{else.} \end{cases} \end{aligned}$$

In other words, $\delta(a)$ is the minimal energy barrier to overcome to be at distance at least $\frac{a}{2}$ from x_0 (when the process is still in the catchment area of x_0). If $a \leq I < b$, such a distance has been attained, and it has been so in a prescribed time window of length at most $2b$. As in the proof of Lemma 5.6 we use Theorem 5.10 to control the probability to reach a given energy level in a given time : for all $a \geq 0$ and $\varepsilon < \varepsilon_0$,

$$\begin{aligned} \mathbb{P}(\theta(a) \leq 2b) &\leq 1 - \exp\left(-\Gamma\left(e^{-\frac{\delta(a)}{\varepsilon}} + 2b\frac{e^{-\frac{\delta(a)}{\varepsilon}}}{\sqrt{\varepsilon}}\right)\right) \\ &\leq \Gamma\left(1 + 2b\varepsilon^{-\frac{1}{2}}\right) e^{-\frac{\delta(a)}{\varepsilon}}. \end{aligned}$$

Putting all things together, when $\varepsilon < 1$,

$$\begin{aligned}
& \mathbb{E} [|X_{t_\varepsilon} - x_0| \mathbf{1}_{t_\varepsilon < \tau}] \\
& \leq \mathbb{E} [I \mathbf{1}_{t_\varepsilon < \tau}] \\
& \leq \sum_{0 \leq k \leq \lceil \varepsilon^{-\frac{1}{2}} \rceil} \sqrt{\varepsilon}(k+1) \mathbb{P} \left(k\sqrt{\varepsilon} \leq I < (k+1)\sqrt{\varepsilon} \right) + \sum_{k \geq 1} (k+1) \mathbb{P} (k \leq I < (k+1)) \\
& \leq 4\Gamma\sqrt{\varepsilon} \sum_{0 \leq k \leq \lceil \varepsilon^{-\frac{1}{2}} \rceil} (k+1)^2 e^{-\frac{\delta(k\sqrt{\varepsilon}/2)}{\varepsilon}} + 4\Gamma \sum_{k \geq 1} (k+1)^2 \varepsilon^{-\frac{1}{2}} e^{-\frac{\delta(k)}{\varepsilon}}. \tag{5.6}
\end{aligned}$$

From the convexity of U , $\delta(k)$ grows faster than linearly, so that, if ε_0 is such that

$$\varepsilon \mapsto \varepsilon^{-\frac{1}{2}} e^{-\frac{\delta(1)}{\varepsilon}}$$

is non-increasing on $(0, \varepsilon_0)$, for all $\varepsilon < \varepsilon_0$,

$$\sum_{k \geq 1} (k+1)^2 \varepsilon^{-\frac{1}{2}} e^{-\frac{\delta(k)}{\varepsilon}} < \sum_{k \geq 1} (k+1)^2 \varepsilon_0^{-\frac{1}{2}} e^{-\frac{\delta(k)}{\varepsilon_0}} < \infty.$$

Thus by the dominated convergence Theorem, the second sum of (5.6) goes to 0 as ε goes to 0. As far as the first sum is concerned, note that, since x_0 is a non-degenerated minimum of U , there exist $\eta > 0$ such that $\delta(s) \geq \eta s^2$ for all $s \in [0, 1]$. This implies

$$\sum_{0 \leq k \leq \lceil \varepsilon^{-\frac{1}{2}} \rceil} (k+1)^2 e^{-\frac{\delta(k\sqrt{\varepsilon}/2)}{\varepsilon}} \leq \sum_{k \geq 0} (k+1)^2 e^{-\eta k^2} < \infty.$$

Reinjected in (5.6), this finishes to prove $\mathbb{E} [|X_{t_\varepsilon} - x_0| \mathbf{1}_{t_\varepsilon < \tau}]$ goes to 0; $\mathbb{E} [|X_{t_\varepsilon} - x_2| \mathbf{1}_{\tau < t_\varepsilon < \tau(2)}]$ is treated the same way, and at the end of the day $\mathbb{E} [|X_{t_\varepsilon} - W|]$ goes to 0. For ε small enough one can thus find a coupling (X, W) with marginal laws $\mathcal{L}(X_{t_\varepsilon})$ and $e^{-t}\delta_{x_0} + (1 - e^{-t})\delta_{x_2}$ such that $\mathbb{E} [|X - W|]$ is arbitrarily small, which concludes. \square

In particular, since $\tilde{\eta}_\varepsilon$, the first marginal of the invariant measure η_ε , converges to the Dirac mass on x_2 , and

$$\mathcal{W} \left(e^{-t}\delta_{x_0} + (1 - e^{-t})\delta_{x_2}, \delta_{x_2} \right) = e^{-t}|x_2 - x_0|,$$

we get

Corollary 5.8. *Assume $U(x_2) < U(x_1)$. Then*

$$\mathcal{W} (\mu_{t_\varepsilon}, \tilde{\eta}_\varepsilon) \xrightarrow{\varepsilon \rightarrow 0} e^{-t}|x_2 - x_0|.$$

5.3 Necessary and sufficient conditions for the cooling schedule

We are interested in the behaviour of the inhomogeneous PDMP with generator

$$L_t f(x, y) = yf'(x, y) + \frac{1}{\varepsilon_t} (yU'(x))_+ (f(x, -y) - f(x, y)).$$

where the cooling schedule $t \mapsto \varepsilon_t$ is supposed non-increasing and positive. The process can be explicitly constructed in the following way. Let $(E_k)_{k \geq 0}$ be a family of i.i.d. variable with exponential law. Set $(X_0, Y_0) = (x_0, y_0)$ and $T_0 = 0$. Suppose the process is already constructed up to a jump time T_i , $i \geq 0$, and is independant from $(E_k)_{k \geq i}$ up to T_i . Let

$$\begin{aligned} T_{i+1} &= \inf \left\{ t > T_i, \int_{T_i}^t \frac{(Y_{T_i} U'(X_{T_i} + Y_{T_i}(s - T_i)))_+ ds}{\varepsilon_s} \geq E_i \right\} \\ X_s &= X_{T_i} + Y_{T_i}(s - T_i) \quad \text{for } s \in [T_i, T_{i+1}] \\ Y_s &= Y_{T_i} \quad \text{for } s \in [T_i, T_{i+1}] \\ Y_{T_{i+1}} &= -Y_{T_i}. \end{aligned}$$

In a finite time interval $[0, T]$, ε_t is bounded by ε_T and thus there cannot be an infinite number of jumps in a finite time. In particular the sequence of $(T_i)_{i \geq 0}$ goes to infinity and the process is well defined at all times.

Our method to prove Theorem 5.3 is adapted to our settings from the work of Hajek ([53]) on simulated annealing on a discrete space. We consider a smooth Morse potential U on \mathbb{R} with a finite number of local extrema, and convex at infinity. If x is a minimum of depth E , denote by C_x the set of all points which are reachable from x at height strictly less than E . We call C_x the cusp of x (see the grey area of Figure 1.1 in Section 1.3).

More generally we call cusp an interval $C = (z_l, z_r)$ with $U(z_l) = U(z_r)$ and for all $x \in C$, $U(x) < U(z_l)$. The depth d of C is defined as

$$d = U(z_l) - \min\{U(x), x \in C\}.$$

We note $B = \{z \in C, U(z) = \min\{U(x), x \in C\}\}$ the bottom of the cusp. Note that a minimum x is always in the bottom of C_x , and that conversely if z is in the bottom of C_x then $C_x = C_z$. Obviously the depth of C_x equals the depth of x .

We want to bound the time the process spends in a cusp C , depending on the depth d of the latter. Nevertheless this is impossible to do so if we only assume the initial position is in C : we should put aside the cases where the process starts near the boundary of C , with a velocity directed toward the exit.

We denote by \mathcal{N} the set of local extrema of U in C , and

$$u = U(z_l) - \max\{U(z), z \in \mathcal{N}\}.$$

The maximum over \mathcal{N} is necessarily attained on a local maximum, except when there is no local maximum between z_l and z_r , in which case \mathcal{N} is a single point, which is a local minimum, and $u = d$. We define

$$x_l = \inf\{x \in C, U(z_l) - U(x) \geq u\} \quad x_r = \sup\{x \in C, U(z_r) - U(x) \geq u\}$$

and

$$A_C = (z_l, x_l) \times \{-1\} \cup (x_r, z_r) \times \{1\} \subset \mathbb{R} \times \{\pm 1\}.$$

The process can only leave C from A_C , and can only enter it through $\bar{A}_C = (C \times \{\pm 1\}) \setminus A_C$. We will note $A_C(t)$ (resp. $\bar{A}_C(t)$) the event $(X_t, Y_t) \in A_C$ (resp. \bar{A}_C).

Let $t_0 > 0$ and

$$\tau_C = \inf\{t > t_0, X_t \notin C\}.$$

In order to prove Theorem 5.3, we will establish the two following intermediate results :

Theorem 5.9. *Let C be a cusp of depth d . There exist $\varepsilon_0, c > 0$ such that for all time $t_0 > 0$, all end $z \in \{z_l, z_r\}$ and all cooling schedule $t \mapsto \varepsilon_t$, if $\bar{A}_C(t_0)$ holds, $\varepsilon_{t_0} \leq \varepsilon_0$ and*

$$\int_{t_0}^{\infty} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds = \infty,$$

then

$$\mathbb{E} \left[\int_{t_0}^{\tau_C} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds \right] \leq c \quad (5.7)$$

$$\mathbb{P}(X_{\tau_C} = z) \geq \frac{1}{c} \quad (5.8)$$

Theorem 5.10. *Let C be a cusp of depth d . There exist $\Gamma, \varepsilon_0 > 0$ such that for all $t_0 > 0$, such that for all times $t_0 > 0, r \geq t_0$ and all cooling schedule $t \mapsto \varepsilon_t$, if $\bar{A}_C(t_0)$ holds and $\varepsilon_{t_0} \leq \varepsilon_0$ then*

$$\mathbb{P}(\tau_C \geq r) \geq \exp \left(-\Gamma \left(e^{-\frac{d}{\varepsilon_{t_0}}} + \int_{t_0}^r \frac{e^{-\frac{d}{\varepsilon_u}}}{\sqrt{\varepsilon_u}} du \right) \right).$$

Note that contrary to Theorem 5.3, these intermediate results do not require the temperature to go to zero. In particular they hold for constant ε .

5.3.1 Proof of Theorem 5.9

This is proved by induction over the number of local minima the cusps C contains. Thus we start with a cusp that contains only one minimum, and so no maximum.

