



Comportement en temps long de processus de Markov

Pierre-André Zitt

► To cite this version:

Pierre-André Zitt. Comportement en temps long de processus de Markov. Probabilités [math.PR]. Université Paris Est, 2014. tel-01132419

HAL Id: tel-01132419

<https://theses.hal.science/tel-01132419>

Submitted on 17 Mar 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**LABORATOIRE D'ANALYSE ET DE
MATHÉMATIQUES APPLIQUÉES**

Sous la co-tutelle de :
CNRS

UPEC • UNIVERSITÉ PARIS-EST CRÉTEIL

UPEM • UNIVERSITÉ PARIS-EST MARNE-LA-VALLÉE

Mémoire d'habilitation à diriger les recherches

Pierre-André Zitt

Comportement en temps long de processus de Markov

Habilitation soutenue le 28 novembre 2014, devant le jury constitué de :

Marc ARNAUDON	Université de Bordeaux I
Bernard BERCU	Université de Bordeaux I
Patrick CATTIAUX	Université Paul Sabatier
Benjamin JOURDAIN	École Nationale des Ponts et Chaussées
Eva LÖCHERBACH	Université de Cergy-Pontoise
Florence MERLEVÈDE	Université Paris-Est Marne-la-Vallée
Sylvie ROELLY	Université de Potsdam

au vu des rapports de

Marc ARNAUDON	Université de Bordeaux I
Eva LÖCHERBACH	Université de Cergy-Pontoise
Prasad TETALI	Georgia Tech

Habilitation à diriger des recherches

Pierre-André Zitt

17 mars 2015

Table des matières

Table des matières	3
Remerciements	5
1 Introduction	7
1.1 Le comportement en temps long de processus de Markov	7
1.2 Des approches classiques	12
1.3 Plan du mémoire	16
2 Autour des inégalités fonctionnelles	19
2.1 Inégalités faibles pour un modèle de mécanique statistique	19
2.2 Spectre essentiel et « super-inégalités de Poincaré »	21
2.3 Temps de retour, fonctions de Lyapunov et inégalités fonctionnelles	23
3 Algorithmes stochastiques	25
3.1 Le recuit simulé	25
3.2 Une estimation de médiane en dimension infinie	27
4 Processus de Markov déterministes par morceaux	31
4.1 Introduction	31
4.2 Le zéro ou l'infini ?	33
4.3 Convergence « abstraite » vers l'équilibre	36
4.4 Convergence vers l'équilibre : couplages astucieux	39
5 Miscellanées	43
5.1 Gaz de Coulomb	43
5.2 Un modèle d'évolution	46
Publications	51
Bibliographie	53

Remerciements

Je tiens tout d’abord à remercier Marc Arnaudon, Eva Löcherbach et Prasad Tetali d’avoir accepté de faire un rapport sur ce mémoire, et Bernard Bercu et Benjamin Jourdain et Florence Merlevède d’avoir bien voulu faire partie de mon jury.

Patrick Cattiaux m’a fait découvrir en thèse le monde des inégalités fonctionnelles, et elles continuent à planer autour de mes travaux, en guettant le bon moment pour faire leur retour. . .

Les années autour de ma thèse ont été riches en voyage à Berlin et Potsdam ; pendant ces séjours, Sylvie Roelly m’a toujours fait bon accueil ¹, et je suis heureux d’avoir enfin une occasion de lui offrir du champagne !

★ ★ ★

Les mathématiques ont en commun avec le cyclisme d’être un sport individuel qui se court en équipe. Parmi mes équipiers, je me dois de mentionner tout particulièrement le duo de duaux topologiques, Djalil Chafaï et Florent Malrieu. L’appétit de Florent pour les mathématiques ², son goût pour les « petits exemples » et son enthousiasme communicatif sont autant de qualités qui ne peuvent que pâlir devant son don inégalé pour le mime de processus stochastiques.

La curiosité pas si ordinaire de Djalil, son souci du détail et de la juste rédaction ³, et son honnêteté frisant parfois ⁴ la provocation en font un collègue extrêmement stimulant. L’équilibre « exploration/exploitation » est souvent difficile à trouver en matière de recherche, et Djalil est une des rares personnes qui incitent à « sortir des puits de potentiel » pour aller voir ce qui se fait un peu plus loin... L’avenir dira si j’arriverai à suivre ses encouragements à « [ne pas] faire du temps long toute [ma] vie »...

Merci également à Michel Benaïm, avec qui j’ai bu du café avant de collaborer, et Stéphane Le Borgne, avec qui j’ai collaboré avant d’avoir l’occasion de prendre un café ; la longue aventure « PDMP » n’aurait pas été la même sans eux.

★ ★ ★

J’ai maîtrisé mes premières conférences à l’université de Bourgogne, à Dijon, où j’ai pu nouer une collaboration fructueuse avec Peggy Cénac et Hervé Cardot. J’ai également eu l’opportunité d’y contribuer au lancement de l’initiative d’excellence PODEX — POts D’EXcellence — initiative mettant à profit toutes les bonnes pratiques de *crowdsourcing*, pour encourager les synergies dans une logique d’ouverture à l’international, ce qui a permis l’obtention d’excellents livrables, en particulier sous forme de tajines.

1. Les esprits chagrins regretteront peut-être que son influence ne s’étende pas jusqu’au choix de la décoration intérieure des chambres d’hôtel de Potsdam.

2. et la fondue savoyarde

3. Malgré nos désaccords sur le rapport signal/bruit optimal !

4. Souvent ?

Le passage de Dijon⁵ à Marne-la-Vallée⁶ n'a pas eu que des inconvénients. En plus d'un univers accueillant pour les mauvais jeux de mots, en particulier sur les sigles — où d'autre voit-on un LAMA s'écrier GEMECOD avant de prendre un ASPRO ? — j'ai également (re)-trouvé avec grand plaisir mon « frère de thèse » Nathaël, et pu discuter, de mathématiques et d'autres choses, avec de nombreux collègues.

Parmi ces collègues, Matthieu Fradelizi a eu la mauvaise idée de prendre un café alors que j'imprimais un brouillon de ce rapport, puis d'accepter de relire le dit brouillon. Il doit donc être tenu responsable de toutes les coquilles qui auraient survécu à son œil expert.

★ ★ ★

À la technique, les fidèles vim, $\text{\LaTeX}$ et TikZ ont récemment accueilli les petits git et Julia, au désespoir croissant de svn et Scilab, dont je n'oublie pas les bons et loyaux services.

★ ★ ★

Enfin, pour me soutenir dans le processus en temps long qu'est la recherche en mathématique, processus que l'on pourrait en première approximation modéliser par une alternance entre les deux états « ça y est mon calcul marche » / « ah non en fait il marche pas »⁷, la solide équipe familiale, au premier rang de laquelle figure ma géographe préférée, a toujours été présente. Elle s'est même récemment enrichie d'une jeune collaboratrice, qui compense son manque d'expérience par une extrême curiosité et une bonne humeur communicative. Un grand merci à tout ce petit monde...

Pierre-André Zitt, Champs-sur-Marne, novembre 2014

5. « Ah, le pays de la moutarde ? » / « Ah, le pays du bon vin... »

6. « Ah, tu es chez Mickey ? »

7. L'estimation des taux de sauts et de la mesure invariante est laissée en exercice.

Chapitre 1

Introduction

1.1 Le comportement en temps long de processus de Markov

a) Quelques exemples

Le thème central de mes travaux est l'étude du comportement en temps long de processus aléatoires. Considérons pour fixer les idées un processus de Markov $(X_t)_{t \in \mathbb{R}_+}$ vivant dans $\mathbb{R}^d$, et notons $\mathbb{P}_x[\cdot]$, respectivement $\mathbb{P}_\mu[\cdot]$, la loi du processus partant de x (resp. de μ), et $\mathbb{E}_x, \mathbb{E}_\mu$ les espérances correspondantes. La loi d'un tel processus peut être décrite (au moins formellement) par son générateur L , agissant sur un ensemble de fonctions $f : \mathbb{R}^d \rightarrow \mathbb{R}$, de telle manière que

$$\mathbb{E}_x[f(X_{t+dt})] = f(x) + Lf(x) \cdot dt + o(dt).$$

Rappelons qu'une mesure μ est **invariante** pour le processus si, sous $\mathbb{P}_\mu$, X_t suit la loi μ , et qu'elle est dite **réversible** si sous $\mathbb{P}_\mu$, (X_0, X_t) et (X_t, X_0) ont la même loi. On utilisera la notation de semi-groupe νP_t pour désigner la loi de X_t quand X_0 suit la loi ν . Donnons quelques exemples :

Le processus d'Ornstein–Uhlenbeck. C'est le processus gaussien sur $\mathbb{R}$ défini par le générateur

$$Lf(x) = f''(x) - xf'(x).$$

Ce processus résout l'équation différentielle stochastique

$$dX_t = \sqrt{2}dB_t - X_t dt.$$

Sa dynamique met en compétition deux ingrédients :

- un bruit brownien qui tend à faire visiter tout l'espace ;
- une « dérive » (*drift*) qui ramène exponentiellement vite vers 0, la solution de l'équation ordinaire sans le terme brownien étant $x_t = x_0 e^{-t}$.

Ce processus, illustré dans la figure 1.1, est suffisamment simple pour nous permettre d'illustrer toutes les approches développées plus bas : il est réversible par rapport à la mesure gaussienne standard, sa décomposition spectrale est entièrement connue, il vérifie les inégalités fonctionnelles usuelles avec des constantes optimales souvent explicites, on peut étudier sa convergence par un couplage simple...

Des particules en interaction. Plaçons maintenant n particules dans l'espace $\mathbb{R}^3$. Supposons que chaque particule subit trois influences :

- un bruit brownien,
- une dérive qui la ramène en 0,
- une force répulsive électrique exercée par les autres particules, d'intensité inversement proportionnelle au carré de la distance entre particules.

Plus précisément, en notant X_t^i la position de la i^e particule au temps t , le système suit le système d'équations différentielles stochastiques

$$\forall i \in 1, \dots, n \quad dX_t^i = \sqrt{2}dB_t^i - X_t^i dt + \sum_{j \neq i} \frac{(X_t^i - X_t^j)}{|X_t^i - X_t^j|^3} dt.$$

Sans la troisième force, chaque particule évolue indépendamment suivant une dynamique d'Ornstein–Uhlenbeck ; la troisième force pousse les particules à se repousser mutuellement.

Le processus TCP. Revenons désormais en dimension 1 et considérons le processus sur $\mathbb{R}_+$ de générateur

$$Lf(x) = f'(x) + x(f(x/2) - f(x)).$$

Ce processus, issu de la modélisation du protocole informatique TCP¹, est bien différent des précédents : s'il y a toujours une dérive, qui emmène le processus à vitesse 1 vers $+\infty$, la diffusion a été remplacée par des **sauts** : à un taux proportionnel à sa position x , le processus saute en $x/2$.

Un exemple de flots modulés. Dans $\mathbb{R}^2$, on se donne deux matrices fixées A_0 et A_1 . À chacune de ces matrices est associé le flot Φ^i de l'équation linéaire (déterministe) :

$$dx_t = A_i x_t dt.$$

On construit maintenant un processus aléatoire en alternant les deux flots Φ^0 et Φ^1 pendant des temps exponentiels : un exemple est illustré dans la figure 1.2.

Questions. Dans tous ces exemples on cherche à savoir **ce que fait le processus quand t tend vers l'infini**. Comme dans le cas classique des chaînes de Markov, il y a deux comportements extrêmes :

- le processus est absorbé par un sous-ensemble et n'en sort plus ; parfois il converge même presque sûrement vers un point. C'est le type de comportement recherché si le processus décrit un algorithme stochastique² !
- le processus continue à visiter tout l'espace, en favorisant certaines régions.

Dans la plupart des exemples que nous considérerons, il existe une unique mesure de probabilité invariante pour le processus et l'on est dans le deuxième cas. La question principale est alors de savoir **en quel sens et à quelle vitesse** le processus X_t **converge vers cette mesure limite**.

1. Le sigle TCP signifie *Transmission Control Protocol*. Nous revenons plus en détail sur cette modélisation dans la section 4.1.a.

2. Nous verrons plus bas que ces algorithmes sont souvent des processus inhomogènes en temps, qui « ralentissent » en temps grand, ce qui favorise la convergence.

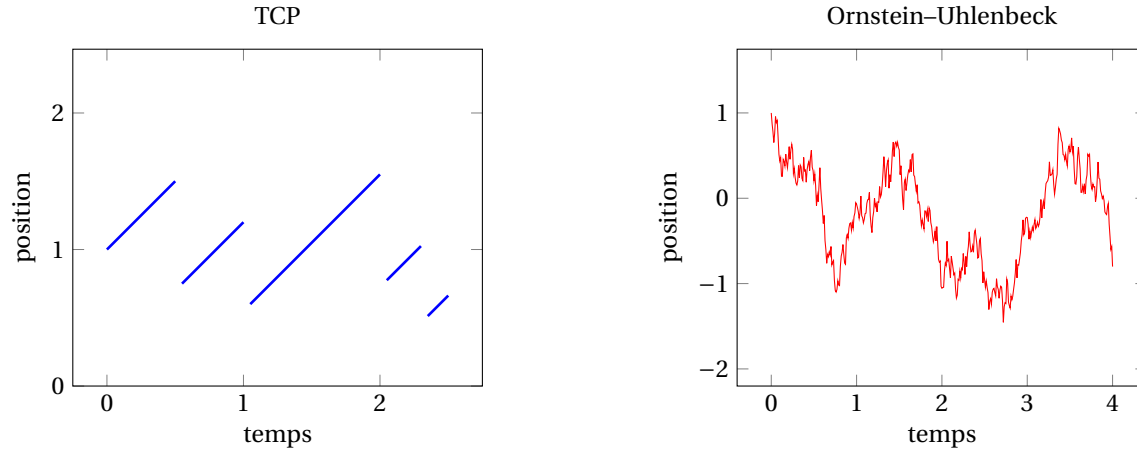


FIGURE 1.1 – Deux exemples de processus

À gauche, une trajectoire du processus TCP : la position augmente linéairement, et en des temps aléatoires elle est divisée par deux. À droite, une trajectoire du processus d'Ornstein-Uhlenbeck, très irrégulière et bruitée.

b) Convergence en loi et distances entre probabilité

Dans le cas absorbant, typique des algorithmes stochastiques, on peut chercher dans un premier temps à montrer la **convergence presque-sûre** du processus vers un point particulier. Dans ce cas, la vitesse de convergence peut être établie en normalisant correctement l'erreur pour obtenir un analogue du théorème limite central : cette approche sera illustrée dans la section 3.2 pour une approximation de médiane dans les espaces de Hilbert.

Dans le cas où le processus continue à visiter tout l'espace, et admet une probabilité invariante, on étudie la convergence étroite de la loi de X_t vers la mesure invariante. Toute étude de vitesse demande une quantification de la convergence par l'introduction d'une **distance** entre mesures de probabilité. Les distances classiques que nous utiliserons se définissent naturellement en termes de **couplage** de lois.

Définition 1.1.1 (Couplage). Soit ν et $\tilde{\nu}$ deux mesures de probabilité sur un espace E .

Un couplage de ν et $\tilde{\nu}$ est une mesure de probabilité π sur $E \times E$ de marginales ν et $\tilde{\nu}$: pour toute parties $A, \tilde{A}$ mesurables,

$$\pi(A \times E) = \nu(A), \quad \pi(E \times \tilde{A}) = \tilde{\nu}(\tilde{A}).$$

De façon équivalente un couplage est la loi d'un couple de variables aléatoires $(X, \tilde{X})$ tel que X suit la loi ν et $\tilde{X}$ la loi $\tilde{\nu}$.

À partir de cette notion on définit une première distance.