A simple cusp

We fix throughout this section a cusp C of depth d with only one local minimum x_0 of U . Note that in this case $u = d$, and $x_l = x_r = x_0$. We will prove Theorem 5.9 in this situation, which is pretty similar to the settings of Section 5.2. Let $\tau_0 = t_0$ and

$$\tau_{i+1} = \min(\inf\{t > \tau_i, X_t = x_0\}, \tau_C).$$

We first prove :

Lemma 5.11. *There exists c_1 and $\varepsilon_0 > 0$ such that for all $\varepsilon_{t_0} \leq \varepsilon_0$, for all $i \geq 1$,*

$$\mathbb{E} \left[\tau_{i+1} - \tau_i \mid \bar{A}_C(\tau_i), \tau_i \right] \leq c_1 \sqrt{\varepsilon_{\tau_i}}.$$

Démonstration. At time $\tau_i, i \geq 1$, given $\bar{A}_C(\tau_i)$, the position X_{τ_i} is x_0 and the velocity Y_{τ_i} is either 1 or -1. Writing

$$\mathbb{E} \left[\tau_{i+1} - \tau_i \mid \bar{A}_C(\tau_i), \tau_i \right] = \mathbb{E} \left[\mathbb{E} \left[\tau_{i+1} - \tau_i \mid \bar{A}_C(\tau_i), \tau_i, Y_{\tau_i} \right] \mid \bar{A}_C(\tau_i), \tau_i \right],$$

both cases $Y_{\tau_i} = -1$ or 1 are treated the same way. For instance

$$\begin{aligned} \mathbb{E} \left[\tau_{i+1} - \tau_i \mid \tau_i, (X_{\tau_i}, Y_{\tau_i}) = (x_0, -1) \right] &= \mathbb{P}(\tau_{i+1} = \tau_C \mid \tau_i, (X_{\tau_i}, Y_{\tau_i}) = (x_0, -1)) |x_0 - z_l| \\ &\quad + \mathbb{E} \left[(\tau_{i+1} - \tau_i) \mathbf{1}_{\tau_{i+1} \neq \tau_C} \mid \tau_i, (X_{\tau_i}, Y_{\tau_i}) = (x_0, -1) \right]. \end{aligned}$$

Since the cooling schedule is non-increasing, the probability to reach the boundary and the expectation of the time of jump are bounded by the corresponding quantities at constant temperature ε_{τ_i} . Indeed, if $E \sim \mathcal{E}(1)$, conditionally to τ_i and to $(X_{\tau_i}, Y_{\tau_i}) = (x_0, -1)$,

$$\begin{aligned} \mathbb{P}(\tau_{i+1} = \tau_C) &= \mathbb{P}\left(E \geq \int_0^{x_0 - z_l} \frac{U'(x_0 - s)}{\varepsilon_{\tau_i + s}} ds\right) \\ &\leq \mathbb{P}\left(E \geq \frac{1}{\varepsilon_{\tau_i}} \int_0^{x_0 - z_l} U'(x_0 - s) ds\right) \\ &= e^{-\frac{d}{\varepsilon_{\tau_i}}} \end{aligned}$$

and, if T_i is the next jump time starting from $(\tilde{X}_{\tau_i}, \tilde{Y}_{\tau_i}) = (x_0, -1)$ for the process at fixed temperature ε_{τ_i} ,

$$\begin{aligned} \mathbb{E}\left[(\tau_{i+1} - \tau_i) 1_{\tau_{i+1} \neq \tau_C}\right] &= 2 \int_0^{x_0 - z_l} \mathbb{P}\left(\frac{\tau_{i+1} - \tau_i}{2} \geq u\right) du \\ &= 2 \int_0^{x_0 - z_l} \mathbb{P}\left(E \geq \int_0^u \frac{U'(x_0 - s)}{\varepsilon_{\tau_i + s}} ds\right) du \\ &\leq 2 \int_0^{x_0 - z_l} \mathbb{P}\left(E \geq \frac{1}{\varepsilon_{\tau_i}} \int_0^u U'(x_0 - s) ds\right) du \\ &= 2\mathbb{E}[T_i 1_{T_i < x_0 - z_l}] \\ &= 2 \int_0^{x_0 - z_l} \frac{sU'(x_0 - s)}{\varepsilon_{\tau_i}} e^{-\frac{U(x_0 - s) - U(x_0)}{\varepsilon_{\tau_i}}} ds. \end{aligned}$$

Thanks to Lemma 5.4 there exist ε_0 and c_1 such that for all $\varepsilon \leq \varepsilon_0$,

$$e^{-\frac{d}{\varepsilon}} |x_0 - z_l| + 2 \int_0^{x_0 - z_l} \frac{sU'(x_0 - s)}{\varepsilon} e^{-\frac{U(x_0 - s) - U(x_0)}{\varepsilon}} ds \leq c_1 \sqrt{\varepsilon},$$

which concludes the proof since $\varepsilon_{\tau_i} \leq \varepsilon_{t_0} \leq \varepsilon_0$. \square

Proof of Part (5.7) of Theorem 5.9 for a simple cusp. Note that, choosing ε_0 small enough, for all cooling schedule $\varepsilon_t \leq \varepsilon_0$, the map

$$s \mapsto (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}}$$

is non-increasing, so that

$$\begin{aligned} \mathbb{E}\left[\int_{\tau_i}^{\tau_{i+1}} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds \middle| \tau_i, \bar{A}_C(\tau_i)\right] &\leq (\varepsilon_{\tau_i})^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_{\tau_i}}} \mathbb{E}[\tau_{i+1} - \tau_i \mid \tau_i, \bar{A}_C(\tau_i)] \\ &\leq c_1 e^{-\frac{d}{\varepsilon_{\tau_i}}}. \end{aligned}$$

We define

$$\Phi(i, j) = \mathbb{E}\left[\int_{\tau_i}^{\tau_j} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} 1_{s < \tau_C} ds \middle| \tau_{i-1}, \bar{A}_C(\tau_{i-1})\right].$$

Obviously $\Phi(j, j) = 0$; suppose $\Phi(i+1, j) \leq c_1$ has already been proved for some $i < j$. Under the event $\bar{A}_C(\tau_{i-1})$, there are two possibilities : either the process escapes the cusp C between the times τ_{i-1} and τ_i , in which case $\tau_i = \tau_C = \tau_j$ and the integral appearing in the definition

of $\Phi(i, j)$ vanishes; or the attempt to exit C between τ_{i-1} and τ_i fails and the process falls down to $X_{\tau_i} = x_0$. To sum up,

$$\begin{aligned}\Phi(i, j) &= \mathbb{P}(\tau_i < \tau_C \mid \tau_{i-1}, \bar{A}_C(\tau_{i-1})) \mathbb{E} \left[\int_{\tau_i}^{\tau_j} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} 1_{s < \tau_C} ds \mid \tau_{i-1}, \bar{A}_C(\tau_i) \right] \\ &\leq \left(1 - e^{-\frac{d}{\varepsilon_{\tau_{i-1}}}} \right) \mathbb{E} \left[\mathbb{E} \left[\int_{\tau_i}^{\tau_{i+1}} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} 1_{s < \tau_C} ds \mid \tau_i, \bar{A}_C(\tau_i) \right] + \Phi(i+1, j) \mid \tau_{i-1} \right] \\ &\leq \left(1 - e^{-\frac{d}{\varepsilon_{\tau_{i-1}}}} \right) \mathbb{E} \left[c_1 e^{-\frac{d}{\varepsilon_{\tau_i}}} + c_1 \mid \tau_{i-1} \right] \\ &\leq \left(1 - e^{-\frac{d}{\varepsilon_{\tau_{i-1}}}} \right) \left(c_1 e^{-\frac{d}{\varepsilon_{\tau_{i-1}}}} + c_1 \right) \leq c_1.\end{aligned}$$

This proves by induction that $\Phi(1, j) \leq c_1$ for all $j \geq 1$. As was already pointed out there can only be a finite number of jumps in a finite time, so that the sequence $(\tau_i)_{i \geq 0}$ almost surely converges to τ_C (which, at this point, may be infinite). The monotone convergence theorem yields

$$\mathbb{E} \left[\int_{\tau_1}^{\tau_C} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds \mid \bar{A}_C(t_0) \right] \leq c_1.$$

On the other hand, as soon as $\varepsilon_t \leq \varepsilon_0$,

$$\mathbb{E} \left[\int_{t_0}^{\tau_1} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds \mid \bar{A}_C(t_0) \right] \leq (\varepsilon_0)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_0}} \max(|x_0 - z_l|, |z_r - x_0|).$$

Bringing all the pieces together, there exists a constant $c > 0$ such that

$$\mathbb{E} \left[\int_{t_0}^{\tau_C} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds \mid \bar{A}_C(t_0) \right] \leq c.$$

□

Note that if the cooling schedule is such that $\int_{t_0}^{\infty} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds = \infty$, this first result implies that τ_C is almost surely finite, in which case (X_{τ_C}, Y_{τ_C}) is well defined, and the second half of Theorem 5.9 makes sense.

Proof of Part (5.8) of Theorem 5.9 for a simple cusp. We note p_i^l (resp. p_i^r) the probability to exit C in one shot, meaning before reaching again x_0 , starting at $(X_{\tau_i}, Y_{\tau_i}) = (x_0, -1)$ (resp. $(x_0, 1)$), with $X_{\tau_C} = z_l$ (resp. z_r). Namely, if $E \sim \mathcal{E}(1)$,

$$\begin{aligned}p_i^l &= \mathbb{P} \left(\int_0^{x_0 - z_l} \frac{1}{\varepsilon_{\tau_i + s}} U'(x_0 - s) ds \leq E \right) \\ p_i^r &= \mathbb{P} \left(\int_0^{z_r - x_0} \frac{1}{\varepsilon_{\tau_i + s}} U'(x_0 + s) ds \leq E \right)\end{aligned}$$

From the non-increasivity of the cooling schedule,

$$e^{-\frac{d}{\varepsilon_{\tau_{i+1}}}} \leq p_i^l, p_i^r \leq e^{-\frac{d}{\varepsilon_{\tau_i}}}.$$

In particular $p_i^l \geq p_{i+1}^r$, which means between two consecutive attempts to exit the cusp, the second one is always less likely to succeed than the first one, and thus if $(X_{\tau_i}, Y_{\tau_i}) = (x_0, -1)$ then the probability that the exit point will be z_l is greater than $\frac{1}{2}$. On the other hand,

$$\begin{aligned}\mathbb{P}(X_{\tau_C} = z_l \mid (X_{\tau_1}, Y_{\tau_1}) = (x_0, 1)) &= (1 - p_1^r) \mathbb{P}(X_{\tau_C} = z_l \mid (X_{\tau_2}, Y_{\tau_2}) = (x_0, -1)) \\ &\geq \frac{1}{2} \left(1 - e^{-\frac{d}{\varepsilon_{t_0}}} \right).\end{aligned}$$

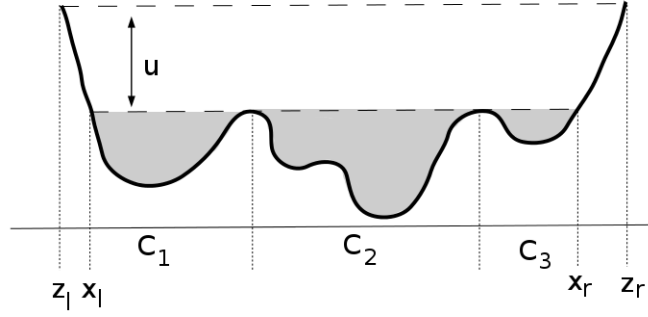


FIGURE 5.1 – The cusp C is divided in several smaller cusps.

Under the event $\bar{A}_C(t_0)$, necessarily $\tau_1 < \tau_C$, namely $X_{\tau_1} = x_0$ and Y_{τ_1} is either 1 or -1; both cases have been treated, and as soon as $\varepsilon_{t_0} \leq \varepsilon_0$,

$$\mathbb{P}\left(X_{\tau_C} = z_l \mid \bar{A}_C(t_0)\right) \geq \frac{1}{2} \left(1 - e^{-\frac{d}{\varepsilon_0}}\right).$$

The case of z_r is symmetric. □

Induction

In this section we consider a general cusp of depth d and we prove Theorem 5.9 under the induction assumption that it holds for any cusp with strictly fewer local maxima than C . There is a finite number of maxima z in C for which $u = U(z_l) - U(z)$, and each connected component of $(x_l, x_r) \setminus \{z, u = U(z_l) - U(z)\}$ is a cusp of depth at most $g = d - u$ (cf Fig. 5.1). We call $(C_i)_{i=1..n}$ these cusps of depth $g = d - u$. We can consider ε_0 and M such that for all $\varepsilon_t \leq \varepsilon_0$, for all $i = 1..n$, for all z_i which is an end of C_i and for all t_0 ,

$$\begin{aligned} \mathbb{P}\left(X_{\tau_{C_i}} = x_i \mid \bar{A}_{C_i}(t_0)\right) &\geq \frac{1}{M} \\ \text{and } \mathbb{E}\left[\int_{t_0}^{\tau_{C_i}} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{g}{\varepsilon_s}} ds \mid \bar{A}_{C_i}(t_0)\right] &\leq M \\ \Rightarrow \mathbb{E}\left[\int_{t_0}^{\tau_{C_i}} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds \mid \bar{A}_{C_i}(t_0)\right] &\leq M e^{-\frac{u}{\varepsilon_{t_0}}}. \end{aligned}$$

Proof of Theorem 5.9, part (5.7). We note $J_0 = t_0$,

$$\begin{aligned} K_i &= \min(\inf\{t > J_{i-1}, A_C(t)\}, \tau_C) \\ J_i &= \min(\inf\{t > K_i, \bar{A}_C(t)\}, \tau_C). \end{aligned}$$

At time K_i , the process is ready to escape : it is located near the boundary of C , and its velocity is directed toward the exit. At time J_i , if it failed to leave the cusp, it goes back into the C_j 's (see Fig. 5.1)

and, as previously,

$$\begin{aligned}
\Phi(i, j) &:= \mathbb{E} \left[\int_{K_i}^{K_j} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds \middle| K_{i-1}, \tau_C > K_{i-1} \right] \\
&= \mathbb{P}(\tau_C > K_i | K_{i-1}, \tau_C > K_{i-1}) \mathbb{E} \left[\int_{K_i}^{K_j} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds \middle| K_{i-1}, \tau_C > K_i \right] \\
&\leq \left(1 - e^{-\frac{u}{\varepsilon K_{i-1}}} \right) \mathbb{E} \left[\int_{K_i}^{J_i} + \int_{J_i}^{K_{i+1}} + \int_{K_{i+1}}^{K_j} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds \middle| K_{i-1}, \tau_C > K_i \right].
\end{aligned}$$

We treat the three parts of the integral one after another. From Lemma 5.11, there exist a constant c_1 (depending on C but not on $(\varepsilon_t)_{t \geq t_0}$) such that

$$\mathbb{E} \left[\int_{K_i}^{J_i} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds \middle| K_i, \tau_C > K_i \right] \leq c_1 e^{-\frac{u}{\varepsilon K_i}}.$$

As far as the time interval $[J_i, K_{i+1}]$ is concerned, here is what happens : if $J_i = \tau_C$, the process has escaped from C , hence $K_{i+1} = \tau_C$ and the integral vanishes. Else the process goes back down to a cusp C_i , and then for a while pass from a C_j to another, until it reaches A_C again. The contribution of the time passed in each C_i to the integral can be controled thanks to the induction assumption. It remains to make sure the ping pong between the C_j 's lasts not too long. When the process exits C_i , the probability it does so through a given end is bounded below thanks to the induction assumption. Thus the expectation of the number N of jumps from a C_i to another C_j before the process reaches A_C is bounded by a constant D which does not depend on the cooling schedule. Hence, denoting by $\tau_{C_i}^*$ an entry time of the process in C_i ,

$$\begin{aligned}
&\mathbb{E} \left[\int_{J_i}^{K_{i+1}} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds \middle| K_i, \tau_C > K_i \right] \\
&\leq \mathbb{E} \left[N \max_i \mathbb{E} \left[\int_{\tau_{C_i}^*}^{\tau_{C_i}} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds \middle| \tau_{C_i}^* > K_i, \bar{A}_{C_i}(\tau_{C_i}^*) \right] \middle| K_i, \tau_C > J_i \right] \\
&\leq DM e^{-\frac{u}{\varepsilon K_i}}.
\end{aligned}$$

Since $\Phi(j, j) = 0$, as an induction assumption we can suppose

$$\Phi(i+1, j) \leq DM + c_1,$$

and thus we get

$$\begin{aligned}
\Phi(i, j) &\leq \left(1 - e^{-\frac{u}{\varepsilon K_{i-1}}} \right) \mathbb{E} \left[e^{-\frac{u}{\varepsilon K_i}} (DM + c_1) + \Phi(i+1, j) \middle| K_{i-1}, \tau_C > K_i \right] \\
&\leq \left(1 - e^{-\frac{u}{\varepsilon K_{i-1}}} \right) \left(e^{-\frac{u}{\varepsilon K_{i-1}}} + 1 \right) (DM + c_1) \\
&\leq DM + c_1.
\end{aligned}$$

The monotone convergence Theorem yields

$$\mathbb{E} \left[\int_{K_1}^{\tau_C} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds \middle| \bar{A}_C(t_0) \right] \leq DM + c_1.$$

In the other hand in the event $\bar{A}_C(t_0)$, either the process start in a C_i , or it reaches a C_i in a time bounded by $\max(|x_l - z_l|, |z_r - x_r|)$. In both cases with the previous argument we used for $[J_i, K_{i+1}]$,

$$\mathbb{E} \left[\int_{t_0}^{K_1} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds \middle| \bar{A}_C(t_0) \right] \leq DM + (\varepsilon_{t_0})^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_{t_0}}} \max(|x_l - z_l|, |z_r - x_r|)$$

and at the end of the day part (5.7) of Theorem 5.9 is proved with

$$c = 2DM + c_1 + (\varepsilon_{t_0})^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_{t_0}}} \max(|x_l - z_l|, |z_r - x_r|).$$

□

The same remark as in Section 5.3.1 holds : when $\int_{t_0}^{\infty} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds = \infty$, the first part of Theorem 5.9 implies that τ_C is almost surely finite, and X_{τ_C} well-defined.

Proof of Theorem 5.9, part (5.8). The situation is very similar to the simple cusp one. Take $M'_0 = t_0$ and

$$\begin{aligned} M_i &= \min(\inf\{t > M'_i, A_C(t)\}, \tau_C) \\ M'_{i+1} &= \min(\inf\{t > M_i, \bar{A}_C(t)\}, \tau_C). \end{aligned}$$

Given the sequence $(M_i)_{i \geq 1}$, $(X_{M_i})_{i \geq 1}$ is an inhomogeneous Markov chain in $\{z_l, x_l, x_r, z_r\}$. Let

$$\begin{aligned} p_i^l &= \mathbb{P} \left(\int_0^{x_l - z_l} \frac{1}{\varepsilon_{M_i + s}} U'(x_0 - s) ds \leq E \right) \\ p_i^r &= \mathbb{P} \left(\int_0^{z_r - x_r} \frac{1}{\varepsilon_{M_i + s}} U'(x_0 + s) ds \leq E \right) \end{aligned}$$

The induction assumption on the C_j 's implies that the transitions of this chain from x_r to x_r and vice-versa are bounded below by a constant $h > 0$ which does not depend on the cooling schedule. Hence $(X_{M_i})_{i \geq 1}$ is more likely to reach z_l before z_r than the chain $(\tilde{X}_{M_i})_{i \geq 1}$ with transition

$$\begin{aligned} \mathbb{P}(x_l \rightarrow x_r) &= 1 - p_i^l & \mathbb{P}(x_r \rightarrow x_l) &= h \\ \mathbb{P}(x_l \rightarrow z_l) &= p_i^l & \mathbb{P}(x_r \rightarrow z_r) &= p_i^r \\ \mathbb{P}(x_l \rightarrow x_l) &= 0 & \mathbb{P}(x_r \rightarrow x_r) &= 1 - h - p_i^r. \end{aligned}$$

Similarly to the situation in Section 5.3.1, $p_i^l \geq p_j^r$ if $j > i$, and $p_1^r \leq e^{-\frac{u}{\varepsilon_0}}$. The probability that \tilde{X} , starting from $\tilde{X}_{M_{i+1}} = x_r$, hits z_r before x_l is

$$\sum_{j \geq 1} p_{i+j}^r \prod_{1 \leq k < j} (1 - h - p_k^r) \leq \sum_{j \geq 1} p_{i+1}^r (1 - h)^{j-1} \leq \frac{p_i^l}{1 - h}.$$

In particular this is less than $\frac{e^{-\frac{u}{\varepsilon_0}}}{1 - h}$, so that the probability that there exist $k_0 \geq 0$ such that $\tilde{X}_{M_{k_0}} = x_l$ is bounded below. Hence to bound the probability to reach z_l before z_r we can assume the initial point is x_l . Let

$$k_{i+1} = \inf\{n > k_i, \tilde{X}_{M_n} \neq x_r\}.$$

Given the times (M_{k_i}) , the chain \tilde{X} is more likely to reach z_l before z_r than the chain X' with transition

$$\mathbb{P}(x_l \rightarrow z_l) = p_{k_i}^l \quad \mathbb{P}(x_l \rightarrow z_r) = \frac{p_{k_i}^l}{1-h} \quad \mathbb{P}(x_l \rightarrow x_l) = 1 - p_{k_i}^l - \frac{p_{k_i}^l}{1-h}.$$

Finally X' goes to z_l rather than z_r with a probability $\frac{1}{1+\frac{1}{1-h}}$, which does not depend on the cooling schedule, which concludes the proof. \square

5.3.2 Proof of Theorem 5.10

We start by some preliminary lemmas. Let C be a cusp of depth d , recall that its bottom is $B = \{x \in C, U(z) \geq U(x) \forall z \in C\}$ and let $x_b = \min B$. We will give an upper bound of the probability that the process, starting at time t_0 at point $(x_b, -1)$, reaches z_l before $(x_b, 1)$. Obviously, if there were no local maximum between x_b and z_l , this probability would be bounded by $e^{-\frac{d}{\varepsilon t_0}}$. To prove a similar bound in more general cases, we decompose C in the following fashion : let

$$J = \{x \in [z_l, x_b], \text{ s.t. } \forall z \in (x, x_b], U(z) < U(x)\}.$$

On the set J , U is non-increasing, and the connected component of the closure of $[z_l, x_b] \setminus J$ are cusps whose right end are local maxima. We call $u_1 > u_2 > \dots > u_q$ this local maxima, C_1, \dots, C_{q-1} the corresponding cusps, $u_0 = x_b$, $C_0 = \{x_b\}$, $u_{q+1} = z_l$ and $C_{q+1} = \{z_l\}$. The point of taking the connected component of *the closure* of $[z_l, x_b] \setminus J$ was to ensure $U(u_i) > U(u_j)$ if $i > j$. We say $U(u_i)$ is the energy level of C_i , denoted by E_i , and we note

$$\delta_i = E_{i+1} - E_i.$$

Note that $\sum_{i=0}^q \delta_i = E_{q+1} - E_0 = d$. The situation is represented in Figure 5.2. Let $t_0 > 0$, $1 \leq i < q$ and suppose $\bar{A}_{C_i}(t_0)$ holds. Let

$$s = \inf\{t > t_0 \text{ s.t. } \exists j \neq i, X_t \in C_j\}.$$

Lemma 5.12. *There exists $c_2 > 0$ which depends on C but not on the cooling schedule so that*

$$\mathbb{P}(U(X_s) = E_{i+1}) \leq c_2 e^{-\frac{\delta_i}{\varepsilon t_0}} \quad (5.9)$$

Démonstration. Since there is a finite number of cusps C_i , it is enough to show such a constant exist for any one of them. Starting from A_{C_i} , if the process exits C_i to the right, it will deterministically falls down to C_{i-1} , which means $U(X_s) = E_{i-1}$. If the process exits to the left, it reaches C_{i+1} in one shot with probability less than $e^{-\frac{\delta_i}{\varepsilon t_0}}$, else goes back down to \bar{A}_{C_i} . Since the probability the process escapes to the left, starting from \bar{A}_{C_i} , is bounded by a constant $h > 0$ thanks to Theorem 5.9,

$$\mathbb{P}(U(X_s) = E_{i+1}) \leq \sum_{k \geq 0} (1-h)^k e^{-\frac{\delta_i}{\varepsilon t_0}} = \frac{e^{-\frac{\delta_i}{\varepsilon t_0}}}{1-h}.$$

\square

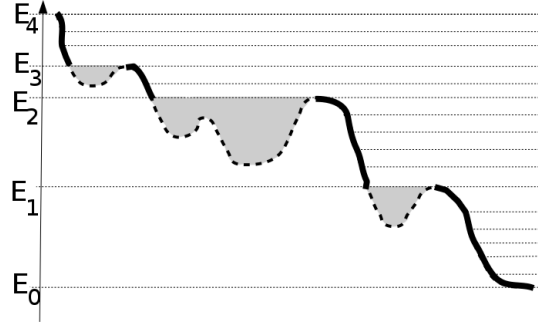


FIGURE 5.2 – If the light comes from the right, J , in bold line, is the sunny part, and the C_i 's are in the shadow.