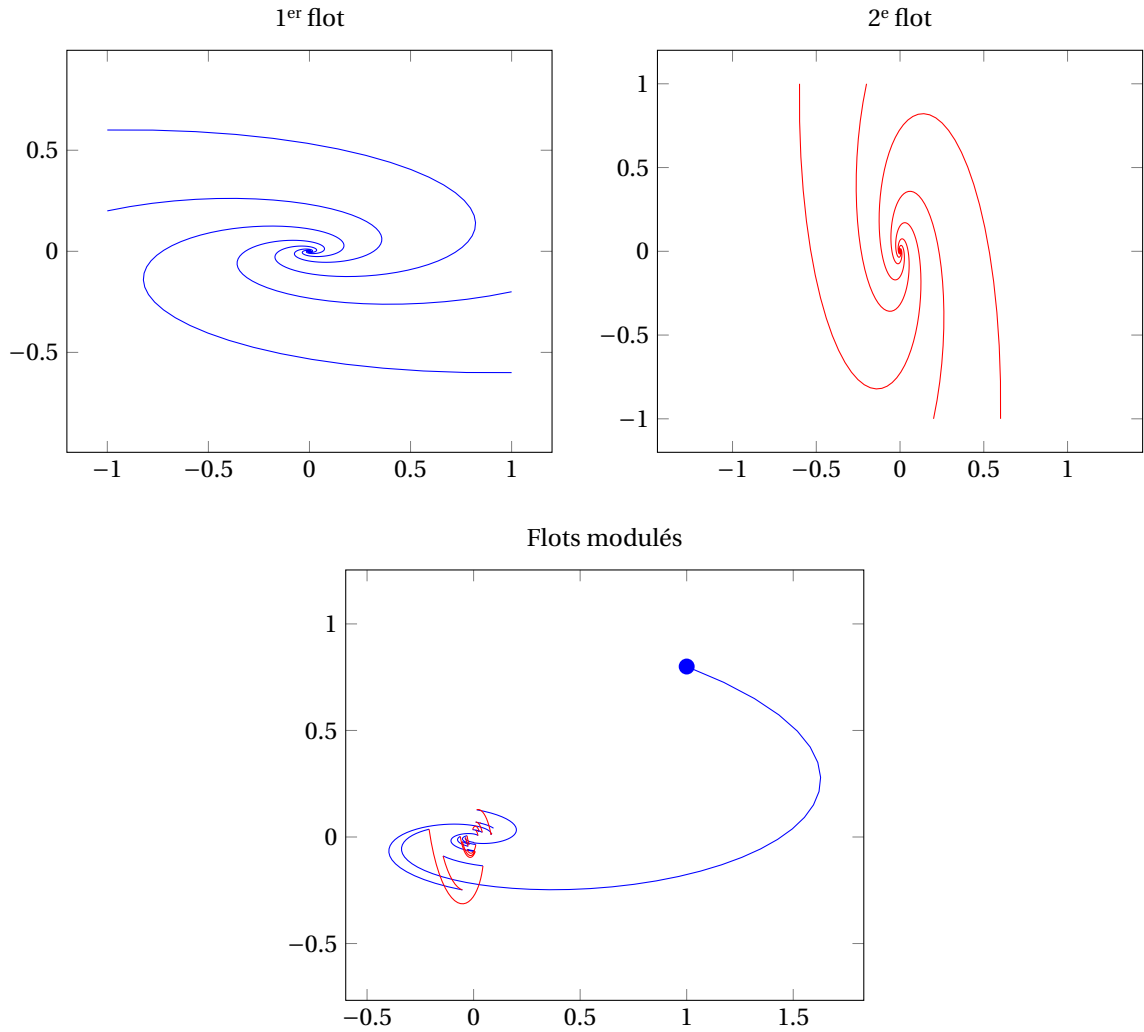


FIGURE 1.2 – Une trajectoire du processus de « flots modulés ».

En haut à gauche (resp. à droite), on a représenté plusieurs trajectoires du flot déterministe $dx_t = Ax_t dt$, ici avec $A = A_0 = \begin{bmatrix} -1 & 6 \\ -2/3 & -1 \end{bmatrix}$ (resp. $A = A_1 = \begin{bmatrix} -1 & -2/3 \\ 6 & -1 \end{bmatrix}$). Pour ces valeurs, les trajectoires convergent « en spirale » vers 0. En bas, le processus suit alternativement les deux flots, pendant des temps aléatoires de loi exponentielle.

Définition 1.1.2 (Distance en variation totale). *La distance en variation totale entre ν et $\tilde{\nu}$ est donnée par :*

$$\begin{aligned}\|\nu - \tilde{\nu}\|_{TV} &= \inf \{ \mathbb{P}(X \neq \tilde{X}) : X \sim \nu, \tilde{X} \sim \tilde{\nu} \} \\ &= \sup \left\{ \nu(A) - \tilde{\nu}(A) : A \in \mathcal{B}(\mathbb{R}^d) \right\} \\ &= \frac{1}{2} \sup \left\{ \int f d\nu - \int f d\tilde{\nu} : f \text{ bornée par } 1 \right\}.\end{aligned}$$

Remarque 1.1.3 (Variation totale et densité). Si ν et $\tilde{\nu}$ admettent les densités g et $\tilde{g}$ par rapport à une mesure de référence μ , la distance $\|\nu - \tilde{\nu}\|_{TV}$ se réécrit $(1/2) \int |g - \tilde{g}| d\mu$. En particulier si $\tilde{\nu}$ est absolument continue par rapport à ν , de densité f , alors

$$\|\tilde{\nu} - \nu\|_{TV} = \frac{1}{2} \int |f - 1| d\nu. \quad (1.1)$$

La distance suivante n'est définie que sur un sous-ensemble des mesures de probabilité.

Définition 1.1.4. *Pour tout $p \geq 1$, soit $\mathcal{P}_p(\mathbb{R}^d)$ l'ensemble des mesures de probabilité sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ admettant un moment d'ordre p : $\nu \in \mathcal{P}_p(\mathbb{R}^d)$ si*

$$\int_{\mathbb{R}^d} |x|^p \nu(dx) < +\infty.$$

Définition 1.1.5 (Distance de Wasserstein). *La distance de Wasserstein (d'ordre 1) est définie sur $\mathcal{P}_1(\mathbb{R}^d)$ par :*

$$\begin{aligned}W_1(\nu, \tilde{\nu}) &= \inf \{ \mathbb{E}(|X - \tilde{X}|) : X \sim \nu, \tilde{X} \sim \tilde{\nu} \} \\ &= \sup \left\{ \int f d\nu - \int f d\tilde{\nu} : f \text{ 1-Lipschitz} \right\}.\end{aligned}$$

Remarque 1.1.6. La formule analogue pour la distance de Wasserstein d'ordre p , W_p , est :

$$W_p(\nu, \tilde{\nu}) = \inf \left\{ \left[\mathbb{E}(|X - \tilde{X}|^p) \right]^{1/p} : X \sim \nu, \tilde{X} \sim \tilde{\nu} \right\}.$$

Elle admet également une formulation duale (voir [Vil03]).

Remarque 1.1.7. La distance en variation totale est particulièrement adaptée au cas discret. Dans un espace continu elle peut être aveugle à certaines convergences : le processus déterministe $x_t = e^{-t}$ converge vers 0 en distance de Wasserstein (de tout ordre), mais pas en variation totale !

Par la définition même, on peut obtenir des bornes explicites sur les distances entre mesures dès lors que l'on sait **construire des bons couplages** : pour contrôler la distance de Wasserstein il faut « **rapprocher** » les variables, en construisant des couplages $(X, \tilde{X})$ où X et $\tilde{X}$ sont souvent proches ; pour contrôler la distance en variation totale il faut réussir à « **coller** » les deux variables en les rendant égales suffisamment souvent.

Pour étudier la convergence des processus, on est amené à construire des couplages entre les lois du processus partant de différentes lois initiales. Rappelons que si X_0 suit la loi ν , on note νP_t la loi de X_t . Pour $\nu, \tilde{\nu}$ deux mesures initiales, on cherchera donc des couplages de νP_t et $\tilde{\nu} P_t$; en appliquant ceci à $\tilde{\nu} = \mu$, la mesure invariante, on obtiendra un couplage entre νP_t et μ , et donc des bornes sur la vitesse de convergence en distance de Wasserstein ou en variation totale.

Nous reviendrons sur ces questions dans les sections 1.2.b et 1.2.d, et verrons des exemples des deux types de couplages dans le chapitre 4.

c) Convergence du semi-groupe

Plutôt que de regarder comme précédemment la convergence de la loi de X_t vers sa mesure invariante, on peut tester cette convergence sur des fonctions en étudiant le **semi-groupe** P_t , défini par

$$P_t f : x \mapsto \mathbb{E}_x [f(X_t)].$$

Dans le cas d'une unique mesure de probabilité invariante μ , la fonction $P_t f$ a vocation à converger vers la fonction constante $\int f d\mu$: cette convergence fonctionnelle peut être comprise en différents sens, et étudiée par différents outils. La notion la plus simple est sans doute de mesurer la convergence dans l'espace de Hilbert $L^2(\mu)$: cette notion est intrinsèquement liée à l'approche spectrale présentée plus bas dans la section 1.2.a, et à l'inégalité fonctionnelle de Poincaré (voir la section 1.2.c). Il s'agit par exemple d'établir un résultat du type :

$$\forall f \in L^2, \quad \left\| P_t f - \int f d\mu \right\|_{2,\mu}^2 \leq \exp(-2ct) \left\| f - \int f d\mu \right\|_{2,\mu}^2. \quad (1.2)$$

De tels résultats de convergence sont intéressants pour eux-mêmes ; ils sont souvent établis comme conséquence d'une **inégalité fonctionnelle**, comme nous le verrons plus loin. Dans le cas d'une dynamique **réversible**, ils se traduisent directement par une **convergence sur les lois**, au sens de la section précédente. Supposons par exemple l'inégalité (1.2) vérifiée. Supposons que la loi ν de X_0 admet la densité f_0 par rapport à μ . La loi νP_t de X_t admet alors, par réversibilité, la densité $P_t f_0$ par rapport à μ . L'expression (1.1) de la distance en variation totale donne :

$$\begin{aligned} \|\nu P_t - \mu\|_{\text{TV}} &= (1/2) \int |P_t f_0 - 1| d\mu \\ &\leq (1/2) \left\| P_t f_0 - \int f_0 d\mu \right\|_{2,\mu} \\ &\leq (1/2) \exp(-2ct) \|f_0 - 1\|_{2,\mu}, \end{aligned}$$

où l'on a appliqué l'inégalité de Cauchy-Schwarz puis l'hypothèse (1.2).

Remarque 1.1.8 (Conditions sur la loi initiale). La convergence n'est obtenue que pour certaines lois initiales, ici les lois dont la densité f_0 est dans $L^2(\mu)$. On peut ici jouer sur différentes inégalités fonctionnelles, impliquant différentes convergences du semi-groupe, ce qui permet de renforcer ou d'affaiblir les hypothèses nécessaires sur f_0 . Ce point est illustré par exemple dans [Roy99, Section 3.2] et nous y reviendrons plus loin.

1.2 Des approches classiques

a) L'approche spectrale.

Il s'agit d'étudier les opérateurs L et P_t via une décomposition spectrale. Cette approche peut donner des résultats très fins... quand on réussit à l'appliquer :

- elle nécessite que l'on puisse obtenir des renseignements sur le spectre ; ceci fonctionne bien dans des modèles « miraculeux » comme le processus d'Ornstein–Uhlenbeck ou des modèles très symétriques comme des marches aléatoires sur des groupes.
- elle est particulièrement adaptée à des processus réversibles, pour lesquels les opérateurs L et P_t sont auto-adjoints dans l'espace de Hilbert $L^2(\mu)$, ce qui permet d'appliquer le théorème spectral.

On pourra trouver dans le livre [LPW09] (en particulier les chapitres 12 et 13) diverses applications de cette technique pour l'étude de chaînes de Markov, et de nombreuses références. Citons en particulier les travaux de Diaconis sur les marches sur les groupes [Dia88], ou l'expression de la loi du temps d'absorption d'une chaîne absorbante comme somme de variables exponentielles dont les paramètres sont donnés par le spectre [DF90, Mic10].

Dans la suite nous n'utiliserons pas directement cette approche. Cependant, nous verrons que la connaissance du spectre du générateur est intimement liée à plusieurs inégalités fonctionnelles également utilisées pour montrer des convergences. Le lien avec l'inégalité de Poincaré est bien connu, mais nous reviendrons dans la section 2.2 sur l'interprétation spectrale d'autres inégalités plus fortes.

b) Le couplage

Une des conditions les plus fortes pour montrer la convergence vers l'équilibre d'un processus est la « condition de Döblin » :

Définition 1.2.1 (Minoration). *Soit F une partie de l'espace d'états E . Le processus X_t vérifie la condition de minoration sur F s'il existe $\epsilon > 0$ et $t_0 > 0$, et une mesure de probabilité ν sur E , tels que pour tout $x \in F$,*

$$\mathbb{P}_x [X_{t_0} \in \cdot] \geq \epsilon \nu(\cdot). \quad (1.3)$$

Cette condition permet en particulier de construire, pour tout couple de points (x, y) dans F , un couplage « collant » π des deux lois $\delta_x P_{t_0}$ et $\delta_y P_{t_0}$ tel que $\pi(X = Y) \geq \epsilon$. Intuitivement, en un temps t_0 , on peut « coller » deux copies du processus partant de deux points différents x et y , avec une probabilité au moins ϵ . Si le couplage échoue, on peut réessayer le couplage sur la période $]t_0, 2t_0]$. En itérant on peut alors obtenir :

Théorème 1.2.2 (Chaîne de Döblin). *Si la condition (1.3) est vérifiée uniformément sur tout l'espace E , X_t converge exponentiellement vite en variation totale.*

Cette condition d'uniformité est en général beaucoup trop forte. Prenons l'exemple du processus d'Ornstein–Uhlenbeck. Comme la loi $\delta_x P_t$ est la gaussienne $\mathcal{N}(xe^{-t}, 1 - e^{-2t})$, on voit facilement que pour tout t_0 , (1.3) est vérifiée uniformément sur tout compact F , mais n'est pas vérifiée sur $\mathbb{R}$ tout entier. Nous verrons un peu plus loin une façon usuelle de contourner ce problème.

c) Inégalités fonctionnelles.

Pour étudier la convergence d'un point de vue plus analytique, de nombreuses inégalités reliant différentes quantités définies à partir de la mesure invariante et du générateur ont été introduites. Ce champ est extrêmement vaste : nous renvoyons le lecteur au livre introductif [ABC⁺00], à la monographie récente [BGL14] et à leurs très nombreuses références ; citons également sur les inégalités de transport le survol [GL10].

Présentons d'abord deux inégalités fondamentales.

Pour les définir, on notera $\text{Var}_\mu(f)$ la variance d'une fonction $f \in L^2$ et $\text{Ent}_\mu(f)$ l'entropie d'une fonction positive, définies respectivement par :

$$\begin{aligned} \text{Var}_\mu(f) &= \int \left(f - \int f d\mu \right)^2 d\mu, \\ \text{Ent}_\mu(f) &= \int f \log f d\mu - \left(\int f d\mu \right) \log \left(\int f d\mu \right). \end{aligned}$$

Définition 1.2.3 (Inégalités fonctionnelles). *Soit μ une mesure invariante pour une dynamique sur $\mathbb{R}^d$ engendrée par un générateur L . On dit que μ vérifie une inégalité de Poincaré, resp. Sobolev logarithmique, s'il existe une constante C_P (resp. C_{LS}) telle que, pour toute f suffisamment régulière,*

$$\mathrm{Var}_\mu(f) \leq C_P(\mu) \int f(-L)f d\mu, \quad (1.4)$$

$$\text{resp. } \mathrm{Ent}_\mu(f^2) \leq C_{LS}(\mu) \int f(-L)f d\mu. \quad (1.5)$$

Ces deux inégalités sont très étudiées, en particulier pour les liens qu'elles ont avec différentes propriétés de convergence du processus. Prenons le cadre simple d'une diffusion avec dérive, solution de l'équation différentielle stochastique :

$$dX_t = dB_t - \nabla V(X_t)dt,$$

où B est un mouvement brownien et V une fonction à valeurs réelles. Si V vérifie de bonnes conditions, la mesure $\mu(dx) = \exp(-V)dx$ peut se normaliser en une mesure de probabilité réversible pour le processus. Les inégalités fonctionnelles entraînent alors différents **contrôles sur la convergence du processus** vers cette mesure invariante : l'inégalité de Poincaré est par exemple équivalente à la convergence exponentielle du semi-groupe dans $L^2(\mu)$, écrite plus haut dans l'équation (1.2). Nous reviendrons sur d'autres exemples, faisant intervenir diverses inégalités fonctionnelles, dans le chapitre 2 de ce mémoire.

Cette technique, bien rodée sur les diffusions, est plus complexe à mettre en œuvre pour les processus dits hypo-elliptiques, où le bruit brownien n'influe pas directement sur toutes les composantes (voir le mémoire [Vil09] et ses références). Le manque de bruit dans les processus déterministes par morceaux complique là aussi cette approche : nous reviendrons sur ce point dans la section 4.1.

d) Fonctions de Lyapunov.

Cadre déterministe

Commençons par un exemple élémentaire : dans l'espace $\mathbb{R}^d$, on se donne une matrice carrée A et on suit le flot linéaire

$$dx_t = Ax_t dt.$$

Le comportement asymptotique des solutions est donné par le signe des parties réelles des valeurs propres de la matrice A . En particulier, si elles sont toutes strictement négatives — on dit que A est une **matrice de Hurwitz** — toutes les trajectoires convergent vers 0 à vitesse exponentielle. Une façon de voir cette convergence est d'introduire une norme adaptée : on peut montrer qu'il existe une unique matrice symétrique définie positive P solution de l'**équation de Lyapunov**

$$A^T P + P A = -I.$$

En notant $\|x\|_P$ la norme sur $\mathbb{R}^d$ induite par la matrice P , la fonction $V(x) = x^T P x = \|x\|_P^2$ est alors une **fonction de Lyapunov** — déterministe — pour le système, c'est à dire que V décroît le long des trajectoires :

$$\begin{aligned} \frac{d}{dt}(V(x_t)) &= x_t^T A^T P x_t + x_t^T P A x_t = -\|x\|_2^2 \\ &\leq -c_A V(x_t) \end{aligned}$$

par équivalence des normes, d'où la décroissance exponentielle

$$V(x_t) \leq \exp(-c_A t) V(x_0).$$

Notons que la norme euclidienne converge également exponentiellement vite : toujours par équivalence des normes,

$$\|x_t\|_2 \leq C_A \exp(-c_A t/2) \|x_0\|_2.$$

Cependant, contrairement à $V(x_t)$, rien ne dit que $t \mapsto \|x_t\|_2$ est une fonction décroissante : on peut le constater sur les trajectoires illustrées en haut de la figure 1.2, p. 10.