Remark : The constant c_2 is defined from h , which depends only on the C_i 's. Thus If $(D_l)_{l \geq 0}$ is a family of imbricated cusps such that for all $k \geq 0$ all the minima of D_k belongs to $\bigcap_{l \geq 0} D_l$, the constant c_2 may be defined uniformly on k so that (5.9) holds for all D_k 's. Since q (the number of small cusps) is also the same for all D_k 's, this remark will extend to the next result.

This lemma implies the process, starting from \bar{A}_{C_i} , is less likely to hit $(z_i, -1)$ before $(x_0, 1)$ than the birth-death process $(W_n)_{n \geq 0}$ on $[0, q + 1]$ with transition probabilities

$$\mathbb{P}(j \rightarrow j + 1) = c_2 e^{-\frac{\delta_j}{\varepsilon_{t_0}}} \quad \mathbb{P}(j \rightarrow j - 1) = \left(1 - c_2 e^{-\frac{\delta_j}{\varepsilon_{t_0}}}\right) \quad j = 1, \dots, q$$

is to hit $q + 1$ before 0, starting at i .

Lemma 5.13. For ε_{t_0} small enough,

$$\mathbb{P}(W \text{ hits } q + 1 \text{ before } 0 \mid W_0 = 1) \leq (2c_2)^q e^{-\frac{E_{q+1} - E_1}{\varepsilon_{t_0}}}$$

As a consequence,

$$\mathbb{P}(X \text{ hits } z_i \text{ before } x_b \mid \bar{A}_{C_1}(t_0)) \leq (2c_2)^q e^{-\frac{E_{q+1} - E_1}{\varepsilon_{t_0}}}$$

and

$$\mathbb{P}((X, Y) \text{ hits } (z_i, -1) \text{ before } (x_b, 1) \mid (X_{t_0}, Y_{t_0}) = (x_b, 1)) \leq (2c_2)^q e^{-\frac{d}{\varepsilon_{t_0}}}$$

Démonstration. More generally, let

$$r_i = \mathbb{P}(W \text{ hits } q + 1 \text{ before } i - 1 \mid W_0 = i).$$

If W hits q before $i - 1$, necessarily $W_1 = i + 1$ (which occurs with probability $c_2 e^{-\frac{\delta_i}{\varepsilon_{t_0}}}$) and either the chain stays above i (with probability r_{i+1}), either it goes back at some point to i and then we are back to the initial problem. Thus

$$r_i = c_2 e^{-\frac{\delta_i}{\varepsilon_{t_0}}} (r_{i+1} + (1 - r_{i+1})r_i).$$

Suppose ε_{t_0} is small enough to have $c_2 e^{-\frac{\delta_i}{\varepsilon_{t_0}}} \leq \frac{1}{2}$ for all $i = 1, \dots, q$, so that

$$r_i \leq 2c_2 e^{-\frac{\delta_i}{\varepsilon_{t_0}}} r_{i+1}$$

which concludes since $r_{q+1} = 1$. \square

We only considered here the left part of C , but the same goes for the right one. If $x'_b = \max B$ then there exist c'_2 and q' such that

$$\mathbb{P}((X, Y) \text{ hits } (z_r, 1) \text{ before } (x'_b, -1) \mid (X_{t_0}, Y_{t_0}) = (x'_b, 1)) \leq (2c'_2)^{q'} e^{-\frac{d}{\varepsilon_{t_0}}},$$

and we write $c_3 = \max((2c_2)^q, (2c'_2)^{q'})$.

Proof of Theorem 5.10. Let $a > 0$, $S_0 = t_0$, $K_0 = 0$ and

$$\begin{aligned} S_{i+1} &= \min(\inf\{t > S_i, X_t \in \{x_b, x'_b\}\}, \tau_C), \\ K_{i+1} &= \inf\{n > K_i, S_n \geq S_{K_i} + a\sqrt{\varepsilon_{K_n}}\}. \end{aligned}$$

A straightforward adaptation of Lemma 5.11 yields

$$\mathbb{E}[S_{i+1} - S_i \mid S_{i+1}] \geq c_4 \sqrt{\varepsilon_{i+1}}$$

for some $c_4 \geq 0$ as long as $\varepsilon_{t_0} \leq \varepsilon_0$, and thus there exist $m > 0$ which does not depend on the cooling schedule such that

$$\mathbb{E}[K_{i+1} - K_i] \leq m.$$

Hence

$$\begin{aligned} \mathbb{P}(S_{K_{i+1}} = \tau_C \mid S_{K_i} < \tau_C, S_{K_i}) &\leq \mathbb{E}\left[(K_{i+1} - K_i)c_3 e^{-\frac{d}{\varepsilon_{S_{K_i}}}} \mid S_{K_i} < \tau_C, S_{K_i}\right] \\ &\leq mc_3 e^{-\frac{d}{\varepsilon_{S_{K_i}}}} \end{aligned}$$

and so, if $mc_3 e^{-\frac{d}{\varepsilon_0}}$ is less than the positive solution of $1 - z = e^{-2z}$,

$$\begin{aligned} \mathbb{P}(S_{K_i} < \tau_C) &\geq \mathbb{E}\left[\prod_{j=0}^i \left(1 - mc_3 e^{-\frac{d}{\varepsilon_{S_{K_j}}}}\right)\right] \\ &\geq \mathbb{E}\left[\exp\left(-2mc_3 \sum_{j=0}^i e^{-\frac{d}{\varepsilon_{S_{K_j}}}}\right)\right] \\ &= \mathbb{E}\left[\exp\left(-2mc_3 \left(e^{-\frac{d}{\varepsilon_{t_0}}} + \frac{1}{a} \sum_{j=0}^i \left(\frac{e^{-\frac{d}{\varepsilon_{S_{K_j}}}}}{\sqrt{\varepsilon_{S_{K_j}}}}\right) a\varepsilon_{S_{K_j}}\right)\right)\right] \\ &\geq \mathbb{E}\left[\exp\left(-2mc_3 \left(e^{-\frac{d}{\varepsilon_{t_0}}} + \frac{1}{a} \int_{t_0}^{S_{K_i}} \frac{e^{-\frac{d}{\varepsilon_u}}}{\sqrt{\varepsilon_u}} du\right)\right)\right]. \end{aligned}$$

Finally for any $r > 0$,

$$\begin{aligned} \mathbb{P}(r < \tau_C) &= \mathbb{P}(S_{K_i} < \tau_C \forall S_{K_i} \leq r) \\ &\geq \mathbb{E}\left[\exp\left(-2mc_3 \left(e^{-\frac{d}{\varepsilon_{t_0}}} + \frac{1}{a} \int_{t_0}^r \frac{e^{-\frac{d}{\varepsilon_u}}}{\sqrt{\varepsilon_u}} du\right)\right)\right]. \end{aligned}$$

\square

Remark : Here the constant c_4 only depends on $U''(x_b)$ and $U''(x'_b)$, and m only depends on c_4 . Furthermore, as has already been noticed, c_3 only depends on the internal sub-cusps of C . Thus if $(D_l)_{l \geq 0}$ is a family of cusp so that for all $k \geq 0$ all the minima of D_k belongs to $\bigcap_{l \geq 0} D_l$, the constant Γ in Theorem 5.10 may be chosen uniformly over k . And if U is a potential with a finite number of local minima, going to $+\infty$ at $\pm\infty$, Γ may be chosen uniformly over all cusps of U .

5.3.3 Proof of the NS condition

Now that Theorems 5.9 and 5.10 are established, we recall (and slightly adapt) the arguments from [53] to prove Theorem 5.3.

Let $E \geq 0$ and

$$\begin{aligned} W_E &= \{x \in \mathbb{R}, x \text{ is a local minimum of depth strictly larger than } E\} \\ R_E &= \{x \in \mathbb{R}, x \text{ is reachable from } y \text{ at height } V(y) + E \text{ for some } y \in W_E\} \\ J &= \{x \in \mathbb{R}, W_E \text{ is reachable from } x \text{ at height } V(x)\}. \end{aligned}$$

If $x \notin J$, then the set of points which are reachable from x at height $V(x)$ is a cusp of depth at most E (else its bottom would be in W_E and reachable from x at height $V(x)$) which does not intersect J (if a point y were in the intersection, we could reach y from x at height $V(x)$ and then reach a point in W_E from y at height $V(y) \leq V(x)$). Thus the connected component of $\mathbb{R} \setminus J$ which contains x is itself a cusp of depth at most E . Moreover one of its end is a local maximum (else it could be thickened without intersecting J) which means there are finitely many connected components of $\mathbb{R} \setminus J$; we call them C_1, \dots, C_n .

Lemma 5.14. *Let $E > 0$. If*

$$\int_0^\infty (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{E}{\varepsilon_s}} ds = \infty$$

then

$$\lim_{t \rightarrow \infty} \mathbb{P}(X_t \in R_E) = 1.$$

Démonstration. Note that there are a finite number of local minima in W_E , so that there exists $\gamma_0 > 0$ such that all those are of depth larger than $E + \gamma_0$. We thicken R_E , for $\gamma < \gamma_0$, as

$$R_{E,\gamma} = \{x \in \mathbb{R}, x \text{ is reachable from } y \text{ at height } V(y) + E + \gamma \text{ for some } y \in W_E\}.$$

Since $R_E = \bigcap_{\gamma > 0} R_{E,\gamma}$, we only need to prove the result for an arbitrary $\gamma < \gamma_0$. Let $t_0 \geq 0$, $A_0 = t_0$ and

$$\begin{aligned} B_i &= \inf\{t > A_i, X_t \notin J\}, \\ A_{i+1} &= \min\{t > B_i, X_t \in J\}, \\ \alpha &= \inf\{k \geq 0, X_{A_k} \in W_E\}, \\ \beta &= \inf\{k > \alpha, X_{B_k} \notin R_{E,\gamma}\}. \end{aligned}$$

Since there is only a finite number of C_j , we can consider ε_0 and $c > 0$ in Theorem 5.9 such that if $\varepsilon_{t_0} \leq \varepsilon_0$, for all $i \geq 0$,

$$\mathbb{E} \left[\int_{B_i}^{A_{i+1}} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{E}{\varepsilon_s}} ds \middle| B_i \right] \leq c.$$

(Note that this would be false for $E = 0$).

On the other hand, note that on each connected component of $J \setminus W_E$, U is monotonous. There are two such components which are infinite, and where the potential is convex, so that the expected time the process stays there is bounded by a constant which only depends on ε_0 . The time the process stays in a compact connected component of $J \setminus W_E$ is bounded by twice the length of the component, and there is a finite number of such components. Thus there exist c' such that if $\varepsilon_{t_0} \leq \varepsilon_0$, for all $i \geq 0$,

$$\mathbb{E} \left[\int_{A_i}^{B_i} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{E}{\varepsilon_s}} ds \middle| A_i \right] \leq c'.$$

Moreover as a consequence of part (5.8) of Theorem 5.9, $\mathbb{E}[\alpha]$ is bounded by a constant which does not depend on the cooling schedule, and finally there exists $\tilde{c} > 0$ such that

$$\mathbb{E} \left[\int_{t_0}^{A_\alpha} (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{E}{\varepsilon_s}} ds \right] \leq \tilde{c},$$

so that

$$\mathbb{P}(A_\alpha \geq r) \leq \frac{\tilde{c}}{\int_{t_0}^r (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{E}{\varepsilon_s}} ds}.$$

At time A_α , the process attains the bottom of a cusp C (and therefore is in \bar{A}_C) of depth $E + \gamma$ which is included in $R_{E,\gamma}$. Thanks to Theorem 5.10,

$$\begin{aligned} \mathbb{P}(A_\beta \geq r) &\geq \exp \left(-\Gamma \left(e^{-\frac{E+\gamma}{\varepsilon_{t_0}}} + \int_{t_0}^r \frac{e^{-\frac{E+\gamma}{\varepsilon_u}}}{\sqrt{\varepsilon_u}} du \right) \right) \\ &\geq \exp \left(-\Gamma e^{-\frac{\gamma}{\varepsilon_{t_0}}} \left(e^{-\frac{E}{\varepsilon_{t_0}}} + \int_{t_0}^r \frac{e^{-\frac{E}{\varepsilon_u}}}{\sqrt{\varepsilon_u}} du \right) \right). \end{aligned}$$

Thus for any t_0 and $r \geq t_0$,

$$\begin{aligned} \mathbb{P}(X_r \in R_{E,\gamma}) &\geq \mathbb{P}(A_\beta \geq r \geq A_\alpha) \\ &\geq \exp \left(-\Gamma e^{-\frac{\gamma}{\varepsilon_{t_0}}} \left(e^{-\frac{E}{\varepsilon_{t_0}}} + \int_{t_0}^r \frac{e^{-\frac{E}{\varepsilon_u}}}{\sqrt{\varepsilon_u}} du \right) \right) - \frac{\tilde{c}}{\int_{t_0}^r (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{E}{\varepsilon_s}} ds}. \end{aligned}$$

Let $h(t)$ be defined for any $t \geq 0$ by

$$\int_t^{h(t)} \frac{e^{-\frac{E}{\varepsilon_u}}}{\sqrt{\varepsilon_u}} du = \frac{1}{\varepsilon_t}.$$

As a strictly increasing function it is invertible, and in particular $(t \rightarrow \infty) \Leftrightarrow (h(t) \rightarrow \infty)$.

$$\mathbb{P}\left(X_{h(t)} \in R_{E,\gamma}\right) \geq \exp\left(-\Gamma e^{-\frac{\gamma}{\varepsilon_t}} \left(e^{-\frac{E}{\varepsilon_t}} + \frac{1}{\varepsilon_t}\right)\right) - \tilde{c}\varepsilon_t$$

$$\xrightarrow{h(t) \rightarrow \infty} 1.$$

□

proof of Theorem 5.3. We treat first the case of fast cooling, namely we assume that for all $\delta > 0$,

$$\int_0^\infty (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{\delta}{\varepsilon_s}} ds < \infty.$$

Let x be a local minimum of U , $\delta > 0$ and $C_\delta(x)$ be the set of points which are reachable from x at height $U(x) + \delta$. Any neighborhood of x contains $C_\delta(x)$ for δ small enough. If at some time t_0 the process enters $C_\delta(x)$, Theorem 5.10 yields

$$\mathbb{P}(X_t \in C_\delta(x) \forall t \geq t_0) \geq \exp\left(-\Gamma \left(e^{-\frac{\delta}{\varepsilon_{t_0}}} + \int_{t_0}^\infty \frac{e^{-\frac{\delta}{\varepsilon_u}}}{\sqrt{\varepsilon_u}} du\right)\right) > 0.$$

Thus each time the process reach a local minimum x , it has a positive probability to stay trapped forever in a neighborhood of x . If it escapes, almost surely it will reach another local minimum later. Thus the probability that it get trapped at some time is 1, and the probability that it's already been trapped at time t goes to 1 as t goes to infinity, so that, if S is a neighborhood of all local minima of U ,

$$\lim_{t \rightarrow \infty} \mathbb{P}(X_t \in S) = 1.$$

Moreover for any local minimum x , as the temperature is not allowed to vanish, there is a non-zero probability to reach x , and so to stay trapped in $C_\delta(x)$, which yields

$$\liminf_{t \rightarrow \infty} \mathbb{P}(X_t \in C_\delta(x)) > 0.$$

Thus we have proved the part 1 and the “if” (\Rightarrow) half of part 2 of Theorem 5.3 in the case of fast cooling. Finally the “only if” half of part 2 is tautological in this case.

Concerning the part 3, recall we have supposed there is at least one non-global minimum \tilde{x} , near which the process has a non-zero probability to stay forever. In this event $U(X_t) \geq \min_{\mathbb{R}} U + \delta$ for δ small enough, so that

$$\lim_{t \rightarrow \infty} \mathbb{P}\left(U(X_t) < \min_{\mathbb{R}} U + \delta\right) \leq 1 - \mathbb{P}(X_t \text{ gets trapped near } \tilde{x}) < 1.$$

Now we turn to slow cooling, namely we suppose there exists $F > 0$ such that $\forall \delta \neq F$,

$$\left(\int_0^\infty (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{\delta}{\varepsilon_s}} ds = \infty\right) \Leftrightarrow (\delta < F).$$

Then for all $\delta < F$, according to Lemma 5.14,

$$\lim_{t \rightarrow \infty} \mathbb{P}(X_t \in R_\delta) = 1.$$

Since any neighborhood of all local minima contains R_δ for δ small enough, part 1 of Theorem 5.3 is proved.

Let $E > 0$ be such that

$$\int_0^\infty (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{E}{\varepsilon_s}} ds = \infty,$$

and S be a neighborhood of all minima of depth E such that S^c is a neighborhood of all other minima. We want to prove

$$\lim_{t \rightarrow \infty} \mathbb{P}(X_t \in S) = 0.$$

Lemma 5.14 together with the first part of Theorem 5.3 we have just proved implies that for any neighborhood \mathcal{U} of the minima in R_E ,

$$\lim_{t \rightarrow \infty} \mathbb{P}(X_t \in \mathcal{U}) = 1.$$

Thus, since S^c is a neighborhood of all the minima it contains, it is enough to prove that R_E does not contain any minimum of depth exactly E .

Let x be such a minimum, and let z be such that $U(z) < U(x)$ and z is reachable from x at height $U(x) + E$. Suppose $x \in R_E$, and let $y \in W_E$ be such that x is reachable from y at height $U(y) + E$.

If $U(y) < U(x)$, since y is reachable from x at depth $U(y) + E < U(x) + E$, by definition of the depth of a local minimum, it means x is of depth strictly less than E , which is a contradiction.

In the other hand if $U(y) \geq U(x)$, then z is reachable from y at height $U(y) + E$, while $U(z) < U(y)$, which is contradictory with the fact that y is of depth strictly larger than E .

This means R_E does not contain any local minimum of depth E .

At this point we have proved part 1 and implication \Leftarrow of part 2 of Theorem 5.3, which we will use to prove the converse.

Now suppose

$$\int_0^\infty (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{E}{\varepsilon_s}} ds < \infty.$$

In particular $E \geq F$. Let x be a minimum of depth E . As a first step, assume

$$\int_0^\infty (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{F}{\varepsilon_s}} ds < \infty.$$

Let C be the set of all points which are reachable from x at height strictly less than $U(x) + F$ (if $F = E$, $C = C_x$). Then C is a cusp of depth F , whose bottom B is constituted of minima of depth exactly E , and the depth of any other minimum in C is strictly less than F (since B is reachable from them without leaving C).

From Theorem 5.10, the process has a non-zero probability to stay trapped forever in C , so that

$$\liminf_{t \rightarrow \infty} \mathbb{P}(X_t \in C) > 0.$$

In the other hand, if $S \subset C$ is a neighborhood of B , since the depth d of any minima in $C \setminus S$ satisfies $d < F$ and so

$$\int_0^\infty (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{d}{\varepsilon_s}} ds = \infty,$$

from part 1 and implication \Leftarrow of part 2 of Theorem 5.3

$$\lim_{t \rightarrow \infty} \mathbb{P}(X_t \in C \setminus S) = 0.$$

Thus

$$\liminf_{t \rightarrow \infty} \mathbb{P}(X_t \in S) > 0$$

which ends the proof of part 2.

A slight adaptation is needed when

$$\int_0^\infty (\varepsilon_s)^{-\frac{1}{2}} e^{-\frac{F}{\varepsilon_s}} ds = \infty.$$

In this case, necessarily $F < E$. Let $\eta \in (0, F - E)$ and C be the set of all points which are reachable from x at height $U(x) + F + \eta$. Then C is a cusp of depth $F + \eta$ whose bottom B is constituted of minima of depth exactly E , and whose other minima are all of depth strictly less than $F + \eta$ (since B is reachable from them without leaving C). Since there is only a finite number of such minima, in fact if η is small enough these non-global minima in C are even of depth less than F (possibly equal).

Since $F + \eta > F$, the process has a non-zero probability to stay trapped forever in C . In the other hand, since all non-global minima in C are of depth less than F ,

$$\lim_{t \rightarrow \infty} \mathbb{P}(X_t \in C \setminus S) = 0.$$

as soon as S is a neighborhood of B . Thus the same conclusion holds.

As in the fast cooling case, part 3 is a direct consequence of parts 1 and 2 and of the presence of at least one non-global minimum. \square

5.4 Non-minimal rate

As have been seen in Chapter 3, the measure $\mu_\varepsilon = e^{-\frac{U(x)}{\varepsilon}} dx \otimes \frac{\delta_{1+\delta-1}}{2}$ is left invariant by the generator

$$Lf(x, y) = yf'(x, y) + \lambda(x, y)(f(x, -y) - f(x, y)) \quad (5.10)$$

if and only if $\lambda(x, y) = \frac{(yU'(x))_+}{\varepsilon} + r(x)$ where r can be any non-negative function. A positive r is a residual rate of jump which brings randomness in the system at any time. The previous works on explicit estimations of convergence to equilibrium for the RTP ([46, 22, 76]) all assume r is bounded below by a positive constant r_* .

If on the other hand r is bounded above, uniformly in ε and x , by a constant r^* , we may think it doesn't change the qualitative behaviour of the process. Indeed, at low temperature, it only adds random jumps at exponential (macroscopic, in the sense : of order of magnitude independant from ε) times to the minimal-rate dynamics, and the latter accounts both for the way the process overcome an energy barrier, and the metastable behaviour in the vicinity of a local minimum, namely : many microscopic excursions of length of order $\sqrt{\varepsilon}$. In the rest of this section, we will make this statement more precise, and prove it.

We will consider the context of Section 5.2, that is the homogeneous generator given by (5.1), and a double-well potential U with its three local extrema $x_0 < x_1 < x_2$. We suppose $(X_0, Y_0) = (x_0, 1)$, and we are interested in the following :

$$\begin{aligned}\tau &= \inf \{t > 0, X_t \in \{x_0, x_1\}\} \\ p_{x_0} &= \mathbb{P}(X_\tau = x_1).\end{aligned}$$

Proposition 5.15. *Whatever the residual rate of jump $x \mapsto r(x)$,*

$$p_{x_0} = \frac{e^{-\frac{U(x_1)-U(x_0)}{\varepsilon}}}{1 + \int_{x_0}^{x_1} r(z) e^{-\frac{U(x_1)-U(z)}{\varepsilon}} dz}$$

Remarks :

- A positive residual rate of jump can only worsen the probability to overcome an energy barrier in one shot.
- A positive residual rate of jump in the neighborhood of x_1 does more harm than the same rate near x_0 .