Cadre aléatoire

Pour un processus aléatoire X_t , il n'y a pas en général de fonction V intéressante telle que $V(X_t)$ décroisse le long des trajectoires. On peut alors demander une certaine décroissance en moyenne en imposant que $P_t V(x)$ soit plus petit que $V(x)$ pour t petit. En passant à la limite on en vient à considérer des fonctions V vérifiant :

$$\forall x, \quad LV(x) \leq 0.$$

Cette condition est à la fois trop forte (par la condition d'uniformité en x) et trop faible : la seule contrainte de signe ne donne pas facilement de résultats quantitatifs.

On en vient à la définition usuelle suivante :

Définition 1.2.4 (Fonction de Lyapunov). *Soit K un compact, et $\alpha > 0$, $C \geq 0$ deux constantes. La fonction V est appelée fonction de Lyapunov pour (X_t) si pour tout x , $V(x) \geq 1$ et si la « condition de dérive » suivante est vérifiée :*

$$\forall x, \quad LV(x) \leq -\alpha V(x) + C \mathbb{1}_K(x). \quad (1.6)$$

On peut montrer facilement que l'existence d'une telle fonction permet de **contrôler les temps de retour** dans le compact K .

Théorème 1.2.5 (Moments des temps de retour). *Si V est une fonction de Lyapunov, le temps d'atteinte τ_K du compact K a des moments exponentiels : en particulier,*

$$\mathbb{E}_x [\exp(\alpha \tau_K)] \leq V(x).$$

Informellement, l'équation (1.6) assure que $V(X_t)$ décroît exponentiellement vite (à taux α) tant que X_t n'a pas atteint K . Comme V est plus grand que 1, cette décroissance ne peut pas durer éternellement, donc X_t doit toucher K rapidement.

Ce contrôle des temps de retour est au cœur de la **méthode de Foster–Lyapunov**, étudiée en grand détail en particulier par S. Meyn et R. Tweedie [MT93]. Le point de départ est la remarque faite plus haut sur la condition de minoration (1.3), qui n'est en général vérifiée que si l'on restreint x à un sous-ensemble F de E . Sous cette condition affaiblie, on ne peut *a priori* coupler deux processus (X_t) , (X'_t) au temps t_0 que s'ils partent tous deux dans F . L'idée naturelle est alors de construire un couplage en deux temps : on construit un premier couplage qui **amène les deux processus dans F** en un certain temps T , puis on utilise la condition de minoration pour **tenter de les coller** au temps $T + t_0$. L'existence d'une fonction de Lyapunov permet de contrôler suffisamment le temps T pour obtenir par exemple le résultat suivant :

Théorème 1.2.6 ([DMT95]). *Si la condition de minoration (1.3) est vérifiée sur un ensemble F , et si il existe une fonction de Lyapunov $V \geq 1$ telle que*

$$LV \leq -KV + C\mathbb{1}_F, \quad (1.7)$$

alors X_t converge exponentiellement vite en variation totale.

Remarque 1.2.7. Cette approche est souvent utilisée en temps discret. Dans ce cadre, les ensembles vérifiant l'analogue de la condition de minoration sont appelés « *small sets* ».

On pourra consulter [Ros02] pour une preuve courte du résultat en temps discret, avec de nombreuses références.

Remarque 1.2.8. Un autre point de vue également très fructueux est de faire apparaître des **processus de renouvellement**. Pour des chaînes à espace d'états discret, les trajectoires peuvent être découpées en boucles indépendantes autour d'un point base x , et les temps d'atteinte successifs de x forment naturellement un processus de renouvellement. Dans le cas plus général, la condition de minoration (1.3) peut être utilisée pour construire un processus enrichi, qui se renouvelle avec probabilité ε quand X_t atteint l'ensemble F .

Remarque 1.2.9. Toutes ces idées ont été introduites dans les années 70 en particulier par Griffeath, Athreya et Ney, et Nummelin [Gri78, AN78, Num78]. Elles ont été depuis continuellement étudiées et développées : on pourra consulter [Num84, Lin92, MT93].

La grande force de cette idée est sa généralité : elle s'applique dans de très nombreux cadres. C'est également sa faiblesse : les arguments sont très généraux et l'application à des cas concrets, si elle prouve la convergence exponentielle, ne donne que des estimations très mauvaises de la vitesse (voir [RR96]). Un des objectifs des travaux présentés dans le chapitre 4 est d'améliorer ces estimations dans des cas particuliers de processus déterministes par morceaux.

Les **fonctions de Lyapunov** (au sens de la définition 1.2.4 ou d'une variante d'icelle) sont en fait un **outil extrêmement versatile** : en plus de l'application ci-dessus au **contrôle des temps de retour**, elles peuvent être utilisées pour établir directement des **inégalités fonctionnelles** (voir la section 2.3 plus bas) ; elles sont également très utilisées en « approximation stochastique » (l'étude de processus de type Robbins–Monro ou Kiefer–Wolfowitz, sur lesquels nous revenons plus loin dans la section 3.2), où l'on perturbe une fonction de Lyapunov déterministe, adaptée à un processus « moyen », pour construire une fonction de Lyapunov aléatoire et montrer des **propriétés de stabilité** (voir le livre [KY03] pour plus de détails).

1.3 Plan du mémoire

Après ce tour d'horizon de différents modèles et outils d'étude, présentons l'organisation du mémoire.

Le chapitre 2 évoque les résultats reliés aux **inégalités fonctionnelles** : l'utilisation d'inégalités faibles pour l'étude de convergence d'un modèle de mécanique statistique ([12]), l'étude effectuée dans [11] du lien entre spectre essentiel d'un opérateur et « super-inégalités » de Poincaré, et les nombreux aller-retours entre inégalités fonctionnelles, fonctions de Lyapunov et contrôle de temps de retour dans des compacts ([10]).

Dans le chapitre 3, je rappelle brièvement les résultats de [13] obtenus pendant ma thèse concernant le recuit simulé, et présente les deux articles [9, 8] qui étudient la convergence d'un **algorithme stochastique** de type Robbins–Monro, en dimension infinie.

Le chapitre 4 présente les travaux [7, 2, 4, 6] sur les **processus déterministes par morceaux**.

J'ai enfin rassemblé dans un dernier chapitre la présentation de deux travaux dans des directions relativement différentes : dans [3] nous établissons un principe de grandes déviations pour un système de particules en interactions répulsives ; [1] s'intéresse à un modèle de biologie, adapté du modèle classique de Wright-Fisher.

Chapitre 2

Autour des inégalités fonctionnelles

2.1 Inégalités faibles pour un modèle de mécanique statistique

Nous présentons brièvement dans cette section les résultats contenus dans [12]. On pourra consulter le livre [Roy99] et ses références pour de plus amples détails sur l'origine du modèle et le formalisme de la mécanique statistique.

Dans la lignée de nombreux travaux [Zeg96, BH99, Yos01, Led01], on s'intéresse à un modèle de spins sur $\mathbb{R}^{\mathbb{Z}^d}$: en chaque point i du réseau $\mathbb{Z}^d$ se trouve un spin $x_i \in \mathbb{R}$, et on étudie le modèle défini par le hamiltonien formel suivant :

$$H(\mathbf{x}) = \sum_{i \in \mathbb{Z}^d} V(x_i) + J \sum_{i \sim j} (x_i - x_j)^2,$$

où V est une fonction réelle (l'auto-interaction), J est une constante et $i \sim j$ signifie que i et j sont voisins dans $\mathbb{Z}^d$. Le premier terme, dit d'auto-interaction, est faible quand $V(x_i)$ l'est, le second (si l'on suppose J positif) est petit si les valeurs des spins en deux sites voisins sont proches. Les exemples typiques pour V sont le cas quadratique $V(x) = ax^2$ (qui correspond à un modèle gaussien), et le cas d'un double puits de potentiel $V(x) = ax^4 - bx^2$ (où chaque spin a deux états privilégiés, l'un positif et l'autre négatif). Le hamiltonien¹ d'une configuration $\mathbf{x} \in (\mathbb{R})^{\mathbb{Z}^d}$ est donc petit si $V(x_i)$ est petit pour chaque i , et si les x_i et x_j sont proches quand i et j sont voisins.

À partir du H formel, on définit classiquement des mesures de Gibbs, en volume fini : pour un ensemble de sites Λ fini et une « condition aux bords » $\mathbf{z} \in \mathbb{R}^{\mathbb{Z}^d}$ fixée, la mesure de Gibbs $\mu_{\Lambda, \mathbf{z}}$ n'est autre que la mesure d'équilibre de la dynamique diffusive suivante sur $\mathbb{R}^{\Lambda}$:

$$\forall i \in \Lambda, \quad dX^i = dB_t^i - \nabla_i H(X \cdot \mathbf{z})$$

où $(X \cdot \mathbf{z})$ est la configuration qui coïncide avec X pour $i \in \Lambda$, et avec $\mathbf{z}$ au dehors.

On définit alors les mesures de Gibbs en volume infini par la condition de compatibilité de Dobrushin, Lanford et Ruelle : une mesure μ sur $\mathbb{R}^{\mathbb{Z}^d}$ est dite de Gibbs si pour tout ensemble fini Λ , et toute fonction dans un ensemble suffisamment grand,

$$\int f(\mathbf{x}) d\mu(\mathbf{x}) = \int \int f(\mathbf{x} \cdot \mathbf{z}) d\mu_{\Lambda, \mathbf{z}}(\mathbf{x}) d\mu(\mathbf{z}).$$

1. sans prendre en compte les problèmes de sommabilité

Une des questions centrales est alors de déterminer l'ensemble des mesures de Gibbs en volume infini, et en premier lieu de montrer l'éventuelle **unicité de la mesure de Gibbs**.

Une façon de montrer l'unicité est de supposer que les mesures en volume fini vérifient des inégalités fonctionnelles, avec une certaine uniformité.

Définition 2.1.1 (Inégalité de Beckner). *Pour $a \in (0, 1)$, on dit que μ vérifie une inégalité de Beckner généralisée de constante C si, pour toute f ,*

$$GBI(a) : \sup_{p \in]1, 2[} \frac{\int f^2 d\mu - (\int f^p d\mu)^{2/p}}{(2-p)^a} \leq C_a \int |\nabla f|^2 d\mu. \quad (2.1)$$

Cette inégalité a été introduite dans cette formulation par R. Latała and K. Oleszkiewicz dans [LO00]. En faisant tendre a vers 0 ou vers 1, on peut montrer ([LO00, BCR06]) que l'on récupère respectivement une inégalité de Poincaré ou de Sobolev logarithmique. De plus ces inégalités sont ordonnées, dans le sens où $GBI(a)$ implique $GBI(a')$, dès que $a > a'$. Ces inégalités « interpolent » donc entre les deux inégalités classiques de Poincaré et Sobolev logarithmique.

Théorème 2.1.2 (Un critère d'unicité de la mesure de Gibbs). *[12] Supposons que l'auto-interaction V est de type double puits : $V(x) = ax^4 - bx^2$ pour a et b positifs.*

Supposons que les mesures de Gibbs μ_n , dans les boîtes $[-n, n]^d$, avec condition aux bords nulle, vérifient l'une des deux conditions suivantes :

— *une inégalité de Sobolev logarithmique avec une constante C_n telle que :*

$$C_n \leq \frac{Cn}{\log(n)}$$

— *une inégalité de Beckner généralisée, pour un réel a suffisamment proche de 1, et une constante C_a indépendante de n .*

Alors la mesure de Gibbs en volume infini est unique.

Remarque 2.1.3. Le résultat est en réalité valable pour des V plus généraux, pour lesquels il faut s'assurer que le modèle est bien défini, via une hypothèse sur J . Nous renvoyons à l'article original pour les détails.

Le théorème était déjà connu pour une inégalité de Sobolev logarithmique *uniforme* (i.e. C_n indépendante de n , cf. [Roy99]). Nous améliorons donc le résultat de deux manières, en autorisant les constantes à croître lentement avec la taille de la boîte ou en permettant des inégalités plus faibles que l'inégalité de Sobolev logarithmique.

La preuve du résultat s'inspire de celle de Royer. Elle utilise une **approche dynamique**. Nous avons déjà vu que les mesures en volume fini apparaissent comme équilibre de dynamiques explicites $X_t^{\Lambda, z}$. On peut également introduire directement une dynamique globale X_t sur $(\mathbb{R})^{\mathbb{Z}^d}$, et montrer que cette dynamique est proche de celle de $X_t^{\Lambda, 0}$ pour des temps t de l'ordre du diamètre de la boîte Λ . Les hypothèses d'inégalités fonctionnelles permettent alors de montrer que ce temps suffit pour que le processus soit assez proche de sa mesure invariante $\mu_{\Lambda, 0}$. On peut alors reproduire un argument de [Roy99] pour établir l'unicité de la mesure de Gibbs. Un des points cruciaux est de jouer sur le lien, mentionné dans la remarque 1.1.8 de l'introduction, entre la force de l'inégalité fonctionnelle et les hypothèses correspondantes sur les mesures initiales.

Au vu de ce résultat, on peut naturellement se poser la question suivante, à ma connaissance non-résolue :

Problème ouvert 2.1.4. *Si les mesures de Gibbs μ_n (pour des conditions aux bords nulles) vérifient une inégalité de Poincaré uniforme en n , a-t-on nécessairement unicité de la mesure de Gibbs ?*

Si certains résultats, comme ceux de [Yos99] permettant de retrouver une inégalité de Sobolev logarithmique uniforme à partir d'une inégalité de Poincaré uniforme, semblent plaider pour une réponse positive, la dernière section de [12] présente l'étude d'un modèle-jouet introduit dans [BM02] pour lequel il y a transition de phase mais uniformité de la constante de Poincaré.

2.2 Spectre essentiel et « super-inégalités de Poincaré »

Nous avons rappelé plus haut les liens existant entre inégalités fonctionnelles et propriétés de convergence des semi-groupes. Il est également intéressant d'essayer d'établir des liens avec des propriétés spectrales de l'opérateur qui engendre le semi-groupe.

Le résultat classique dans ce sens concerne l'inégalité de Poincaré. Il est bien connu que cette inégalité est vraie dans le cas des diffusions avec dérive si et seulement si 0 est une valeur propre isolée pour l'opérateur générateur :

$$L = \frac{1}{2}\Delta - \nabla V \nabla.$$

Dans ce cas, propriétés spectrales, inégalité fonctionnelle et convergence (ici au sens L^2) sont équivalentes.

Il est assez naturel de chercher des **interprétations spectrales** pour **d'autres inégalités fonctionnelles**.

On dit que μ vérifie une « **super-inégalité de Poincaré** » si l'on a :

$$\forall r > 0, \exists \beta(r) \in (0, \infty), \forall f, \quad \int f^2 d\mu \leq r \int |\nabla f|^2 d\mu + \beta(r) \left(\int |f| d\mu \right)^2. \quad (2.2)$$

Si ceci n'est valable que pour $r > r_0$, on parlera d'inégalité partielle.

Cette inégalité, introduite par F.-Y. Wang dans [Wan00a] et développée en particulier dans [Wan00b, GW02], est plus forte que l'inégalité de Poincaré, et permet d'établir des bornes sur la décroissance du semi-groupe et la compacité de certaines résolvantes.

Dans l'article [11] j'ai tout d'abord introduit une variante de cette inégalité, en jouant sur la norme dans le terme de droite de (2.2) pour la remplacer par une norme d'Orlicz, intermédiaire entre les normes L^1 et L^2 . Cette variante est « qualitativement » équivalente à l'inégalité d'origine, c'est-à-dire que l'on peut passer de l'une à l'autre en jouant sur la fonction β .

Théorème 2.2.1. *Soit (Φ, Φ^*) une paire duale de fonctions de Young, telle que $\lim_{x \rightarrow \infty} \frac{\Phi^*(x)}{x^2} = \infty$. Supposons de plus que $x \mapsto \Phi^*(\sqrt{x})$ est une fonction de Young. On note $\|\cdot\|_\Phi$ et L_Φ la norme et l'espace d'Orlicz associés à la fonction Φ .*

Alors l'inégalité suivante :

$$\forall r > r_0, \quad \int f^2 d\mu \leq r \int |\nabla f|^2 d\mu + \beta(r) \|f\|_\Phi^2 \quad (2.3)$$

implique l'inégalité classique (2.2), pour $r > 8r_0$, avec une fonction $\tilde{\beta}$, exprimable explicitement en fonction de β et Φ .