Démonstration. If $x \in (x_0, x_1)$, let p_x be the probability that the process $(X_t, Y_t)_{t \geq 0}$, starting from $(x, 1)$, reaches $(x_1, 1)$ before $(x, -1)$. Suppose the process starts at $(x - s, 1)$ for some small $s > 0$. The probability the process goes from $x - s$ to x without any jump is $1 - s \left(\frac{U'(x)}{\varepsilon} + r(x) \right) + o_{s \rightarrow 0}(s)$. The probability it reaches $(x, 1)$ before $(x - s, -1)$ but with at least one jump (and so with at least two jumps) is of order s^2 (when $s \rightarrow 0$). Once the process has reached $(x, 1)$, it has a probability p_x to reach $(x_1, 1)$ before having fallen back to $(x, -1)$. If nevertheless it has fallen back to $(x, -1)$ - which occurs with probability $(1 - p_x)$ - it has a probability $sr(x) + o_{s \rightarrow 0}(s)$ to jump before reaching $(x - s, -1)$, in which case it reaches again $(x, 1)$ with probability $1 + o_{s \rightarrow 0}(1)$. In this latter case, it reaches $(x_1, 1)$ before $(x - s, -1)$ with probability $p_x + o_{s \rightarrow 0}(1)$. Thus everything boils down to

$$\begin{aligned}p_{x-s} &= \left(1 - s \left(\frac{U'(x)}{\varepsilon} + r(x) \right) \right) (p_x + s(1 - p_x)p_x r(x)) + o_{s \rightarrow 0}(s) \\ &= p_x - sp_x \left(\frac{U'(x)}{\varepsilon} + p_x r(x) \right) + o_{s \rightarrow 0}(s).\end{aligned}$$

We shall solve the differential equation, for all $x \in (x_0, x_1)$,

$$\partial_x p_x = p_x \left(\frac{U'(x)}{\varepsilon} + p_x r(x) \right),$$

by looking for a solution of the form $p_x = h(x) e^{-\frac{U(x_1)-U(x)}{\varepsilon}}$. This yields

$$\left(\frac{1}{h} \right)'(x) = -r(x) e^{-\frac{U(x_1)-U(x)}{\varepsilon}},$$

and we conclude with $h(x_1) = p_{x_1} = 1$. □

To study τ , the length of an excursion away from x_0 , we construct the process with generator (5.10) in the following way : let $(E_k)_{k \geq 1}$ and $(F_k)_{k \geq 1}$ be two independent sequences of independent r.v. of law $\mathcal{E}(1)$. Let V_1 and W_1 be defined by

$$\varepsilon E_1 = \int_0^{V_1} (Y_0 U'(X_0 + Y_0 s))_+ ds \quad F_1 = \int_0^{W_1} r(X_0 + Y_0 s) ds.$$

Let $S_1 = \min(V_1, W_1)$. From $T_0 = 0$ to $T_1 = S_1$ the process goes on deterministically, in the sense that $Y_s = Y_0$ and $X_s = X_0 + sY_0$. At time T_1 the process jumps, so that $Y_{T_1} = -Y_0$. If $S_1 = V_1$, the process jumps due to the minimal rate; we will call this a first type jump. If $S_1 = W_1$ the jump is due to the residual rate and we will call this a second type jump. When the process has been defined up to a jump time T_j , we start the same procedure again, replacing E_1 by E_{j+1} and F_1 by F_{j+1} .

Since U is supposed smooth, $U^{(3)}(x)$ is bounded on $[x_0, x_1]$, interval on which U is strictly increasing. Thus it exists $\rho > 0$ such that for all $x \in [x_0, x_1)$, for all positive $s \leq x_1 - x$,

$$U(x + s) \geq U(x) + \rho s^2.$$

We define

$$D_k = \min \left(|x_1 - x_0|, \sqrt{\frac{\varepsilon E_{2k-1}}{\rho}} \right),$$

which is independent from the sequence $(F_k)_{k \geq 0}$. In the other hand, we consider the events

$$A_k = \left\{ F_k < \int_0^{x_1 - x_0} r(z) dz \right\}$$

and the geometric variable

$$N = \max \{k \geq 1, A_{2k}\},$$

which is independent from the sequence $(E_k)_{k \geq 0}$.

Lemme 5.16.

$$\tau \leq 2 \sum_{k=1}^N D_k.$$

Démonstration. From the time 0 to τ , the process alternates between ascending and downward moves. If $X_\tau = x_0$, the process spends as much time going up as going down, and if $X_\tau = x_1$ then it spends more time going up than down; in either case, τ is less than twice the cumulated ascending time. Let n be the number of ascending moves before the time τ . At the end of the k^{th} ascending move, the next time of jump will be defined thanks to the variable F_{2k} (as long as the process does not reach $(x_0, -1)$). In the event \bar{A}_{2k} (namely the negation of A_{2k}) there is no second type jump before the process reaches $(x_0, -1)$. This means $n \leq N$.

Let $(d_k)_{1 \leq k \leq n}$ be the duration of the n ascending moves, and $(Z_k)_{1 \leq k \leq n}$ be the starting point of these moves (for instance $Z_1 = x_0$).

$$\begin{aligned} d_k &\leq V_{2k-1} \\ &= \sup \{t < (x_1 - x), \varepsilon E_{2k-1} > U(Z_k + t) - U(Z_k)\} \\ &\leq \sup \{t < (x_1 - x_0), \varepsilon E_{2k-1} > \rho t^2\} \\ &= D_k. \end{aligned}$$

Finally

$$\tau \leq 2 \sum_{k=1}^n d_k \leq 2 \sum_{k=1}^N D_k.$$

□

Now if r is supposed bounded above by a constant r^* ,

$$\mathbb{E}[N] = 1 + \frac{1}{\mathbb{P}(\bar{A}_2)} \leq 1 + e^{(x_1 - x_0)r^*},$$

and thus, N and the D_k 's being independant, $\mathbb{E}[\tau] \leq c\sqrt{\varepsilon}$ for some constant c . In the other hand,

$$\mathbb{E}[\tau] \geq \mathbb{P}(\bar{A}_1, \bar{A}_2) \mathbb{E}[\tau \mid \bar{A}_1, \bar{A}_2] \geq e^{-2(x_1 - x_0)r^*} \mathbb{E}[\min(V_1, x_1 - x_0)].$$

From Lemma 5.4 we get that $\mathbb{E}[\tau] \geq c'\sqrt{\varepsilon}$ for some $c' > 0$. But if $\mathbb{E}[\tau]$ is of order $\sqrt{\varepsilon}$, it means when ε goes to 0 it is unlikely that a second type jump occurs during an excursion, so that only the asymptotics of $\mathbb{E}[V_1]$ should intervene. Indeed we can prove the following :

Proposition 5.17. *If $r \leq r^*$,*

$$\mathbb{E}[\tau] = \sqrt{\frac{2\pi\varepsilon}{U''(x_0)}} \left(1 + \underset{\varepsilon \rightarrow 0}{o}(1)\right)$$

Démonstration. When $r \leq r^*$, $F_k \leq r^*W_k$. Since in the other hand $V_k \leq D_k$ always holds, the event $B = \{r^*D_1 \leq \min(F_1, F_2, \delta)\}$ (for some $\delta > 0$) implies $V_1 \leq \min(W_1, W_2)$: there is no second type jump during the excursion. We decompose

$$\mathbb{E}[\tau] = \mathbb{E}[\tau 1_B] + \mathbb{E}[\tau 1_{\bar{B}}].$$

First, note that $\mathbb{P}(B)$ goes to 1 when ε goes to 0. Second,

$$\begin{aligned} \mathbb{E}[1_{\bar{B}}\tau] &\leq 2\mathbb{E}\left[1_{\bar{B}} \sum_{k=1}^N D_k\right] \\ &= 2\mathbb{E}[1_{\bar{B}}D_1] + 2\mathbb{P}(\bar{B}) \mathbb{E}[N - 1] \mathbb{E}[D_2] \\ &\leq 2\sqrt{\mathbb{P}(\bar{B}) \mathbb{E}[D_1^2]} + 2\mathbb{P}(\bar{B}) \mathbb{E}[N - 1] \mathbb{E}[D_2] \\ &= \underset{\varepsilon \rightarrow 0}{o}(\sqrt{\varepsilon}). \end{aligned}$$

If $\delta < r^*|x_1 - x_0|$, when B holds, the process goes up until a first time jump occurs, then it falls back to x_0 , and no second type jump occurs : $\tau = 2V_1$, and

$$\begin{aligned} \varepsilon E_1 &= U(x_0 + V_1) - U(x_0) \\ &= \frac{1}{2}U''(x_0) \left(\frac{V_1}{\sqrt{\varepsilon}}\right)^2 \varepsilon + \underset{\varepsilon \rightarrow 0}{o}(\varepsilon) \\ \Rightarrow \frac{V_1}{\sqrt{\varepsilon}} &\xrightarrow{\varepsilon \rightarrow 0} \sqrt{\frac{2E_1}{U''(x_0)}} \quad a.s. \end{aligned}$$

Moreover $1_{\bar{B}} \rightarrow 1$ almost surely, and

$$\frac{V_1}{\sqrt{\varepsilon}} 1_{\bar{B}} \leq \frac{D_1}{\sqrt{\varepsilon}} \leq \sqrt{\frac{E_1}{\rho}}.$$

By the dominated convergence theorem,

$$2\mathbb{E}[V_1 1_{\bar{B}}] = 2\sqrt{\varepsilon}\mathbb{E}\left[\sqrt{\frac{2E_1}{U''(x_0)}}\right] \left(1 + o_{\varepsilon \rightarrow 0}(1)\right),$$

which concludes. □

As a conclusion, note that when $r \leq r^*$,

$$e^{-\frac{U(x_1)-U(x_0)}{\varepsilon}} \geq p_{x_0} \geq \frac{e^{-\frac{U(x_1)-U(x_0)}{\varepsilon}}}{1 + (x_1 - x_0)r^*},$$

and the adaptation of the whole argumentation would prove Theorem 5.3 still holds. In the other hand if r goes to infinity as ε goes to 0, the velocities at two different instants are more and more decorrelated, and we may think the (suitably rescaled) process ends up with a genuine diffusive part, exactly as we have already proved in Chapter 2.

Bilan et perspectives

Au moment de tirer quelques conclusions sur l'ensemble de ce mémoire, un regard un peu rapide - ou un peu trop indulgent - pourrait dresser le bilan positif suivant : l'objectif était d'établir un résultat de convergence pour un algorithme de recuit simulé basé sur une dynamique cinétique, ce qui amenait à l'objectif secondaire de développer de nouvelles méthodes hypocoercives. Nous avons bel et bien une nouvelle méthode (la méthode d'ordre 3 du chapitre 3), et nous avons bien un résultat (le critère nécessaire et suffisant du chapitre 5). Avant même de discuter des mérites et lacunes respectifs de chacune de ces parties, constatons d'emblée ceci : le nouveau résultat n'a pas été démontré par les nouvelles méthodes... Cette scission entre les deux objectifs initiaux du projet de recherche de la thèse amène à deux directions relativement indépendantes quant aux futurs développements, que nous allons donc distinguer.

En premier lieu, penchons-nous sur le recuit simulé par des processus cinétiques.

S'agissant du recuit simulé, il semble que si une vérité est bien établie à la fois pour les théoriciens et les praticiens, et pour toute interpolation entre ces deux points extrêmes, c'est la persistance, malgré les efforts louables depuis plusieurs décennies, d'un fossé net entre la théorie et la pratique. Les résultats de cette thèse sont évidemment entièrement théoriques, et il y a encore de nombreuses questions de cet ordre, dont nous avons déjà parlé : donner des estimations de la probabilité d'avoir atteint un minimum à temps fixé, étudier des schémas géométriques à horizon de temps fini, étudier le RTP en dimension supérieure ; la liste est loin d'être exhaustive. Mais surtout il s'agit de s'interroger sur l'application concrète des idées qui ont été développées ici. C'est d'ailleurs l'occasion de constater et de regretter le manque d'interaction (assumé ; on ne peut après tout pas tout faire en trois ans) durant cette thèse avec des problèmes pratiques concrets. Palier cette lacune est évidemment l'une des perspectives essentielles pour la suite de mon travail de recherche.

Le bilan de la thèse n'est évidemment pas uniformément strictement négatif : après tout, on peut retirer du chapitre 5 l'idée d'une dynamique qui a bel et bien l'avantage d'être très simple à simuler et de franchir les barrières de potentiel plus efficacement que la diffusion réversible. Le fait que notre résultat n'a cours qu'en dimension un ne me semble pas anodin : indépendamment de la difficulté théorique - comprendre comment une ligne brisée peut s'échapper d'un domaine non convexe (à rapprocher des études de billards ?), et quelle est l'influence du taux de saut - il n'est surtout pas clair du tout que le RTP soit une bonne dynamique pour gérer les barrières entropiques. Par exemple si le potentiel est infini en-dehors d'un domaine D et nul sur D , et que le domaine D est constitué de deux boules connectées par un corridor étroit, n'est-ce pas plus compliqué - surtout en grande dimension ? - qu'avec un brownien de se retrouver non seulement à l'entrée du corridor mais en plus avec une vitesse dans la bonne direction ? On peut se dire que, si la vitesse converge vite à l'équilibre (par rapport à la position), se retrouver devant le corridor avec une vitesse uniforme, c'est se comporter comme un

brownien... mais on perd dans ce cas l'avantage donné par l'inertie.

On peut toujours envisager de n'adopter la dynamique du RTP que dans un nombre très petit de dimensions, et de garder un comportement diffusif dans les autres ; il s'agit alors, comme dans le cas de la méthode ABF (cf. [68]), de déterminer les directions d'intérêts. Elles peuvent éventuellement être proposées par les praticiens ; mais cela m'amène à un autre point.

Quelque soit le processus de Markov considéré (continu en la position), s'il met en moyenne un temps constant (indépendant de la température ε donc) à parcourir un chemin de réaction (c'est-à-dire une trajectoire qui part d'une boule B_1 autour d'un minimum local et l'amène à une boule B_2 autour d'un autre minimum local sans jamais repasser par B_1), nécessairement le temps de franchissement de la barrière d'énergie E est, à l'échelle logarithmique au moins, de l'ordre de $e^{-\frac{E}{\varepsilon}}$. En effet (l'argument suivant a déjà été donné au chapitre 5), si le potentiel U n'a aucun autre minimum que les centres de B_1 et B_2 , la probabilité (sous la loi invariante) d'être en-dehors de ces boules est de l'ordre de $\exp\left(-\frac{\min_{(B_1 \cup B_2)^c} U - \min U}{\varepsilon}\right)$, alors que la probabilité d'être dans les boules est d'ordre 1. Le rapport entre ces probabilités doit être égal, par ergodicité, au rapport des temps moyens passés par le processus dans ces régions. Ceci implique qu'entre deux traversées, un temps d'ordre exponentiel s'écoule. En conclusion, on ne peut pas espérer régler le problème de la métastabilité simplement en changeant la dynamique markovienne locale.

De grands sauts dans la position ne constituent pas non plus une bonne solution : que ce soit pour des méthodes bayésiennes en imagerie (cf. la conclusion de la thèse de François Giraud [49]), ou en chimie pour déterminer le repliement d'une molécule, la masse de la loi de Gibbs cible (même à température fixée) se concentre sur des ensembles de mesure de Lebesgue infime. Sauter au hasard (au sens uniforme), en grande dimension, c'est à peu près à coup sûr tomber sur une configuration absurde, qui sera rejetée.

Il s'agit donc, encore une fois, de trouver les bonnes directions selon lesquelles faire évoluer les paramètres ; et pas simplement de trouver ces directions, *in fine*, en ayant marché au hasard. Prenons le problème du repliement de molécule. Si l'on isolait des éléments statistiquement significatifs de la géométrie globale de la molécule, on pourrait concevoir des chemins qui modifierait de façon privilégiée ces paramètres. Alors, la dynamique du RTP pourrait être utilisée le long de ces chemins. Évidemment, je ne suis pas spécialiste du domaine, ce genre de choses existe sûrement déjà ; ce qui me ramène à la perspective numéro un de ce paragraphe : aller voir du côté des applications concrètes. On renvoie le lecteur intéressé à la synthèse [55].

En second lieu, nous n'en avons pas fini avec l'hypocoercivité.

Les perspectives sur cette question à la fin de la thèse sont peu ou prou les mêmes qu'au début. Un point positif, tout de même, est d'être parvenu à capter, quantitativement, les possibles oscillations de la création d'entropie au cours du temps.

Une direction m'intéresse particulièrement pour le futur proche : les techniques de type calcul Γ , à la Baudoin ([11]), effleurées dans le chapitre 4. Un résultat satisfaisant serait par exemple des inégalités locales dans l'esprit du critère de Bakry-Emery (proposition 1.15 de l'introduction) qui pourraient porter, disons, sur le processus de Langevin cinétique

$$L = y\partial_x - (U'(x) + by)\partial_y + \partial_y^2$$

pour un potentiel tel que $U'' \geq a > 0$ (ou, en dimension supérieure, $z^T \text{Hess}_U(x)z \geq a|z|^2$ pour tous x, z), et ne ferait intervenir que a et b . Par exemple dans le cas gaussien $U'(x) = ax$, en

prenant $b = 1$, $a \geq \frac{1}{4}$, et en notant

$$\begin{aligned} Df &= \begin{pmatrix} \frac{1}{2}\partial_y f - \partial_x f \\ \sqrt{a - \frac{1}{4}}\partial_y f \end{pmatrix} \\ \psi_a(s) &= P_s |DP_{t_s} f|^2, \end{aligned}$$

on vérifie aisément que $\psi'_a(s) \geq \psi_a(s)$. La contrainte $a \geq \frac{1}{4}$ n'est pas anodine : c'est exactement la valeur qui marque la frontière entre les deux comportements différents de la convergence à l'équilibre - avec ou sans oscillation -, comme dans le chapitre 2 avec le télégraphe circulaire (cf. le calcul explicite de Gadat et Miclo [47]). Ensuite, plutôt que d'utiliser l'équivalence pour $a > \frac{1}{4}$ de $|Df|^2$ avec $|\nabla f|^2$ (qui donnerait une très mauvaise constante, et rien pour $a = \frac{1}{4}$), on revient au calcul du Γ_2 usuel et l'on contrôle le terme qui nous dérangeait jusqu'ici grâce à l'information sur ψ_a , ce qui amène après quelques manipulations à

$$(\partial_y P_t f)^2 \leq e^{-t} P_t \left((1+t)(\partial_y f)^2 + \frac{(1+t)^3 - 1}{3} |Df|^2 \right).$$

En procédant de même pour le gradient en x , *in fine*, on peut montrer que pour $a \geq \frac{1}{4}$,

$$|\nabla P_t f|^2 \leq e^{-t} \left(1+t + (1+a^2) \frac{(1+t)^3 - 1}{2} \right) P_t |\nabla f|^2.$$

On voit qu'en utilisant des variables intermédiaires, on obtient exactement le bon taux de convergence (cf. [47], encore une fois), et un préfacteur raisonnable.

Évidemment, une méthode qui fonctionne dans le cas gaussien demande à être éprouvée plus généralement. Ceci fournit une autre perspective : trouver de nouveaux bancs d'essais aux méthodes envisagées. On peut citer les chaînes de particules (voir, par exemple, dans [43]) et les browniens cinétiques sur des variétés (cf. [4]) ; mais le champ des possibles est en l'occurrence particulièrement vaste.

Bibliographie

- [1] S. Aida, T. Masuda, and I. Shigekawa. Logarithmic Sobolev inequalities and exponential integrability. *J. Funct. Anal.*, 126(1) :83–101, 1994.
- [2] S. Aida and I. Shigekawa. Logarithmic sobolev inequalities and spectral gaps : Perturbation theory. *Journal of Functional Analysis*, 1994.
- [3] L. Ambrosio, N. Gigli, and G. Savaré. Metric measure spaces with Riemannian Ricci curvature bounded from below. *Duke Math. J.*, 163(7) :1405–1490, 2014.
- [4] J. Angst and C. Tardif. Dévissage of a Poisson boundary under equivariance and regularity conditions. *ArXiv e-prints*, November 2013.
- [5] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques*. Panoramas et synthèses. Société mathématique de France, Paris, 2000.
- [6] A. Arnold and J. Erb. Sharp entropy decay for hypocoercive and non-symmetric Fokker-Planck equations with linear drift. *ArXiv e-prints*, September 2014.
- [7] R. Azaïs, J.-B. Bardet, A. Genadot, N. Krell, and P.-A. Zitt. Piecewise deterministic Markov process - recent results. *ArXiv e-prints*, September 2013.
- [8] D. Bakry, F. Baudoin, M. Bonnefont, and D. Chafaï. On gradient bounds for the heat kernel on the Heisenberg group. *J. Funct. Anal.*, 255(8) :1905–1938, 2008.
- [9] D. Bakry, P. Cattiaux, and A. Guillin. Rate of convergence for ergodic continuous Markov processes : Lyapunov versus Poincaré. *J. Funct. Anal.*, 254(3) :727–759, 2008.
- [10] J.-B. Bardet, A. Christen, A. Guillin, F. Malrieu, and P.-A. Zitt. Total variation estimates for the TCP process. *Electron. J. Probab.*, 18 :no. 10, 21, 2013.
- [11] F. Baudoin. Bakry-Emery meet Villani. *ArXiv e-prints*, August 2013.
- [12] W. Beckner. A generalized Poincaré inequality for Gaussian measures. *Proc. Amer. Math. Soc.*, 105(2) :397–400, 1989.
- [13] M. Beiglböck, C. Léonard, and W. Schachermayer. A general duality theorem for the Monge-Kantorovich transport problem. *Studia Math.*, 209(2) :151–167, 2012.
- [14] M. Benaïm, I. Ciotir, and C.-E. Gauthier. Self-repelling diffusions via an infinite dimensional approach. *ArXiv e-prints*, August 2014.
- [15] M. Benaïm, S. Le Borgne, F. Malrieu, and P.-A. Zitt. Quantitative ergodicity for some switched dynamical systems. *Electron. Commun. Probab.*, 17 :no. 56, 14, 2012.
- [16] M. Benaïm, S. Le Borgne, F. Malrieu, and P.-A. Zitt. On the stability of planar randomly switched systems. *Ann. Appl. Probab.*, 24(1) :292–311, 2014.
- [17] F. Bolley and I. Gentil. Phi-entropy inequalities for diffusion semigroups. *J. Math. Pures Appl. (9)*, 93(5) :449–473, 2010.

- [18] M. Bonnefont and A. Joulin. Intertwining relations for one-dimensional diffusions and application to functional inequalities. *Potential Analysis*, April 2014.
- [19] F. Bouguet. Quantitative speeds of convergence for exposure to food contaminants. *ArXiv e-prints*, October 2013.
- [20] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes. I. Sharp asymptotics for capacities and exit times. *J. Eur. Math. Soc. (JEMS)*, 6(4) :399–424, 2004.
- [21] A. Bovier, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes. II. Precise asymptotics for small eigenvalues. *J. Eur. Math. Soc. (JEMS)*, 7(1) :69–99, 2005.
- [22] V. Calvez, G. Raoul, and C. Schmeiser. Confinement by biased velocity jumps : aggregation of *Escherichia coli*. *ArXiv e-prints*, April 2014.
- [23] O. Catoni. Rough large deviation estimates for simulated annealing : application to exponential schedules. *Ann. Probab.*, 20(3) :1109–1146, 1992.
- [24] P. Cattiaux, A. Guillin, and P.-A. Zitt. Poincaré inequalities and hitting times. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(1) :95–118, 2013.
- [25] F. Cérou, A. Guyader, T. Lelièvre, and F. Malrieu. On the length of one-dimensional reactive paths. *ALEA Lat. Am. J. Probab. Math. Stat.*, 10(1) :359–389, 2013.
- [26] D. Chafaï. Entropies, convexity, and functional inequalities : on Φ -entropies and Φ -Sobolev inequalities. *J. Math. Kyoto Univ.*, 44(2) :325–363, 2004.
- [27] D. Chafaï, F. Malrieu, and K. Paroux. On the long time behavior of the TCP window size process. *Stochastic Process. Appl.*, 120(8) :1518–1534, 2010.
- [28] T-S Chiang, C-R Hwang, and S J Sheu. Diffusion for global optimization in \mathbb{R}^n . *SIAM J. Control Optim.*, 25(3) :737–753, 1987.
- [29] B. Cloez and M. Hairer. Exponential ergodicity for Markov processes with random switching. *ArXiv e-prints*, March 2013.
- [30] O. L. V. Costa and F. Dufour. Stability and ergodicity of piecewise deterministic markov processes. *SIAM Journal on Control and Optimization*, 47 :1053–1077, 2008.
- [31] M. H. A. Davis. *Markov models and optimization*, volume 49 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1993.
- [32] L. Desvillettes and C. Villani. On the trend to global equilibrium in spatially inhomogeneous entropy-dissipating systems : the linear Fokker-Planck equation. *Comm. Pure Appl. Math.*, 54(1) :1–42, 2001.
- [33] L. Desvillettes and C. Villani. On the trend to global equilibrium for spatially inhomogeneous kinetic systems : the Boltzmann equation. *Invent. Math.*, 159(2) :245–316, 2005.
- [34] J.-D. Deuschel and D. W. Stroock. *Large deviations*, volume 137 of *Pure and Applied Mathematics*. Academic Press, Inc., Boston, MA, 1989.
- [35] P. Diaconis, S. Holmes, and R. M. Neal. Analysis of a nonreversible Markov chain sampler. *Ann. Appl. Probab.*, 10(3) :726–752, 2000.
- [36] P. Diaconis and L. Miclo. On the spectral analysis of second-order Markov chains. *Ann. Fac. Sci. Toulouse Math. (6)*, 22(3) :573–621, 2013.
- [37] J. Dieudonné. *Calcul infinitésimal*. Hermann, Paris, 1968.