Les conditions sur Φ impliquent les inclusions $L^\infty \subsetneq L_{\Phi^*} \subsetneq L^2 \subsetneq L_\Phi \subsetneq L^1$, en un sens, on a généralisé l'inégalité (2.2) à toute une famille de normes intermédiaires entre L^1 et L^2 .

Remarque 2.2.2. Le résultat tel qu'il est écrit dans [11] contient une coquille qui m'a été rapportée par Changsong Deng, ce qui rend incorrecte la preuve directe de l'inégalité de Sobolev logarithmique adimensionnelle de la dernière partie. Une telle preuve demanderait probablement une étude spectrale encore plus fine.

J'ai ensuite utilisé ce fait pour fournir une nouvelle preuve d'un résultat de F.-Y. Wang. Rappelons tout d'abord une définition.

Définition 2.2.3 (Spectre essentiel). *Si L est un opérateur auto-adjoint, le spectre de L est l'ensemble des λ pour lesquels l'opérateur $\lambda I - L$ n'admet pas d'inverse. On distingue deux cas :*

- $\lambda I - L$ n'est pas injectif, on dit que λ est valeur propre.
- $\lambda I - L$ est injectif d'image dense, mais son inverse n'est pas borné.

Le spectre discret est l'ensemble des valeurs propres isolées dont l'espace propre est de dimension finie ; le complémentaire du spectre discret est appelé spectre essentiel.

Le spectre essentiel est également la partie du spectre stable par perturbation compacte :

$$\lambda \in \sigma_{\text{ess}}(L) \iff \forall K \text{ opérateur compact, } \lambda \in \sigma(L + K).$$

On dispose alors d'un lien entre super-inégalité de Poincaré et bas du spectre essentiel.

Théorème 2.2.4 ([Wan00a, GW02]). *Si le spectre essentiel de $(-L)$ est inclus dans $[1/r_0, \infty)$ pour un $r_0 \geq 0$, alors $(-L)$ vérifie une super-inégalité de Poincaré (2.2), pour une fonction $r \mapsto \beta(r)$, définie pour tout $r > 8r_0$. En particulier si le spectre essentiel est vide ($r_0 = 0$), $-L$ vérifie une super-inégalité de Poincaré.*

L'utilisation de normes d'Orlicz adaptées permet une preuve rapide de ce résultat.

Disons quelques mots de la technique utilisée pour effectuer le changement de normes du théorème 2.2.1. Il s'agit de traduire l'inégalité fonctionnelle comme une inégalité sur les ensembles. Plus précisément, suivant les idées développées dans [BR03, BCR06] on introduit une notion de capacité des ensembles :

Définition 2.2.5 (Capacité). *Pour tout ensemble A tel que $\mu(A) \leq 1/2$, la capacité de A est définie par :*

$$\text{Cap}_\mu(A) = \inf \left\{ \int |\nabla f|^2 d\mu, \mathbb{1}_A \leq f \leq 1, \mu(\text{Supp}(A)) \leq 1/2 \right\}. \quad (2.4)$$

On dira que μ vérifie l'inégalité capacité-mesure pour (κ, C_κ) si

$$\forall A, \mu(A) \geq \kappa \implies \text{Cap}_\mu(A) \geq C_\kappa \mu(A). \quad (2.5)$$

Nous renvoyons à [BGL14, Chap. 8] pour une présentation détaillée de cette notion et de nombreuses références. La plupart des inégalités fonctionnelles sont reliées à des inégalités de type capacité-mesure : par exemple, si C_{MC} est la constante optimale dans l'inégalité ci-dessus pour $\kappa = 1/2$, et C_P est la constante optimale pour l'inégalité de Poincaré, alors ([BCR06, Proposition 13]) :

$$C_{MC}^{-1} \leq C \leq 4C_{MC}^{-1}. \quad (2.6)$$

Pour les super-inégalités de Poincaré, j'établis dans [11] le résultat suivant :

Théorème 2.2.6. *Si le critère capacité mesure (κ, C_κ) est vérifié, et si $\kappa \mapsto \frac{\kappa}{C_\kappa}$ est décroissante, alors l'inégalité (2.2) est vérifiée avec :*

$$\beta(r) = \frac{1}{\inf\{\kappa, C_\kappa \geq 8/r\}}. \quad (2.7)$$

quand cette quantité est finie. En particulier si C_κ tend vers l'infini, l'inégalité est valable pour tout $r > 0$.

2.3 Temps de retour, fonctions de Lyapunov et inégalités fonctionnelles

En collaboration avec P. Cattiaux et A. Guillin, nous éclaircissons dans l'article [10] les liens existant entre inégalités fonctionnelles, **temps de retour** dans des compacts et **fonctions de Lyapunov**, dans un cadre continu. Cette étude s'inscrit dans la lignée de nombreux travaux récents, utilisant des variantes de la définition 1.2.4 de fonction de Lyapunov, pour établir directement diverses inégalités fonctionnelles ([BBCG08, CGWW09, CGGR10] et l'article de synthèse [CG09]); le lien avec les moments des temps de retour était déjà présent (qualitativement) dans [CK83] et a été également traité dans le cas unidimensionnel dans [LLS11] (voir aussi [LLL11]).

Rappelons que, si l'on note L le générateur du processus, et si K est un compact, on dit que V est une fonction de Lyapunov si il existe des constantes c, C telles que :

$$\forall x \in \mathbb{R}^d, \begin{cases} V(x) \geq 1, \\ LV(x) \leq -cV(x) + C\mathbb{1}_K(x). \end{cases}$$

Notre principal résultat ([10, Théorème 2.3]) est, sous des hypothèses d'hypo-ellipticité, une équivalence entre :

- une **inégalité de Poincaré**,
- l'existence de **moments exponentiels pour les temps de retour** dans un ouvert borné, partant de la mesure invariante :

$$\exists U, \exists \lambda, \quad \mathbb{E}_\mu \left[e^{\lambda T_U} \right] < \infty,$$

- la même condition, partant de n'importe quel point :

$$\exists U, \exists \lambda', \forall x, \quad \mathbb{E}_x \left[e^{\lambda' T_U} \right] < \infty, \tag{2.8}$$

- l'existence d'une **fonction de Lyapunov**,
- la **convergence exponentielle** dans $L^2(\mu)$ vers la mesure invariante.

Beaucoup de ces liens étaient déjà connus, nous les avons rassemblés en simplifiant certaines preuves et en explicitant, quand c'était possible, les constantes obtenues. En particulier, dans le passage d'une inégalité de Poincaré au contrôle des temps de retour, nous obtenons une expression explicite des exposants λ, λ' intervenant dans les moments exponentiels.

On peut alors se demander ce qui se passe si l'on renforce ou affaiblit l'une ou l'autre de ces propriétés équivalentes ; les résultats que nous obtenons dans ce cas ne sont que partiels. En dimension 1 par exemple, nous avons montré sous certaines hypothèses ([10, Proposition 5.3]) une équivalence entre :

- le caractère ultraborné du semi-groupe (si f est dans L^1 , $P_t f$ est bornée),
- l'existence d'une fonction de Lyapunov bornée,
- l'existence d'un intervalle U borné et d'un $\lambda > 0$ tel que :

$$\sup_x \mathbb{E}_x \left[e^{\lambda T_U} \right] < \infty.$$

Problème ouvert 2.3.1 (Temps de retour et ultracontractivité). *Quels liens peut-on faire entre des propriétés du semi-groupe plus fortes que l'inégalité de Poincaré, et des contrôles explicites en fonction de x des moments des temps de retour de l'équation (2.8), en particulier en dimension quelconque ?*

Chapitre 3

Algorithmes stochastiques

3.1 Le recuit simulé

Cette section présente brièvement les résultats obtenus pendant ma thèse et publiés dans [13].

Le « recuit simulé » est un **algorithme stochastique d'optimisation**, introduit dans les années 1980 ([KGV83]) pour trouver des solutions approchées de problèmes complexes. Dans un cadre continu, cet algorithme s'écrit simplement sous forme d'une équation différentielle stochastique inhomogène, et a été en particulier étudié dans [HCS87, Roy89, Mic92].

Si V est la fonction à minimiser (on prendra $V : \mathbb{R}^d \rightarrow \mathbb{R}$, suffisamment régulière), une première idée pour chercher les minima est de descendre le gradient :

$$dX_t = -\nabla V(X_t) dt.$$

Ce processus a l'inconvénient de se « bloquer » dans les minima locaux. Pour empêcher cela, on force le processus à explorer son environnement en rajoutant un bruit brownien, dont on fait décroître l'intensité au cours du temps. L'équation différentielle devient une équation différentielle stochastique,

$$dX_t = \sqrt{\sigma(t)} dB_t - \nabla V(X_t) dt,$$

où la fonction σ , appelée température, est prise décroissante et de limite nulle.

La question est alors de savoir, en fonction du choix de σ , si l'algorithme trouve bien les minimas globaux.

On peut par exemple chercher à identifier une éventuelle **limite de la loi du processus** à l'instant t . Introduisons la mesure « d'équilibre instantané » :

$$\mu_\sigma(dx) = \frac{1}{Z_\sigma} \exp\left(-\frac{V(x)}{\sigma}\right),$$

où Z_σ est une constante de normalisation. Cette mesure est réversible pour la diffusion à température fixée, gouvernée par l'ÉDS

$$dX_t = \sqrt{\sigma} dB_t - \nabla V(X_t) dt.$$

Si V et $|\nabla V|^2 - \Delta V$ tendent vers l'infini, il est connu que la mesure μ_σ vérifie une inégalité de Poincaré, avec une constante C_σ qui vérifie :

$$C_\sigma \approx \exp\left(-\frac{d^*}{\sigma}\right),$$

où d^* est une constante définie explicitement à partir du potentiel V .

Ce comportement asymptotique a permis à L. Miclo de remonter, grâce à une approche par inégalités fonctionnelles inspirée par Holley, Kusuoka et Stroock [HS88, HKS89], un résultat de convergence apparu dans [HCS87].

Théorème 3.1.1 ([Mic92]). *Pour toute constante $c > d^*$, si l'on choisit un refroidissement en*

$$\sigma(t) = \frac{c}{\ln(t)},$$

alors le processus de recuit converge.

Considérons un exemple simple où, en dehors d'un compact, V est une puissance de la distance à 0 :

$$V(x) = |x|^\alpha,$$

avec $\alpha > 0$. Il est facile de constater que la constante d^* évoquée plus haut ne dépend pas du α , et donc du comportement de V à l'infini. Par conséquent les bons choix pour le refroidissement $\sigma(t)$ ne dépendent que du **comportement local de V** , près de son minimum.

Pendant, en raison de la condition sur le gradient, le résultat précédent ne peut s'appliquer que si $\alpha > 1$. J'ai pu relâcher suffisamment les hypothèses pour inclure des croissances en $|x|^\alpha$, $\alpha \in]0, 1[$. Le résultat est le suivant.

Théorème 3.1.2 ([13]). *Si V vérifie les conditions suivantes :*

- $|\nabla V| \rightarrow 0$,
- $\Delta V \leq 0$ à l'infini,
- *la structure des puits de potentiel n'est pas « pathologique »,*

alors le processus de recuit simulé trouve le minimum global pour $c > d^$, et peut se tromper si $c < d^*$.*

La preuve s'inspire de celle de Miclo ([Mic92]). On considère la quantité

$$I_t = \text{Ent}(\mathcal{L}(X_t) | \mu_t),$$

l'entropie relative de la loi du processus par rapport à la mesure invariante « instantanée ». On montre ensuite sa **convergence vers 0** en établissant une **inégalité différentielle**. Comme on sait que μ_t converge faiblement vers une masse de Dirac en le minimum global, on en déduit la convergence voulue.

La dérivation de I_t fait en particulier apparaître un terme d'énergie, $\int |\nabla f|^2 d\mu_t$. Dans notre cas celui-ci ne peut pas être contrôlé par l'inégalité de Poincaré. On la remplace donc par une **inégalité faible de Poincaré** (introduite par Röckner et Wang dans [RW01]) :

$$\forall r > 0, \text{Var}_\mu(f) \leq \alpha(r) \int |\nabla f|^2 d\mu + r \text{Osc}(f)^2,$$

où α est une fonction décroissante de r . Les difficultés viennent alors de l'estimation de la fonction α , que nous retrouvons grâce à un critère « **capacité-mesure** », similaire à celui présenté précédemment dans la définition 2.2.5 à propos des super-inégalités de Poincaré.

Il nous faut également remplacer la norme infinie (difficile à évaluer) par une norme plus forte que L^2 , mais que l'on puisse comparer à l'entropie I_t : nous utilisons ici une norme d'Orlicz bien choisie.

3.2 Une estimation de médiane en dimension infinie

Reprenons le problème précédent, où l'on cherche le minimum d'une fonction V , mais supposons maintenant que cette fonction est convexe. Si l'on a directement accès à V , on peut utiliser des algorithmes d'optimisation déterministes très efficaces et il n'y a pas besoin de « rajouter du bruit » comme l'on a vu précédemment : il n'y a pas de « pièges » dans lesquels l'algorithme peut se perdre.

Dans certaines circonstances cependant, on n'a accès qu'à une version bruitée de la fonction V : trouver son minimum est alors une tâche plus délicate. On peut cependant utiliser la même idée de « pas décroissants » que dans le recuit simulé ; on obtient l'**algorithme de Robbins-Monro**. Cet algorithme, introduit dans [RM51], et l'algorithme voisin de Kiefer et Wolfowitz ([KW52]), ont donné naissance à de nombreuses variantes et à tout un champ de recherches, l'« approximation stochastique ». Nous renvoyons aux monographies [Bor08, KY03] pour de très nombreuses références sur ces algorithmes, leurs variantes ainsi que pour un panorama des différentes techniques d'étude, dans le cadre de la dimension finie.

En collaboration avec H. Cardot et P. Cénac, nous avons utilisé dans [9, 8] l'algorithme de Robbins-Monro pour estimer une « courbe médiane ». Le point de départ est l'étude de données qui ne sont pas simplement vectorielles mais fonctionnelles. Nous avons par exemple eu accès, via la société Médiamétrie, à des courbes d'audience télévisée au cours du temps, pour différents individus. Chaque courbe indique, sur un intervalle $[0, T]$, à quel moment l'individu regarde sa télévision. On considère ces courbes comme des éléments de l'espace de Hilbert $H = L^2([0, T])$.

Pour dégager une tendance centrale dans ces courbes, l'idée naturelle est de considérer la courbe moyenne, dont la définition ne pose pas de problème. Malheureusement, cet indicateur est très sensible aux données atypiques et aux points « aberrants ». On peut lui préférer un indicateur plus robuste, qui généralise la médiane. Si X est une variable aléatoire (X étant donc dans notre cas à valeurs dans l'espace fonctionnel H de dimension infinie), on définit classiquement ([Hal48, Kem87]) la médiane (généralisée) m comme solution du problème de minimisation de la fonction $V(\alpha) = \mathbb{E} [\|X - \alpha\|]$:

$$m = \operatorname{argmin}_{\alpha \in H} \mathbb{E} [\|X - \alpha\|].$$

On connaît des conditions d'existence et d'unicité de cette médiane, et elles sont ici vérifiées.

La question est alors, en supposant que l'on observe des copies indépendantes $(X_n)_{n \in \mathbb{N}}$ de X , de retrouver la solution m du problème d'optimisation. On aimerait pour ce faire avoir accès au gradient de la fonction à minimiser, pour le descendre :

$$-\nabla V(\alpha) = \mathbb{E} \left[\frac{X - \alpha}{\|X - \alpha\|} \right].$$

C'est impossible puisque la distribution de X est supposée inconnue. Cependant, on peut considérer la quantité

$$\frac{X_n - \alpha}{\|X_n - \alpha\|}$$

comme une approximation de la quantité moyenne ∇V .

L'algorithme de Robbins-Monro consiste alors à suivre cette approximation du gradient, avec un pas de plus en plus petit. Partant d'un point initial Z_0 , on définit par récurrence :

$$Z_{n+1} = Z_n + \gamma_n \frac{X_{n+1} - Z_n}{\|X_{n+1} - Z_n\|}, \quad (3.1)$$

où γ_n est une suite de réels qui tend vers 0.

Remarque 3.2.1 (Points médians sur une variété). Si l'on remplace l'espace H par une variété de dimension finie, la recherche des minima de V est rendue beaucoup plus difficile puisque la fonction V n'est plus convexe, et peut avoir des minima locaux même dans le cas où la médiane est bien définie. On trouvera dans [AM12, AM13b, ADPY12, AM13a] l'étude d'un algorithme stochastique proche du recuit simulé pour traiter ces questions.

Pour l'algorithme brut, le bon choix du γ_n est de l'ordre de c/n pour une certaine constante c non-explicite : si γ_n est plus petit on risque de faire trop peu de mouvement : l'algorithme converge, mais pas nécessairement vers la bonne limite ; si γ_n est trop grand au contraire, la vitesse de convergence n'est pas optimale. Une astuce maintenant courante, introduite dans [PJ92], est de considérer alors un processus moyenné : on se fixe un γ_n en c/n^a , pour $a \in (0, 1)$, mais on re-lisse en considérant $\bar{Z}_n = \frac{1}{n} \sum_{i=1}^n Z_i$. Du point de vue pratique l'algorithme obtenu se montre beaucoup moins sensible au choix des paramètres.