- [38] J. Dolbeault, C. Mouhot, and C. Schmeiser. Hypocoercivity for kinetic equations with linear relaxation terms. *C. R. Math. Acad. Sci. Paris*, 347(9-10) :511–516, 2009.
- [39] J. Dolbeault, C. Mouhot, and C. Schmeiser. Hypocoercivity and stability for a class of kinetic models with mass conservation and a confining potential. To appear in Transactions of the American Mathematical Society, 2013.
- [40] R. Douc, G. Fort, and A. Guillin. Subgeometric rates of convergence of f -ergodic strong Markov processes. *Stochastic Process. Appl.*, 119(3) :897–923, 2009.
- [41] D. Down, S. P. Meyn, and R. L. Tweedie. Exponential and uniform ergodicity of Markov processes. *Ann. Probab.*, 23(4) :1671–1691, 1995.
- [42] V. Dumas, F. Guillemin, and P. Robert. A Markovian analysis of additive-increase multiplicative-decrease algorithms. *Adv. in Appl. Probab.*, 34(1) :85–111, 2002.
- [43] J.-P. Eckmann, C.-A. Pillet, and L. Rey-Bellet. Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures. *Comm. Math. Phys.*, 201(3) :657–697, 1999.
- [44] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics : Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.
- [45] F. Filbet, C. Mouhot, and L. Pareschi. Solving the Boltzmann equation in $N \log_2 N$. *SIAM J. Sci. Comput.*, 28(3) :1029–1053 (electronic), 2006.
- [46] J. Fontbona, H. Guérin, and F. Malrieu. Quantitative estimates for the long-time behavior of an ergodic variant of the telegraph process. *Adv. in Appl. Probab.*, 44(4) :977–994, 2012.
- [47] S. Gadat and L. Miclo. Spectral decompositions and \mathbb{L}^2 -operator norms of toy hypocoercive semi-groups. *Kinet. Relat. Models*, 6(2) :317–372, 2013.
- [48] S. Gadat and F. Panloup. Long time behaviour and stationary regime of memory gradient diffusions. *Annales de l’Institut Henri Poincaré*, to appear, 2013.
- [49] F. Giraud. *Analyse des modèles particuliers de Feynman-Kac et application à la résolution de problèmes inverses en électromagnétisme*. PhD thesis, Université de Bordeaux, 2013.
- [50] F. Giraud and P. Del Moral. Non-Asymptotic Analysis of Adaptive and Annealed Feynman-Kac Particle Models. *ArXiv e-prints*, September 2012.
- [51] L. Gross. Logarithmic Sobolev inequalities on loop groups. *J. Funct. Anal.*, 102(2) :268–313, 1991.
- [52] M. Hairer. On Malliavin’s proof of Hörmander’s theorem. *Bull. Sci. Math.*, 135(6-7) :650–666, 2011.
- [53] B. Hajek. Cooling schedules for optimal annealing. *Math. Oper. Res.*, 13(2) :311–329, 1988.
- [54] B. Helffer and F. Nier. *Hypoelliptic estimates and spectral theory for Fokker-Planck operators and Witten Laplacians*, volume 1862 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2005.
- [55] D. Henderson, S. H. Jacobson, and A. W. Johnson. The theory and practice of simulated annealing. In *Handbook of metaheuristics*, volume 57 of *Internat. Ser. Oper. Res. Management Sci.*, pages 287–319. Kluwer Acad. Publ., Boston, MA, 2003.

- [56] F. Hérau. Hypocoercivity and exponential time decay for the linear inhomogeneous relaxation Boltzmann equation. *Asymptot. Anal.*, 46(3-4) :349–359, 2006.
- [57] R. Holley and D. W. Stroock. Simulated annealing via Sobolev inequalities. *Comm. Math. Phys.*, 115(4) :553–569, 1988.
- [58] R. A. Holley, S. Kusuoka, and D. W. Stroock. Asymptotics of the spectral gap with applications to the theory of simulated annealing. *J. Funct. Anal.*, 83(2) :333–347, 1989.
- [59] L. Hörmander. Hypoelliptic second order differential equations. *Acta Math.*, 119 :147–171, 1967.
- [60] V. Kontis, M. Ottobre, and B. Zegarliński. Markov semigroups with hypocoercive-type generator in Infinite Dimensions I : Ergodicity and Smoothing. *ArXiv e-prints*, June 2013.
- [61] V. Kontis, M. Ottobre, and B. Zegarliński. Markov semigroups with hypocoercive-type generator in Infinite Dimensions II : Applications. *ArXiv e-prints*, June 2013.
- [62] K. Kuwada. Duality on gradient estimates and Wasserstein controls. *J. Funct. Anal.*, 258(11) :3758–3774, 2010.
- [63] R. Latała and K. Oleszkiewicz. Between Sobolev and Poincaré. In *Geometric aspects of functional analysis*, volume 1745 of *Lecture Notes in Math.*, pages 147–168. Springer, Berlin, 2000.
- [64] S. D. Lawley, J. C. Mattingly, and M. C. Reed. Sensitivity to switching rates in stochastically switched ODEs. *Commun. Math. Sci.*, 12(7) :1343–1352, 2014.
- [65] M. Ledoux. L’algèbre de Lie des gradients itérés d’un générateur markovien—développements de moyennes et entropies. *Ann. Sci. École Norm. Sup. (4)*, 28(4) :435–460, 1995.
- [66] T. Lelièvre. Two mathematical tools to analyze metastable stochastic processes. *ArXiv e-prints*, January 2012.
- [67] T. Lelièvre, F. Nier, and G. A. Pavliotis. Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *Journal of Statistical Physics, to appear*, December 2012.
- [68] T. Lelièvre, M. Rousset, and G. Stoltz. *Free energy computations : A mathematical perspective*. Imperial College Press, 2010.
- [69] A. H. Löpker and J. S. H. van Leeuwen. Transient moments of the TCP window size process. *J. Appl. Probab.*, 45(1) :163–175, 2008.
- [70] F. Malrieu. Some simple but challenging markov processes. Preprint.
- [71] F. Malrieu and D. Talay. Concentration inequalities for euler schemes. *Monte Carlo and Quasi-Monte Carlo Methods*.
- [72] S. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition, 2009. With a prologue by Peter W. Glynn.
- [73] L. Miclo. Recuit simulé sur R^n . Étude de l’évolution de l’énergie libre. *Ann. Inst. H. Poincaré Probab. Statist.*, 28(2) :235–266, 1992.
- [74] L. Miclo. Une étude des algorithmes de recuit simulé sous-admissibles. *Ann. Fac. Sci. Toulouse Math. (6)*, 4(4) :819–877, 1995.
- [75] L. Miclo and P. Monmarché. Étude spectrale minutieuse de processus moins indécis que les autres. 2078 :459–481, 2013.

- [76] P. Monmarché. Hypocoercive relaxation to equilibrium for some kinetic models. *Kinet. Relat. Models*, 7(2) :341–360, 2014.
- [77] N. Ratanov. Double telegraph processes and complete market models. *Stoch. Anal. Appl.*, 32(4) :555–574, 2014.
- [78] M. Reed and B. Simon. *Methods of modern mathematical physics. IV. Analysis of operators*. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1978.
- [79] G. Royer. A remark on simulated annealing of diffusion processes. *SIAM J. Control Optim.*, 27(6) :1403–1408, 1989.
- [80] G. Royer. *Une initiation aux inégalités de Sobolev logarithmiques*. Société Mathématique de France, 1999.
- [81] W. Rudin. *Functional analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, second edition, 1991.
- [82] L. Rüschendorf. Monge-Kantorovich transportation problem and optimal couplings. *Jahresber. Deutsch. Math.-Verein.*, 109(3) :113–137, 2007.
- [83] A. Scemama, T. Lelièvre, G. Stoltz, and M. Caffarel. An efficient sampling algorithm for variational monte carlo. *Journal of Chemical Physics*, 125, September 2006.
- [84] D. Talay. Stochastic Hamiltonian systems : exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Process. Related Fields*, 8(2) :163–198, 2002. Inhomogeneous random systems (Cergy-Pontoise, 2001).
- [85] C. Villani. Hypocoercive diffusion operators. In *International Congress of Mathematicians. Vol. III*, pages 473–498. Eur. Math. Soc., Zürich, 2006.
- [86] C. Villani. Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950) :iv+141, 2009.
- [87] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.
- [88] P-A Zitt. Annealing diffusions in a potential function with a slow growth. *Stochastic Process. Appl.*, 118(1) :76–119, 2008.

Hypocoercivité : approches alternatives et applications aux algorithmes stochastiques

Auteur : Pierre Monmarché

Directeur de thèse : Laurent Miclo

Date et lieu de soutenance : le 10-12-2014 à l'Institut de Mathématiques de Toulouse.

Discipline : Mathématiques

Résumé

Dans cette thèse, des dynamiques markoviennes alternatives à la diffusion réversible usuelle sont considérées pour échantillonner une mesure de Gibbs dans le cadre d'un algorithme de recuit simulé. Le problème de la convergence, à température fixée, de ces processus vers leur mesure invariante amène à des questions d'hypocoercivité. Dans la mesure où les résultats antérieurs dans le domaine ne donnent pas d'asymptotiques précises du taux de convergence à basse température, de nouvelles méthodes pour obtenir de tels taux explicites sont proposées et étudiées, notamment sur les processus de Markov déterministes par morceaux. Enfin une condition optimale sur le schéma de température d'un recuit simulé basé sur le RTP complètement dégénéré est obtenue en dimension un.

Mots clés : hypocoercivité, processus cinétiques, PDMP, recuit simulé, inégalités fonctionnelles.

Abstract

Some Markov dynamics are considered to sample Gibbs laws in the framework of the simulated annealing algorithm, as possible alternatives to the usual reversible diffusion. The problem of the convergence, at fixed temperature, of these processes toward their equilibrium leads to hypocoercivity questions. Since the previous results in this field do not yield sharp asymptotics for the convergence rate at low temperature, new methods are investigated, in particular in the case of piecewise deterministic Markov processes. Finally an optimal condition is given for the cooling schedule of a simulated annealing based on the fully degenerate RTP in dimension one.

Key words : hypocoercivity, kinetic processes, PDMP, simulated annealing, functional inequalities.

Institut de Mathématiques de Toulouse
118 route de Narbonne
F-31062 Toulouse Cedex

Ecole doctorale EDMITT
Bâtiment 1R1 RDC porte 5
31062 Toulouse cedex