Pour montrer de bons résultats de convergence, les hypothèses nécessaires sont de deux types. Les premières concernent la loi de X et sont relativement faibles, le point essentiel est que la distribution doit être suffisamment « étalée », en ne chargeant pas trop les petits voisinages des points. Il faut ensuite bien choisir les pas γ_n : les résultats sont valables en particulier pour le choix $\gamma_n = \frac{c}{n^a}$, pour $a \in]1/2, 1[$. Sous ces conditions on a le résultat suivant.

Théorème 3.2.2. [9] *Supposons que :*

- la loi de X n'est pas concentrée sur une droite : pour tout $v \in H$, il existe $w \in H$ orthogonal à v et tel que $\text{Var}(\langle w, X \rangle) > 0$.
- la loi de X est un mélange de deux distributions $\mu_X = \lambda \mu_{\text{diff}} + (1 - \lambda) \mu_{\text{disc}}$, telles que :
 - μ_{diff} ne charge pas trop fortement les voisinages des points : si $\mathcal{B}(A)$ est la boule $\{\alpha \in H, \|\alpha\| \leq A\}$, et Y une variable de loi μ_{diff} ,

$$\forall A, \exists C_A \in [0, \infty), \forall \alpha \in \mathcal{B}(A), \quad \mathbb{E} [\|Y - \alpha\|^{-1}] \leq C_A.$$

- $\mu_{\text{disc}} = \sum_i p_i \delta_{\alpha_i}$ est une mesure discrète, dont le support ne contient pas la médiane.
- Il existe A telle que :

$$\exists C_A \in [0, \infty), \forall h \in \mathcal{B}(A), \quad \mathbb{E} [\|X - (m + h)\|^{-2}] \leq C_A. \quad (3.2)$$

Alors :

- l'algorithme est **consistant** : Z_n converge presque sûrement vers la médiane m .
- l'algorithme moyenné est **asymptotiquement normal** : en posant $\bar{Z}_n = \frac{1}{n} \sum Z_i$, l'écart renormalisé $\sqrt{n}(\bar{Z}_n - m)$ a une limite gaussienne.

$$\sqrt{n}(\bar{Z}_n - m) \rightarrow \mathcal{N}(0, \Gamma_m^{-1} \Sigma \Gamma_m^{-1}),$$

où Γ_m et Σ sont définis par

$$\Sigma = \mathbb{E} \left[\frac{(X - m)}{\|X - m\|} \otimes \frac{(X - m)}{\|X - m\|} \right] \quad \Gamma_m = \mathbb{E} \left[\frac{1}{\|X - m\|} \left(I - \frac{(X - m) \otimes (X - m)}{\|X - m\|^2} \right) \right].$$

La philosophie de la plupart des résultats d'approximation stochastique est la suivante. On réécrit l'expression (3.1) donnant Z_{n+1} en fonction de Z_n comme somme de deux termes :

$$Z_{n+1} = (Z_n - \gamma_n \nabla V(Z_n)) + \gamma_n \left(\nabla V(Z_n) - \frac{Z_n - X_{n+1}}{\|Z_n - X_{n+1}\|} \right).$$

Dans cette décomposition,

- le premier terme est obtenu en faisant évoluer Z_n par une dynamique moyenne déterministe, ici la descente du gradient de V ;
- le second est un terme de reste, d'espérance (conditionnelle) nulle, et dont l'effet disparaît à la limite car la somme $\sum \gamma_n^2$ converge.

Tout consiste donc à comparer les trajectoires de Z_n et celles de la dynamique moyenne déterministe ; on parle généralement de « méthode de l'équation différentielle ordinaire ». Les comportements fins près du point cible peuvent dans les bons cas être étudiés en linéarisant l'ÉDO, ce qui permet d'établir la normalité asymptotique — ceci explique en particulier l'apparition de l'opérateur Σ , qui n'est autre que la Hessienne du potentiel V au point minimal m .

Nous suivons ces schémas de preuves en nous appuyant en particulier sur les articles [Duf97, Pel00], qui utilisent dans un cadre fini-dimensionnel des **techniques de martingales**.

La première étape, pour montrer la convergence de l'algorithme, s'adapte relativement aisément. Pour montrer le résultat de vitesse de convergence, on utilise une version du TLC pour des martingales dans des Hilbert ([Jak88]).

La **difficulté principale** vient du caractère **infini-dimensionnel**, qui empêche d'utiliser des arguments de compacité dans le contrôle des restes.

La forme très explicite de la fonction à minimiser permet cependant de contrôler son comportement local, près du minimum, via une décomposition spectrale.

Du point de vue pratique, cet algorithme présente deux avantages importants sur des algorithmes statiques comme celui développé dans [VZ00]. Le premier est que la mise à jour de l'estimation Z_{n+1} se fait en n'utilisant que l'estimation actuelle Z_n et le nouveau point observé. Ainsi, il n'y a pas besoin de garder en mémoire l'ensemble des données et l'algorithme peut **fonctionner « en ligne »**. De plus, cette étape de mise à jour de l'estimateur est très **simple et très rapide**, ce qui permet de traiter en un temps raisonnable une grande quantité de données. Il demande néanmoins un choix de paramètres (le c et le a qui définissent la suite de pas γ_n). Sur des données simulées, nous avons constaté en pratique que les résultats restaient bons pour des plages de paramètres assez larges.

Nous avons également appliqué l'algorithme à des données réelles fournies par Médiamétrie. À partir des courbes d'audience pour quelques milliers d'individus, enregistrées seconde par seconde, sur une durée d'une journée, l'algorithme, codé en R avec appel à des routines C, a traité les données en un temps de l'ordre d'une seconde.

Ces résultats se prêtent à des extensions dans plusieurs directions. Une idée, mise en place dans [CCM10], est d'adapter ces techniques pour faire de la classification. On peut aussi se demander ce qui se passe quand les données ne sont plus supposées indépendantes. Dans l'article [8], nous nous intéressons à l'estimation de **médianes conditionnelles** : plutôt que d'étudier uniquement une variable X , on a accès à un couple (X, Y) , et l'on cherche à estimer la médiane de X conditionnellement aux valeurs de la co-variable Y . Dans notre exemple où X est une courbe d'audience, nous disposons d'un certain nombre de données auxiliaires, comme l'âge Y du téléspectateur, et cherchons à estimer l'audience médiane pour les téléspectateurs d'un certain âge y . Pour tenir compte de cette variable nous introduisons un noyau K , et remplaçons l'algorithme précédent par :

$$Z_{n+1} = Z_n + \gamma_n \frac{1}{h_n} K\left(\frac{Y_{n+1} - y}{h_n}\right) \frac{X_{n+1} - Z_n}{\|X_{n+1} - Z_n\|},$$

où h_n est une deuxième suite déterministe. Autrement dit, l'algorithme Z_n n'évolue que lorsque Y_{n+1} est suffisamment proche de y , le poids étant contrôlé par le noyau K .

Là encore, sous de bonnes hypothèses sur la loi du couple (X, Y) et sur les deux suites γ_n et h_n , nous montrons la convergence d'un algorithme moyenné et sa normalité asymptotique ; nous renvoyons à [8] pour les énoncés précis de ces résultats et l'application aux données d'audience.

Chapitre 4

Processus de Markov déterministes par morceaux

4.1 Introduction

Les processus aléatoires sont usuellement un mélange de trois ingrédients : une dynamique déterministe (terme de dérive), un bruit diffusif (de type brownien) et une dynamique de sauts. Dans les processus déterministes par morceaux (PDMP), le terme de diffusion est absent, et le caractère aléatoire est réduit au choix des moments et des lieux de sauts.

L'étude générale de ces processus a été lancée par Davis ([Dav84, Dav93]). Ils sont présents dans de très nombreux domaines de modélisation : citons la modélisation de réseaux de communications ([DGR02, GRZ04, GR09, GR11]), la biologie moléculaire ([RMC07, BR11]), les modèles de populations ([ODA88, EO05]), la neurobiologie ([BR11, PTW10]), la fiabilité ([DD02]), des modèles de suivi de stocks [BKKP05],... Pour plus de références nous renvoyons au compte-rendu [5].

L'étude de ces processus avec les outils usuels rappelés dans l'introduction présente plusieurs difficultés.

- Ils sont en général **non-réversibles**, ce qui empêche l'utilisation naïve de techniques spectrales.
- Le **manque de bruit** rend difficile la construction de couplages.
- Les deux ingrédients se conjuguent pour expliquer que les inégalités fonctionnelles s'adaptent mal ; en cela les PDMP se rapprochent des processus hypo-elliptiques.

Remarque 4.1.1. On peut tenter de contourner ces obstacles. Pour les deux premiers, des travaux de Gadat, Miclo et Monmarché [MM13, GM13] utilisent une décomposition spectrale explicite sur des exemples simples pour étudier finement le comportement de semi-groupes, spécialement en temps petit. Un travail récent de Monmarché [Mon14] s'intéresse également à la preuve d'inégalités fonctionnelles adaptées. Nous verrons plus loin comment certaines techniques de couplage restent efficaces pour établir la convergence de ces processus.

Avec divers collaborateurs, nous nous sommes intéressés dans [6, 2, 4, 7, 5] à deux familles d'exemples.

a) Autour de TCP

La première famille est reliée au processus « TCP ». Celui-ci vient d'un modèle informatique simple en temps discret. Un agent cherche à transmettre des données par un canal de transmission. À chaque

intervalle de temps, il envoie X_n paquets de données : si la transmission se passe sans erreur, il augmente son débit : $X_{n+1} = X_n + 1$; en cas d'erreur due à la congestion du réseau, l'agent diminue la taille de sa fenêtre de transmission ; dans le modèle le plus simple il la divise par 2 : $X_{n+1} = X_n/2$.

Le modèle n'est complet que si l'on décrit la loi des erreurs de transmission. Dans un modèle naïf, les erreurs arrivent à chaque temps n avec une probabilité fixée p . Un modèle un peu plus réaliste est de dire que chaque paquet envoyé encourt le même risque d'échec de transmission.

Dans la suite nous nous intéresserons à un modèle en temps et espace continu, obtenu par limite d'échelle du modèle discret (voir [DGR02, GRZ04, OKM96] pour plus de détails sur ce passage à la limite et pour des variantes) :

Définition 4.1.2 (Processus TCP). *Soit $\lambda : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ une fonction positive. Le processus TCP à taux λ est le processus sur $\mathbb{R}_+$ de générateur formel :*

$$Lf(x) = f(x) + \lambda(x)(f(x/2) - f(x)).$$

Si λ est constant le processus est extrêmement simple, ceci correspond au modèle naïf où le taux d'erreur ne dépend pas de la quantité de données envoyée. Le modèle plus réaliste correspond au choix $\lambda(x) = x$.

Notons également que ce processus est intimement lié à des modèles biologiques de croissance cellulaire : ici le X_t représente une taille de cellule, qui augmente avec le temps jusqu'à une division cellulaire. Ces modèles font intervenir d'autres questions que nous n'aborderons pas ici ; en particulier il est naturel de suivre non pas une seule particule mais toute une population de particules en croissance/fragmentation. Ceci est par exemple utilisé dans [DHKR12] pour la construction d'estimateurs statistiques de paramètres biologiques (voir aussi [Clo11]) ; ces modèles sont également très étudiés du point de vue des ÉDP (voir par exemple [PR05]).

b) Flots modulés

Soit E un ensemble fini, et pour tout $i \in E$ donnons-nous un champ de vecteurs $F^i : \mathbb{R}^d \rightarrow \mathbb{R}^d$. Supposons pour simplifier ces champs lipschitziens. On peut alors définir sur $\mathbb{R}_+$ le flot de F^i , $\Phi^i = \{\Phi_t^i\}$, en demandant que

$$t \mapsto \Phi_t^i(x) = \Phi^i(t, x)$$

soit l'unique solution du problème de Cauchy $\dot{x} = F^i(x)$, pour la condition initiale $x(0) = x$.

Soit enfin $\lambda(i, j; x)$ une famille de fonctions positives.

Définition 4.1.3 (Flots modulés). *Le processus de flots modulés est le processus $(Z_t = (X_t, I_t))$ à valeurs dans $\mathbb{R}^d \times E$ dont le générateur agit sur des fonctions*

$$\begin{aligned} g : M \times E &\rightarrow \mathbb{R}, \\ (x, i) &\mapsto g(x, i) = g^i(x), \end{aligned}$$

suivant la formule

$$Lg(x, i) = \langle F^i(x), \nabla g^i(x) \rangle + \sum_{j \in E} \lambda(x, i, j)(g^j(x) - g^i(x)). \quad (4.1)$$

Intuitivement, partant de (x, i) , la composante continue X_t du processus évolue suivant le flot déterministe donné par le champ $F^{I_t} = F^i$, jusqu'à ce que la composante discrète I_t saute vers une autre valeur j . La composante X_t suit alors le champ F^j , jusqu'au prochain saut, etc. La composante discrète

saute de i à j avec une intensité $\lambda(x; i, j)$ qui dépend de la composante continue. Nous renvoyons à la figure 1.2, p. 10 pour une illustration détaillée, et aux figures 4.1 et 4.2 plus bas pour d'autres exemples de trajectoires.

Cette classe de processus, malgré sa simplicité, est déjà très riche et donne lieu comme nous le verrons à une grande variété de comportements. En plus de ce caractère de « modèle test », ces processus sont également utilisés en modélisation dans de nombreux domaines, comme la biologie moléculaire ([RMC07, BR11], voir [YZ10] pour d'autres applications). Ces processus peuvent aussi se voir comme résultat d'une itération de fonctions aléatoires (voir la synthèse [DF99] pour plus d'informations sur ce domaine).

c) Plan du chapitre

Les deux sections 4.2 et 4.3.b détaillent plusieurs résultats **qualitatifs** sur le comportement en temps long du modèle de flots modulés. La section 4.4 présente deux résultats **quantitatifs** de convergence vers l'équilibre par couplage, pour le processus TCP et pour les flots modulés.

4.2 Le zéro ou l'infini ?

La première question naturelle est de savoir quels sont les points que le processus peut visiter. Remarquons que l'emplacement du processus au temps t dépend uniquement des flots successifs qu'il a suivis et des temps correspondants. Pour $\mathbf{i}$ une suite finie d'indices $(i_0, \dots, i_{n-1})$ et $\mathbf{u}$ une suite de temps $(u_1, \dots, u_n)$, soit $\phi_{\mathbf{u}}^{\mathbf{i}}$ le « flot composé » :

$$\phi_{\mathbf{u}}^{\mathbf{i}}(x) = \Phi_{u_n}^{i_{n-1}} \circ \dots \circ \Phi_{u_1}^{i_0}(x). \quad (4.2)$$

Savoir où le processus est au temps t est alors naturellement relié à une question de **contrôle** sous-jacente : si l'on peut choisir les moments de saut et les flots suivis, où peut-on amener le processus ?

Nous montrons dans [4] que le **support de la loi des trajectoires** partant de x est exactement l'ensemble des **solutions d'une inclusion différentielle** déterminée uniquement par les champs de vecteurs F^i : de ce point de vue, l'aléatoire ne joue aucun rôle.

L'aléa redevient crucial si l'on considère le **comportement en temps long** du processus : nous illustrons ce point dans [2] sur un exemple de PDMP dans le plan $\mathbb{R}^2$.

Nous avons rappelé dans la section 1.2.d que si l'on considère une seule dynamique linéaire dans le plan, gouvernée par le champ de vecteurs $F^0(x) = A_0 \cdot x$, avec A_0 une matrice de Hurwitz, les trajectoires convergent exponentiellement vite vers 0.

Rajoutons maintenant une seconde matrice A_1 et considérons d'abord le problème de contrôle suivant : si A_0 et A_1 sont des matrices de Hurwitz, quels points peut-on atteindre en suivant alternativement les flots de A_0 et de A_1 ? Dans l'article [BBM09], quatre cas sont identifiés explicitement (en termes d'inégalités entre diverses traces et déterminants). Dans le premier cas il existe une fonction de Lyapunov quadratique commune, c'est à dire qui décroît le long des trajectoires des deux champs. Il est alors clair que pour tout choix des moments de sauts, la trajectoire convergera exponentiellement vite vers 0. Le deuxième cas¹ du théorème de [BBM09] est celui qui nous intéressera le plus : il correspond à un exemple où A_0 et A_1 sont de Hurwitz, mais une combinaison convexe $A_\lambda = (1 - \lambda)A_0 + \lambda A_1$ a une valeur propre strictement positive (disons pour $\lambda = 1/2$). Dans ce cas, les flots, considérés séparément, convergent vers 0, mais on peut faire

1. Le troisième cas est un cas dégénéré, nous reviendrons sur le quatrième cas plus loin.

alterner intelligemment pour faire tendre la trajectoire vers l'infini ! Du point de vue déterministe, dans cette situation, la région accessible du plan n'est pas bornée.

Nous étudions la **version aléatoire** de l'exemple précédent : que se passe-t-il si l'on ne contrôle plus les sauts mais qu'ils arrivent selon un processus de Poisson d'intensité β ?

Théorème 4.2.1 (Le zéro ou l'infini). *Il existe un $\chi(\beta) \in \mathbb{R}$ tel que, pour toute mesure initiale ν ne chargeant pas l'origine,*

$$\frac{1}{t} \log \|X_t\| \xrightarrow[t \rightarrow \infty]{\mathbb{P}_\nu - a.s.} \chi(\beta). \quad (4.3)$$

De plus, il existe deux constantes $0 < \beta_1 \leq \beta_2 < \infty$ telles que :

- si $\beta < \beta_1$, alors $\chi(\beta)$ est strictement négatif et $\|X_t\| \xrightarrow[t \rightarrow \infty]{\mathbb{P}_\nu - a.s.} 0$,
- si $\beta > \beta_2$, $\chi(\beta)$ est strictement positif et $\|X_t\| \xrightarrow[t \rightarrow \infty]{\mathbb{P}_\nu - a.s.} \infty$.

Ce résultat est illustré dans la figure 4.1.

Si l'on garde en tête le problème de contrôle précédent, l'explication intuitive est relativement simple : quand β est petit, le système saute peu, et chaque champ est suivi suffisamment longtemps pour se rapprocher — en norme euclidienne ! — de l'origine. Au contraire, quand β est grand, le processus aléatoire est proche d'une évolution déterministe suivant la matrice $\frac{1}{2}(A_0 + A_1)$; si celle-ci a une valeur propre positive, la norme a tendance à augmenter.

Si la preuve du premier cas est simple, celle de l'explosion est plus difficile. L'argument fondamental est une **décomposition polaire** du processus : en écrivant $X_r = R_t \Theta_t$ pour R_t dans $\mathbb{R}_+$ et Θ_t dans le cercle, on voit aisément que le processus angulaire (Θ_t, I_t) est autonome : c'est un PDMP sur le cercle. On montre également facilement qu'en dehors des instants de saut, $\dot{R}_t = R_t \langle \Theta_t, A_{I_t} \Theta_t \rangle$. Par conséquent $\log(R_t)$ apparaît comme une intégrale : en posant e_θ le vecteur unitaire $(\cos(\theta), \sin(\theta))$ et $\mathcal{A}(\theta, i) = \langle A_i e_\theta, e_\theta \rangle$, on a :

$$\log(R_t) = \log(R_0) + \int_0^t \mathcal{A}(\Theta_s, I_s) ds. \quad (4.4)$$

et $\log(R_t)$ est donc une fonctionnelle additive du PDMP angulaire $((\Theta_s, I_s))_{0 \leq s \leq t}$.

Toute la question se ramène donc à établir un **théorème ergodique** pour le processus (Θ_t, I_t) pour une mesure invariante ν_β , puis à rassembler suffisamment d'information sur la mesure ν_β pour **déterminer le signe** de $\int \mathcal{A}(\theta, i) d\nu_\beta(\theta, i)$.

Au vu du théorème 4.2.1, on peut immédiatement se demander si la transition est unique, c'est-à-dire si $\beta_1 = \beta_2$. Même dans notre cadre simple, nous n'avons pu établir ce résultat que pour des cas particuliers de matrices A_i .

Ce travail a inspiré le résultat encore plus surprenant, dû à S. Lawley, J. Mattingly et M. Reed, et illustré dans la figure 4.2.

Théorème 4.2.2 (Un « intervalle » d'explosion, [LMR14]). *Il existe A_0 et A_1 deux matrices de Hurwitz, dont la moyenne $A_{1/2} = (1/2)(A_0 + A_1)$ est encore Hurwitz, et telles que la norme du processus X_t :*

- converge exponentiellement vite vers 0 si β est suffisamment petit,
- converge exponentiellement vite vers 0 si β est suffisamment grand,
- explose pour au moins un choix de β .

La preuve repose sur la même décomposition polaire que précédemment. Là encore, sauf dans des cas particuliers, il n'est pas sûr qu'il n'y ait que deux valeurs critiques pour β .

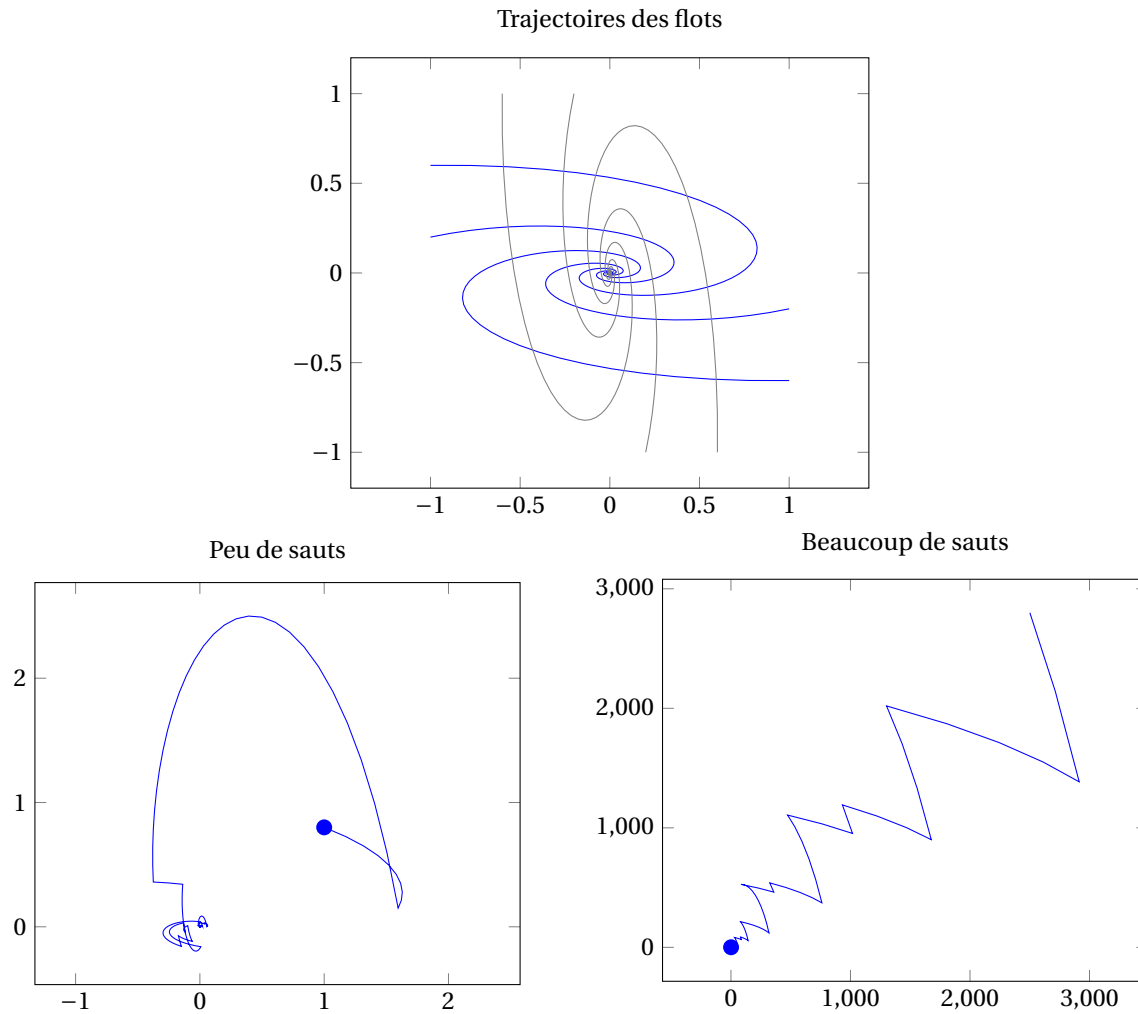


FIGURE 4.1 – Convergence ou divergence pour les flots modulés

Illustration du théorème 4.2.1. Les deux matrices A_0 , A_1 sont de Hurwitz, mais leur demi-somme a une valeur propre positive. Les trajectoires correspondant à chacune des matrices sont illustrées en haut. Si le taux de saut est faible (en bas à gauche), le processus converge vers 0 ; s'il est fort (en bas à droite) les trajectoires ressemblent à celles du système moyenné, et le processus diverge.

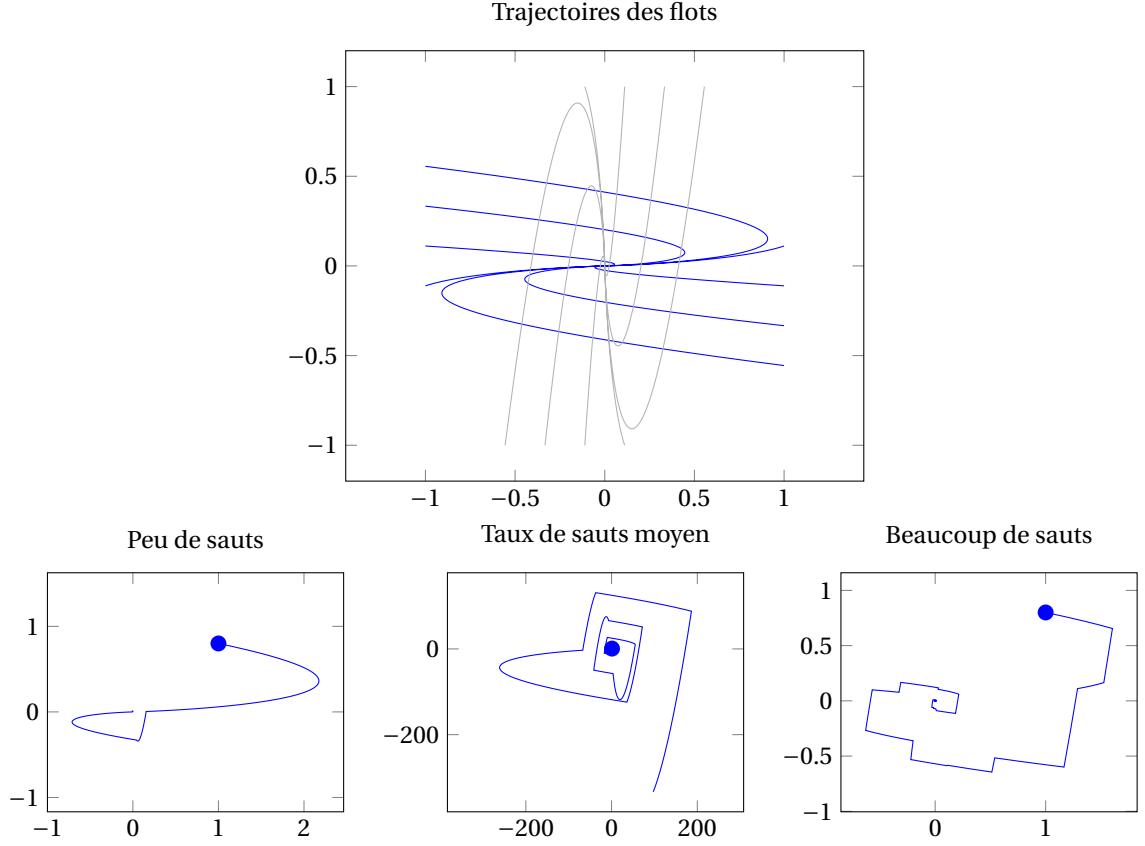


FIGURE 4.2 – Convergence ou divergence pour les flots modulés (suite)

Illustration du théorème 4.2.2 (issu de [LMR14]). Ici, la matrice moyennée est encore de Hurwitz. Il y a donc convergence pour les taux de saut très faibles (à gauche) ou très forts (à droite). Pour au moins une valeur intermédiaire, le processus diverge.

4.3 Convergence « abstraite » vers l'équilibre

a) Un ensemble de points récurrents

Dans la suite nous excluons les questions de divergence vers l'infini en supposant qu'il existe un compact M positivement invariant pour les flots Φ^i :

$$\forall i \in E, \forall t \geq 0, \quad \Phi_t^i(M) \subset M. \quad (4.5)$$

Sous cette condition, les taux de saut $\lambda(i, j, \cdot)$ sont majorés uniformément par un $\lambda < \infty$. Ceci permet une construction explicite du processus à temps continu via une chaîne de Markov auxiliaire $\tilde{Z}_n = (\tilde{X}_n, \tilde{I}_n)$ sur $M \times E$ et une « temporisation » poissonnienne d'intensité constante λ . D'autres chaînes ont été étudiées (voir en particulier [DC99, CD08]), celle-ci semble dans ce cas restreint particulièrement adaptée, le

caractère constant de l'intensité rendant beaucoup de calculs plus agréables. On a par exemple une bijection explicite entre les mesures invariantes pour la chaîne et pour le processus.

Introduisons maintenant un « ensemble limite » de points dont on peut toujours se rapprocher, quel que soit le point de départ.

Définition 4.3.1 (Orbite d'un point). *Soit $\mathbb{T}$ l'ensemble des couples $(\mathbf{i}, \mathbf{u})$ où $\mathbf{i}$ est une suite finie d'indices et $\mathbf{u}$ une suite finie de temps, de même longueur que $\mathbf{i}$. L'orbite positive de x est l'ensemble :*

$$\gamma^+(x) = \left\{ \phi_{\mathbf{u}}^{\mathbf{i}}(x) : (\mathbf{i}, \mathbf{u}) \in \mathbb{T} \right\},$$

c'est l'ensemble des points que l'on peut atteindre en partant de x et en suivant des flots arbitraires pendant des temps arbitraires.

Cette orbite est un ensemble déterministe, et il est clair qu'une trajectoire du processus partant de x reste dans cet ensemble.

Définition 4.3.2 (Ensemble accessible). *L'ensemble accessible est le compact (éventuellement vide) défini par*

$$\Gamma = \bigcap_{x \in M} \overline{\gamma^+(x)}.$$

Cet ensemble, défini uniquement à partir des flots déterministes, a des liens forts avec le processus ; en particulier il donne des renseignements sur le support des mesures invariantes.

Proposition 4.3.3 (Ensemble accessible et mesures invariantes).

- (i) *Si $\Gamma \neq \emptyset$, alors pour toute mesure invariante μ , $\Gamma \times E \subset \text{Supp}(\mu)$. De plus il y a égalité pour au moins une mesure invariante ; en cas d'unicité on a donc $\text{Supp}(\mu) = \Gamma \times E$.*
- (ii) *Si Γ est d'intérieur non-vide, toute mesure invariante a pour support $\Gamma \times E$.*

Le premier résultat est optimal : nous construisons dans [4] un exemple où, selon les taux de sauts, il existe une ou trois mesures invariantes ; quand il y en a trois, une seule a pour support $\Gamma \times E$.

Remarque 4.3.4. On peut également construire un exemple où Γ est d'intérieur non-vide mais pour lequel il y a plusieurs mesures invariantes.

Cet ensemble Γ a encore d'autres propriétés intéressantes : on a par exemple le résultat de récurrence suivant pour la chaîne discrète $\tilde{Z}_N$:

Proposition 4.3.5 (Récurrence pour les points de Γ). *Si $p \in \Gamma$, tout voisinage $\mathcal{U}$ de p est récurrent : pour tout point $(x, i) \in M \times E$,*

$$\mathbb{P}_{x,i} [\exists n \in \mathbb{N}, \tilde{X}_n \in \mathcal{U}] = 1.$$

b) Régularité et convergence « abstraite » vers l'équilibre

Les résultats décrits ci-dessus donnent une piste pour montrer l'unicité de la mesure invariante et la convergence du processus vers celle-ci : adapter l'argument classique de Foster–Lyapunov rappelé dans l'introduction (Théorème 1.2.6). Soit en effet p un point de Γ .

- Par la propriété de récurrence² de la proposition 4.3.5, le processus revient toujours près de p : ceci remplace la condition de Lyapunov ;

2. Plus précisément, par l'analogue de cette propriété pour le processus à temps continu.



FIGURE 4.3 – Une récurrence de Hanoi

Si l'on sait engendrer une famille de vecteurs $\mathcal{H}$ en suivant les indices $\mathbf{i}$ pendant des temps $\mathbf{u}$, alors on peut engendrer F^i , les $[F^i, H]$ et les vecteurs de $\mathcal{H}$ en suivant la séquence $(\mathbf{i}, \mathbf{u})$, puis le champ i , puis la séquence $\mathbf{i}, \mathbf{u}$ de nouveau, à condition de choisir le temps u beaucoup plus grands que tous les autres temps.

— on peut espérer coupler (coller) deux copies du processus partant près de p en un temps $t_0 > 0$.

Ce dernier point n'est pas évident et n'a pas de raison d'être vrai en général : l'**absence de diffusion** se fait ici cruellement sentir. Pour assurer ce couplage, on cherche alors à établir des **conditions** pour que, partant d'un point, **la loi de la position** X_t ait une partie **absolument continue** par rapport à la mesure de Lebesgue.

Le résultat est un **analogue du critère classique de Hörmander** pour la régularité des diffusions hypo-elliptiques. Pour l'exprimer, donnons quelques notations. Soit $\mathcal{F}_0$ la collection de champs de vecteurs $\{F^i : i \in E\}$. Soit $\mathcal{F}_k = \mathcal{F}_{k-1} \cup \{[F^i, V], V \in \mathcal{F}_{k-1}\}$ (où $[F, G]$ est le crochet de Lie de F et G) et $\mathcal{F}_k(x)$ l'espace vectoriel engendré par $\{V(x), V \in \mathcal{F}_k\}$.

De même, partant de $\mathcal{G}_0 = \{F^i - F^j, i \neq j\}$, définissons $\mathcal{G}_k$ en prenant itérativement les crochets de Lie avec les $\{F^i : i \in E\}$, et soit $\mathcal{G}_k(x)$ le sous-espace de $\mathbb{R}^d$ correspondant.

Théorème 4.3.6 (Critère de type Hörmander, [BH12], [4]). *Si pour un entier k , $\mathcal{F}_k(x_0) = \mathbb{R}^d$, on dit que la « condition de crochets faible » est vérifiée. Alors il existe un l tel que la loi de $\tilde{X}_l$ partant de x ait une partie absolument continue par rapport à la mesure de Lebesgue, uniformément pour x dans un voisinage de x_0 .*

Si pour un entier k , $\mathcal{G}_k(x_0) = \mathbb{R}^d$, la « condition de crochets forte » est vérifiée, et il existe un t_0 tel que la loi de X_t partant de x ait une partie absolument continue par rapport à la mesure de Lebesgue, uniformément par rapport à x dans un voisinage de x_0 et $t \in [t_0, t_0 + \varepsilon]$.

Remarque 4.3.7. Il est facile de vérifier que $\mathcal{G}_k(x)$ est un sous-espace de $\mathcal{F}_k(x)$, la condition « forte » implique donc bien la condition « faible ». On peut construire un exemple où la condition faible est vérifiée, mais pas la forte.

Ce théorème est une version légèrement renforcée du théorème 2 de [BH12]. La preuve de [BH12] est très élégante mais non constructive. Nous montrons dans [4] comment construire explicitement la suite de sauts à faire pour créer de la densité par rapport à la mesure de Lebesgue, en jouant sur les alternances de flots et les échelles de temps.

Cette « création » de densité, reliée au fait que le processus revient nécessairement près de Γ , permet de mener à bien la stratégie expliquée ci-dessus et de montrer le résultat suivant.

Théorème 4.3.8 (Convergence en variation totale). *Si la condition de crochets faible est vérifiée en un point $p \in \Gamma$, alors la chaîne $\tilde{Z}$ admet une unique mesure invariante $\tilde{\pi}$, absolument continue par rapport à la mesure de Lebesgue $\lambda_{M \times E}$ sur $M \times E$. De plus il existe deux constantes $c > 1$ et $\rho \in (0, 1)$ telles que, pour tout n ,*

$$\|\mathbb{P}[\tilde{Z}_n \in \cdot] - \tilde{\pi}\|_{TV} \leq c\rho^n$$

où $\|\cdot\|_{TV}$ est la norme de variation totale.

Nous montrons également un résultat similaire pour le processus à temps continu, en supposant vérifiée la condition de crochets forte.

4.4 Convergence vers l'équilibre : couplages astucieux

Les considérations précédentes peuvent parfois permettre de montrer une convergence exponentielle vers l'équilibre. Si l'on s'intéresse à la **vitesse explicite** de convergence, les résultats ne sont que peu applicables :

- dans le cas des flots modulés, on utilise des arguments de compacité qui empêchent toute estimation de la vitesse ;
- dans le cas du processus TCP, des bornes de convergence explicites peuvent être obtenus par la méthode de Foster-Lyapunov (voir par exemple [RR96]). Le calcul des bornes obtenues donne malheureusement de très mauvais résultats, avec des estimations du temps de relaxation de l'ordre de 10^{15} ... Ces estimations peuvent être améliorées pour des processus stochastiquement monotones, mais cette propriété n'est pas vérifiée pour le processus TCP.

L'amélioration de l'estimation de ces constantes passe alors par la construction de **couplages mieux adaptés**, prenant en compte les spécificités des processus. Nous avons mené une telle étude dans deux cas.

a) De bonnes vitesses pour TCP

Reprenons l'exemple du processus TCP, de générateur

$$Lf(x) = f'(x) + \lambda(x)(f(x/2) - f(x)).$$

On peut montrer facilement que, pour un taux de saut $\lambda(x)$ constant ou linéaire ($\lambda(x) = x$), le processus est bien défini en tout temps (il ne peut pas exploser en temps fini), et qu'il admet une unique mesure de probabilité invariante. Cette mesure a même une densité explicitable sous forme de série ([DGR02]).

Pour quantifier la convergence, D. Chafaï, F. Malrieu et K. Paroux ont introduit dans [CMP10] un **couplage** entre deux réalisations (X_t, X'_t) du processus. Dans le cas des taux constants, les instants de sauts du processus sont simplement donnés par un processus de Poisson. Le couplage le plus simple est alors de prendre les mêmes sauts pour les deux processus. La différence $D_t = X_t - X'_t$ est laissée invariante par la dérive, et est divisée par 2 à chaque saut ; elle suit donc la loi de $(X_0 - X'_0)(1/2)^{N_t}$ où N_t est un processus de Poisson d'intensité λ . En particulier, $|D_t|$ décroît et on a

Théorème 4.4.1 (Convergence en vitesse de Wasserstein pour le processus naïf, [CMP10]). *La distance de Wasserstein entre les lois des processus partant de deux points x, x' décroît exponentiellement vite :*

$$W_1(\delta_x P_t, \delta_{x'} P_t) \leq |x - x'| \exp(-\lambda t/2).$$

Pour le « vrai » processus TCP (à taux linéaire), deux processus (X_t, X'_t) partant de deux points différents x et x' n'ont pas les mêmes taux de sauts : on ne peut pas construire de couplage qui garantisse que tous les sauts soient simultanés. Toujours dans [CMP10], les auteurs proposent un couplage Markovien qui **maximise les sauts simultanés**. Le générateur du couple est donné, pour $x < x'$, par

$$\begin{aligned} Lf(x, x') &= \partial_1 f(x, x') + \partial_2 f(x, x') && \text{(dérives linéaires)} \\ &+ x(f(x/2, x'/2) - f(x, x')) && \text{(sauts simultanés au taux } x = \min(x, x')) \\ &+ (x' - x)(f(x, x'/2) - f(x, x')) && \text{(le processus du haut saute seul)} \end{aligned}$$

La dynamique de la différence $|D_t| = |X_t - X'_t|$ est beaucoup moins sympathique dans ce cas : elle est toujours invariante par la dérive, et est divisée par 2 en cas de sauts simultanés, mais elle est multipliée par $|X_t - X'_t|/2$ qui peut être arbitrairement grand, si le processus du haut saute seul !

En utilisant ce couplage, les auteurs de [CMP10] montrent, pour le vrai processus TCP, un résultat de convergence à vitesse polynomiale ([CMP10, Théorème 2.5]).

Avec J.-B. Bardet, A. Christen, A. Guillin et F. Malrieu, nous avons renforcé et étendu ces résultats de convergence sur deux plans. Nous avons tout d'abord réussi à montrer que, pour le vrai TCP, la convergence en distance de Wasserstein est bien exponentielle :

Théorème 4.4.2 (Convergence en distance de Wasserstein pour TCP). *[6] Il existe un $\lambda \approx 0.12$ et, pour tout $\tilde{\lambda} < \lambda$, tout $p \geq 1$ et tout $t_0 > 0$, une constante $C = C(p, \tilde{\lambda}, t_0)$ telle que, pour toute mesures initiales $\nu, \tilde{\nu}$ et tout $t \geq t_0$,*

$$W_p(\nu P_t, \tilde{\nu} P_t) \leq C \exp\left(-\frac{\tilde{\lambda}}{p} t\right).$$

Nous avons également obtenu des bornes pour la convergence en variation totale, tant pour le processus naïf que pour le vrai TCP :

Théorème 4.4.3 (Convergence en variation totale pour TCP). *[6] Pour tout $\tilde{\lambda} < \lambda \approx 0.12$ et tout $t_0 > 0$, il existe C tel que pour toutes mesures initiales $\nu, \tilde{\nu}$, et pour tout $t \geq t_0$,*

$$\|\nu P_t - \tilde{\nu} P_t\|_{TV} \leq C \exp\left(-\frac{2\tilde{\lambda}}{3} t\right).$$

Remarquons que la convergence en variation totale pour le processus à taux constant est également étudiée dans [PR05, LP09], par des méthodes d'ÉDP, qui donnent ici des résultats légèrement moins forts.

L'idée de la preuve du théorème 4.4.2 est de construire une fonction $V(X_t, X'_t)$ qui contrôle la distance entre X_t et X'_t , et qui décroît (en moyenne) au cours du temps. Deux problèmes se posent :

- si X_t et X'_t sont très proches l'un de l'autre, un saut isolé du plus grand est peu probable mais fait exploser la distance ;
- si X_t et X'_t sont tous deux très proches de 0, leur taux de saut est très faible et ils restent à distance constante l'un de l'autre.

Pour contourner le premier problème, on étudie l'évolution d'une fonction bien choisie de la distance (du type $\sqrt{x-y}$) ; pour le deuxième on arrive à utiliser le fait que la dérive fait sortir de la zone où les sauts sont rares, en multipliant la fonction « distance » par un poids bien choisi.

Pour obtenir les bornes explicites en variation totale du théorème 4.4.3 l'idée est de construire un **nouveau couplage** à partir du précédent. Le couplage de [CMP10], qui maximise les sauts simultanés, n'a en effet aucune chance de coller les deux processus : il est donc inutilisable tel quel pour montrer une estimation de variation totale.

Comme on l'a vu dans l'introduction, l'approche classique du problème consiste à identifier un ensemble K sur lequel la condition de minoration est vérifiée, puis à coupler (typiquement de manière indépendante) deux copies jusqu'à ce qu'elles touchent K toutes les deux et à essayer de les coller à ce moment. Dans le cas particulier de notre processus, on peut facilement montrer une condition de minoration, non pas quand x et x' sont tous deux dans K , mais quand ils sont **proches l'un de l'autre**. On construit un couplage de la façon suivante (pour le processus à taux constant, le cas général étant similaire). On tire d'abord les instants de sauts $T_1, \dots, T_{N_t}$ suivant un processus de Poisson d'intensité λ . Si $N_t = 0$ les processus ne sautent pas et on ne peut pas les coupler. Si $N_t \geq 1$:

- jusqu'au temps $U = T_{N_t-1}$ on fait faire aux deux processus des sauts simultanés ; si N_t est suffisamment grand, les deux processus sont alors très proches au temps U ;
- On construit les trajectoires sur $[U, t]$ en utilisant le meilleur couplage (au sens de la variation totale) entre les lois $\delta_{X_t} P_{t-U}$ et $\delta_{X'_t} P_{t-U}$.

Cette idée d'un **couplage en deux temps** (rapprochement puis collage) s'applique également à d'autres modèles, où elle donne également des estimations de vitesse beaucoup plus raisonnables que la méthode classique.

b) Couplage pour des flots modulés

Dans [7], nous établissons des résultats quantitatifs de convergence à l'équilibre pour des cas particuliers de flots modulés. Nous considérons d'abord un cas simple où les taux de sauts $\lambda(i, j; x)$ de I ne dépendent pas de la position x . Le processus I pris tout seul est alors Markovien, et la position (X_t) s'écrit comme fonction (déterministe) de (I_t) . On suppose que I est une chaîne de Markov irréductible sur E et on note ν sa mesure invariante.

Hypothèse 4.4.4 (Dissipation en moyenne). *Pour tout $i \in E$, il existe un $\alpha(i) \in \mathbb{R}$ tel que :*

$$\langle x - \tilde{x}, F^i(x) - F^i(\tilde{x}) \rangle \leq -\alpha(i)|x - \tilde{x}|^2, \quad x, \tilde{x} \in \mathbb{R}^d, \quad (4.6)$$

et de plus

$$\sum_{i \in E} \alpha(i) \nu(i) > 0. \quad (4.7)$$

Sous cette hypothèse on obtient :

Théorème 4.4.5. *Soit $p < q < \kappa$, et supposons que μ_0 et $\tilde{\mu}_0$ ont un moment d'ordre q majoré par une constante m . Alors il existe une constante $C(p, m)$ telle que*

$$\mathcal{W}_p(\mu_0 P_t, \tilde{\mu}_0 P_t) \leq C(p, m) \exp(-\theta_p t),$$

où la constante θ_p ne dépend que de p et de la chaîne de Markov I .

Il y a en particulier convergence exponentielle en distance de Wasserstein vers l'unique mesure invariante.

Remarque 4.4.6. Les expressions de θ_p et $C(p, m)$ sont explicitées dans [7].

Remarque 4.4.7 (Dissipation moyenne). L'équation (4.7) fait intervenir la moyenne des quantités (signées) $\alpha(i)$; on couvre donc des cas où certains flots font diverger la position vers l'infini, mais où ils sont compensés par d'autres, qui sont suivis plus longtemps car les indices correspondants ont plus de poids dans la mesure invariante ν .

La preuve de ce résultat est relativement simple et se fait par couplage. Partant de (x, i) , $(\tilde{x}, \tilde{i})$, on fait évoluer les indices I_t, I'_t par couplage indépendant jusqu'à ce qu'ils se collent, puis on les laisse ensemble. Après le temps de coalescence T , les parties continues X_t et X'_t évoluent toujours suivant le même champ, ce qui permet d'appliquer la contraction (4.6).

Nous considérons ensuite le cas plus complexe où les taux de saut peuvent dépendre de la position.

Hypothèse 4.4.8. Il existe $\underline{a} > 0$ et $\kappa > 0$ tels que, pour tout $x, \tilde{x} \in \mathbb{R}^d$ et $i, j \in E$,

$$a(x, i, j) \geq \underline{a} \quad \text{et} \quad \sum_{j \in E} |a(x, i, j) - a(\tilde{x}, i, j)| \leq \kappa |x - \tilde{x}|.$$

De plus, il existe $\alpha > 0$ tel que

$$\langle x - \tilde{x}, F^i(x) - F^i(\tilde{x}) \rangle \leq -\alpha |x - \tilde{x}|^2, \quad x, \tilde{x} \in \mathbb{R}^d, \quad i \in E. \quad (4.8)$$

Ces hypothèses renforcées impliquent en particulier que, pour tout $i \in E$,

$$|\varphi_t^i(x) - \varphi_t^i(\tilde{x})| \leq e^{-\alpha t} |x - \tilde{x}|, \quad x, \tilde{x} \in \mathbb{R}^d.$$

Les champs F^i ont donc exactement un point stable $\sigma(i)$, pour lequel :

$$|\varphi_t^i(x) - \sigma(i)| \leq e^{-\alpha t} |x - \sigma(i)|.$$

On peut alors facilement montrer qu'il existe nécessairement une boule $B(0, r)$ stable par tous les flots.

Théorème 4.4.9. Il existe deux constantes c et γ telles que, si μ_0 et $\tilde{\mu}_0$ sont à support dans $B(0, r)$,

$$\mathcal{W}_1(\eta_t, \tilde{\eta}_t) \leq (1 + 2r)(1 + ct) \exp\left(-\frac{\alpha}{1 + \alpha/\gamma} t\right)$$

où α est le taux de dissipation (4.8).

Là encore, ceci implique l'unicité de la mesure invariante et donne une vitesse de convergence. Les constantes c et γ peuvent là aussi être explicitées.

Remarque 4.4.10. Ces conditions peuvent sembler très restrictives, mais à ce niveau de généralité il est difficile d'espérer mieux : les exemples de la section 4.2 montrent que l'alternance de deux flots linéaires, convergeant tous deux exponentiellement vite vers 0, peut diverger presque-sûrement vers plus l'infini quand ils n'ont pas de fonction de Lyapunov commune...

La preuve du théorème 4.4.9 se fait là encore par couplage, mais l'argument est plus délicat à mettre en œuvre : comme les taux de saut dépendent de la position, le processus I_t seul n'est plus Markovien, et on ne peut plus construire un couplage coalescent simple. On procède alors de la manière suivante :

- tant que I_t et $\tilde{I}_t$ diffèrent, on fait évoluer les deux processus indépendamment,
- si I_t et $\tilde{I}_t$ coïncident, on couple les processus de manière à maximiser la probabilité qu'ils sautent ensemble vers le même point.

Comme les composantes continues X_t et X'_t se rapprochent exponentiellement vite tant que les indices I_t et I'_t coïncident, et comme les taux de sauts sont proches quand X_t et X'_t sont proches, on parvient à montrer que les processus restent ensemble avec une grande probabilité.

Chapitre 5

Miscellanées

5.1 Gaz de Coulomb

Les résultats de cette section, obtenus en collaboration avec D. Chafaï et N. Gozlan, sont issus de [3].

a) Un modèle de particules

Considérons une généralisation du modèle de particules en interaction évoqué brièvement dans l'introduction : étant donné deux fonctions $V : \mathbb{R}^d \rightarrow \mathbb{R}$ et $W : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$, et deux familles (q_N, β_N) de réels, on regarde le système d'équations différentielles stochastiques couplées

$$dX_{i,t} = \sqrt{\frac{2}{\beta_N}} dB_{i,t} - q_N \nabla V(X_{i,t}) - \sum_{j \neq i} q_N^2 \nabla_1 W(X_{i,t}, X_{j,t}) dt$$

où $(B_t^{(1)})_{t \geq 0}, \dots, (B_t^{(N)})_{t \geq 0}$ sont des mouvements brownien indépendants sur $\mathbb{R}^d$. Dans ce modèle :

- chaque X_i représente la position d'une particule, de charge électrique q_N , se déplaçant dans l'espace $\mathbb{R}^d$;
- les particules sont toutes soumises à un champ électrique, gradient du potentiel électrique V ;
- elles interagissent par paires suivant le potentiel d'interaction W ;
- elles sont soumises à une agitation thermique brownienne, à la température $1/\beta_N$.

Le cas mentionné dans l'introduction correspond, en dimension $d = 3$, à un potentiel de confinement V quadratique, et à une interaction répulsive de type « Coulombien » :

$$W(x, y) = \begin{cases} |x - y|, & \text{pour } d = 1, \\ -\log(|x - y|), & \text{pour } d = 2, \\ \frac{1}{|x - y|^{d-2}}, & \text{pour } d \geq 3. \end{cases}$$

Notons que ces interactions, naturelles pour le modèle de répulsion électrique, explosent sur la diagonale. En dimension 1 et 2, on obtient des processus très liés aux modèles de matrice aléatoires : le processus en dimension 1 est par exemple un mouvement Brownien de Dyson ([AGZ10]).

La dynamique dans $(\mathbb{R}^d)^n$ admet une mesure d'équilibre P_N . En posant

$$H_N(x_1, \dots, x_N) := \sum_{i=1}^N q_N V(x_i) + \sum_{i<j} q_N^2 W(x_i, x_j)$$

le Hamiltonien et

$$Z_N := \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \exp(-\beta_N H_N(x_1, \dots, x_N)) dx_1 \cdots dx_N < \infty,$$

la fonction de partition, on sait que P_N est la mesure sur $(\mathbb{R}^d)^N$ définie par

$$dP_N(x_1, \dots, x_N) := \frac{\exp(-\beta_N H_N(x_1, \dots, x_N))}{Z_N} dx_1 \cdots dx_N. \quad (5.1)$$

On cherche à étudier le **comportement asymptotique** de P_N quand le **nombre de particules** tend vers l'infini. Les symétries du modèle suggèrent d'étudier le comportement sous cette mesure de probabilité de la mesure empirique $\mu_N := \frac{1}{N} \sum_{i=1}^N \delta_{x_i}$. En effet, l'énergie d'une configuration ne dépend que de cette mesure :

$$\begin{aligned} H_N(x_1, \dots, x_N) &= \frac{1}{N} \sum_{i=1}^N V(x_i) + \frac{1}{N^2} \sum_{i<j} W(x_i, x_j) \\ &= \int V(x) d\mu_N(x) + \frac{1}{2} \iint_{\neq} W(x, y) d\mu_N(x) d\mu_N(y) \end{aligned}$$

où l'indice « $\neq$ » indique qu'on intègre hors-diagonale. En notant $\mathcal{M}_1(\mathbb{R}^d)$ l'ensemble des mesures de probabilité sur $\mathbb{R}^d$, l'expression de H_N suggère de définir une fonctionnelle limite $I : \mathcal{M}_1(\mathbb{R}^d) \rightarrow (-\infty, +\infty]$ par

$$I(\mu) := \frac{1}{2} \iint (V(x) + V(y) + W(x, y)) d\mu(x) d\mu(y).$$

L'énergie H_N est alors « presque » égale à $I(\mu_N)$ (si l'on ne prend pas en compte les termes infinis sur la diagonale). Quand le nombre de particules est grand, le système devrait donc **favoriser les configurations de faible énergie**. Nous quantifions ce résultat par un **principe de grande déviations** et une **étude des minimiseurs de I** .

Là encore, les cas de la dimension 1 et 2 sont très étudiés dans la littérature, en particulier pour leurs liens avec des modèles de matrices aléatoires (nous renvoyons à l'article [3] pour de très nombreuses références).

b) Grandes déviations

Le premier résultat que nous obtenons est un principe de grandes déviations, un équivalent en dimension supérieure à 3 de théorèmes connus en théorie des matrices aléatoires ([BAG97, BAZ98, PH98, Har12]). On munit $\mathcal{M}_1(\mathbb{R}^d)$ de la topologie faible, et pour tout $A \subset \mathcal{M}_1(\mathbb{R}^d)$ on note $\text{int}(A)$, $\text{clo}(A)$ l'intérieur et la fermeture de A .

Théorème 5.1.1 (Grandes déviations). *Si $\beta_N \gg N \log(N)$, et sous certaines hypothèses sur V et W , la loi de μ_N sous la mesure P_N suit un principe de grandes déviations, à vitesse β_N , pour la fonction de taux*

$$I_\star := I - \inf_{\mathcal{M}_1(\mathbb{R}^d)} I.$$

Plus précisément pour tout borélien $A \subset \mathcal{M}_1(\mathbb{R}^d)$,

$$\begin{aligned} - \inf_{\mu \in \text{int}(A)} I_\star(\mu) &\leq \liminf_{N \rightarrow \infty} \frac{\log P_N(\mu_N \in A)}{\beta_N} \\ &\leq \limsup_{N \rightarrow \infty} \frac{\log P_N(\mu_N \in A)}{\beta_N} \leq - \inf_{\mu \in \text{clo}(A)} I_\star(\mu). \end{aligned}$$

Remarque 5.1.2. Les hypothèses de ce théorème sont relativement faibles, et couvrent des cas beaucoup plus généraux que les interactions de type coulombien.

La preuve de ce résultat suit un schéma classique ([DZ10, Chap. 4]) : on montre successivement, pour les mesures non-normalisées $Z_N Q_N$, une propriété de tension exponentielle, une borne supérieure sur les boules ouvertes¹, puis une borne inférieure de grandes déviations. Comme souvent, c'est ce dernier point qui est le plus technique, et qui demande des arguments de régularisation.

c) Propriétés de la mesure d'équilibre

Si l'on considère un potentiel d'interaction de Coulomb, on peut montrer que la fonctionnelle $I_\star$ est minimisée par une unique mesure, et en donner une caractérisation. Notons, pour $x \in \mathbb{R}^d$,

$$U^\mu(x) := \int W(x - y) d\mu(y)$$

le **potentiel** (électrique) créé par la mesure μ . Nous montrons une caractérisation du minimiseur par le potentiel qu'il crée, donnant ainsi l'analogie de résultats de Frostman ([Fro35]) dans le cas compact sans potentiel extérieur V , et de Dragnev et Saff sur les sphères ([DS07]).

Théorème 5.1.3 (Gaz de Riesz). *Pour $W(x, y)$ le potentiel de Coulomb,*

1. *la fonctionnelle I est strictement convexe là où elle est définie,*
2. *elle admet un unique minimiseur $\mu_\star$,*

$$I(\mu_\star) = \inf_{\mu \in \mathcal{M}_1(\mathbb{R}^d)} I(\mu);$$

3. *Si les $(\mu_N)_N$ sont définis sur un même espace de probabilité, et si $\beta_N \gg N \log(N)$, alors μ_N tend presque sûrement vers $\mu_\star$;*
4. *le minimiseur $\mu_\star$ est à support compact ;*
5. *Si une mesure μ à support compact vérifie*

$$U^\mu + V \leq C \quad \text{sur } \text{Supp}(\mu), \quad (5.2)$$

$$U^\mu + V \geq C \quad \text{quasi-partout } \text{Supp}(\mu_\star), \quad (5.3)$$

pour une certaine constante C alors $\mu = \mu_\star$.

Remarque 5.1.4. L'article original [3] établit ce résultat dans le cadre plus général des potentiels dits de Riesz.

1. Les boules sont prises au sens de la distance de Fortet–Mourier, qui métrise la convergence faible.

Corollaire 5.1.5 (Gaz de Coulomb pour un potentiel radial). *Si V est quadratique et W de type Coulombien, la mesure d'équilibre $\mu_\star$ est la mesure uniforme sur une boule.*

Ce corollaire, présent dans [LG10, Proposition 2.13], est l'analogie de la « loi du cercle » pour l'ensemble complexe de Ginibre (voir par exemple [BAZ98, PH98]).

Dans le cas d'un potentiel de Riesz, nous répondons aussi à une question inverse : étant donné une **mesure cible** μ à support compact, absolument continue, de densité dans L^p pour $p > d/\alpha$, on peut **construire** explicitement un **potentiel** de confinement V tel que μ soit l'unique mesure minimisant I .

Les preuves de ces résultats utilisent des outils et propriétés classiques de théorie du potentiel.

Parmi les nombreuses ouvertures possibles mentionnées dans [3], il paraît particulièrement intéressant, en lien avec les chapitres précédents, de chercher une bonne manière de **simuler** les mesures d'équilibre μ_N . Dans le cas $d = 2$, $W(x, y) = -2 \log |x - y|$, le processus est déterminantal et on dispose d'un algorithme exact ([HKPV06, Algorithme 18], et [SZT09]). Dans le cas général le problème est plus difficile. On peut cependant espérer utiliser un processus pour lequel P_N est invariant (l'ÉDS de départ ou une de ses variantes) et l'utiliser pour une méthode de Monte-Carlo, dans l'esprit des algorithmes « MALA » (*Metropolis adjusted Langevin approach*, [RR01]) ou « HMC » (*Hamiltonian Monte-Carlo*, [Nea11]), une méthode dont les liens avec les PDMPs mériteraient approfondissement).

Ces questions sont reliées au problème extrêmement riche de la recherche de points « bien répartis » sur la sphère, connu (suivant les auteurs et les définitions de « bien répartis ») comme problème de Thomson, problème de Tammes, recherche de points de Fekete, recherche de *spherical codes*, etc. : nous renvoyons à [SK97, HS04] pour une introduction au sujet. Une des façons pour une configuration d'être « bien répartie » est de minimiser l'interaction $\sum W(x_i, x_j)$ pour un W bien choisi (Coulombien, de Riesz, ...).

Problème ouvert 5.1.6 (Simulation). *Quels algorithmes mettre en œuvre pour simuler efficacement P_N ? Peut-on coupler cette approche avec un « refroidissement » pour trouver des points bien répartis sur la sphère via un algorithme de type recuit simulé ?*

Cette piste, évoquée dans [Nur95], ne semble pas avoir été développée, en particulier du point de vue théorique.

5.2 Un modèle d'évolution

Nous décrivons ici les résultats de [1], obtenus en collaboration avec L. Goudenège.

a) Génétique des populations — le modèle de Wright-Fisher

La **génétique des populations** vise à étudier les changements de fréquence des gènes au cours du temps, en fonction de la sélection naturelle, de la « dérive génétique » (les effets aléatoires) et des mutations. Nous considérons ici le cas le plus simple, où un gène possède deux allèles (par convention « noir » et « blanc »), et l'on étudie l'évolution au cours du temps de la proportion d'individus portant l'allèle « blanc ».

Si l'allèle n'a aucun effet sur le phénotype des individus, la situation peut être décrite par le modèle classique de Wright-Fisher. Ce modèle est suffisamment simple pour que l'on puisse l'analyser très en détails mathématiquement (cf. par exemple les chapitres correspondants dans [Dur08] ou [Eth11]). Rappelons son principe. On considère une population de taille N , et on note la proportion $X_k^N \in \{0, 1/N, \dots, 1\}$ d'allèles « blanc » à la génération k . Pour passer d'une génération à l'autre on procède en deux étapes :

1. Chaque individu présent à la génération k pond un très grand nombre M d'œufs, du même type que lui (blanc ou noir).
2. Parmi ces œufs, N sont choisis uniformément pour donner la génération suivante.

Comme M est supposé très grand, ceci revient à dire que chaque individu de la génération $k + 1$ choisit au hasard uniformément un parent parmi les n individus présents à la génération k , et copie son type. Ce modèle peut être enrichi pour prendre en compte la sélection² : la façon usuelle de le faire est de dire que seule une certaine fraction s_w des œufs blancs (respectivement une fraction s_b des œufs noirs) survivent jusqu'à la maturité.

La question la plus naturelle dans ce modèle est de trouver, partant d'une proportion x d'individus « blancs », la probabilité que l'allèle correspondant envahisse la population. On peut ensuite se demander quel temps est nécessaire pour que l'invasion ait lieu.

Quand la population devient grande, il n'y a pas de formule close pour ces quantités (sauf dans le cas facile sans sélection). On utilise typiquement des formules approchées, justifiées par une **approximation diffusiv** : d'une manière un peu analogue au résultat classique de Donsker, on montre que, dans la **limite de sélection faible** où le rapport des probabilités de survie s_w/s_b s'écrit $1 + \beta/N$, la suite de processus (Y^N) définie par $Y_t^N = X_{[tN]}^N$ converge vers une diffusion sur $[0, 1]$ totalement explicite, gouvernée par l'ÉDS $dX_t = \sqrt{a(x)}dB_t + b(X_t)dt$ de générateur $L = \frac{1}{2}a(x)\partial_{xx} + b(x)\partial_x$, où

$$\begin{cases} a(x) = x(1-x) \\ b(x) = \beta x(1-x). \end{cases}$$

Remarque 5.2.1 (Sélection). Le terme de sélection, et son échelle en $1/N$, semble ici très arbitraire. Une façon de comprendre ce résultat est que cette échelle est la bonne pour qu'on observe encore à la limite une compétition entre le bruit (le terme brownien, également appelé dérive génétique) et la sélection (le terme de dérive de la diffusion).

Un des intérêts du modèle étudié dans [1] est qu'il fait apparaître « naturellement » l'équivalent d'un **terme de sélection dépendant de la fréquence**, à partir d'une description directe de l'interaction entre les individus de la population.

b) La question biologique : évolution des comportements nuisibles

Dans de nombreux contextes biologiques, des individus agissent d'une manière qui nuit aux autres sans donner d'avantage direct. Par exemple, les mâles de certaines espèces d'invertébrés ont une réserve spermatique limitée ; de façon surprenante, on a observé qu'ils continuaient à tenter de se reproduire avec des femelles alors que leur réserve était épuisée [DB06, SHR08]. Ce comportement n'apporte aucun bénéfice direct en terme de fertilisation. Toutefois, pour ces espèces, la copulation fait décroître ou stoppe la réceptivité sexuelle de la femelle : le comportement nuisible semble donc pouvoir être avantageux dans la mesure où l'individu empêche ses compétiteurs de produire des œufs, et augmente donc le nombre **relatif** de ses propres œufs.

Pour mieux comprendre ce type de comportement et tenter de quantifier l'avantage créé par la nuisance, nous étudions un modèle très simple d'une telle interaction, à partir de laquelle nous définissons un modèle de type Wright-Fisher, mais qui fait apparaître un terme sélectif dont l'intensité dépend de la fréquence.

2. On peut aussi introduire des mutations mais nous ne considérerons pas ce cas ici.

Le modèle élémentaire, mis au point par F.-X. Dechaume-Moncharmont et M. Galipaud³, est le suivant. On considère une urne contenant w boules blanches et b boules noires, représentant les différents mâles disponibles. Pendant une saison de reproduction, f femelles viennent successivement tenter de se reproduire : ces tentatives sont représentées par des tirages successifs dans l'urne. À chaque tirage :

- Si la boule est blanche, la reproduction réussit **et la boule est retirée de l'urne** (le mâle ne cherche plus à se reproduire).
- Si la boule est noire, **on la remet dans l'urne**. La reproduction réussit si c'est la première fois que cette boule est tirée, elle échoue sinon.

Pour comparer les stratégies, notons $p_w(w, b, f)$ (resp. $p_b(w, b, f)$) les probabilités d'accès à la reproduction (sur l'ensemble de la saison) pour une boule blanche (resp. noire) : ces quantités peuvent être vues comme une quantification de la *fitness* respective des deux allèles. On peut relativement facilement montrer que $p_w < p_b$, on peut alors se demander quel est l'ordre de grandeur de la différence $p_b - p_w$. Nous avons pu montrer que cette différence est de l'ordre de $1/N$. Nous obtenons en fait un résultat beaucoup plus précis pour la limite de grande population où w , b et f tendent vers l'infini en gardant leur rapport constant.

Théorème 5.2.2 (Limite de grande population, énoncé informel). *Il existe des fonctions v et $\tilde{v}$, quasi-explicites, telles que, si w , b , f tendent vers l'infini avec $N = w + b + f$ et $\frac{1}{N}(w, b, f) \rightarrow (x, y, z)$,*

$$\begin{aligned} p_b(w, b, f) &= v(x, y, z) + O(1/N), \\ p_w(w, b, f) &= v(x, y, z) + O(1/N), \\ p_b(w, b, f) - p_w(w, b, f) &= \frac{1}{N} \tilde{v}(x, y, z) + O(1/N^2). \end{aligned}$$

La preuve de ce résultat s'appuie sur l'interprétation d'une **relation de récurrence** sur p_b comme **schéma numérique** pour une ÉDP linéaire simple ; la fonction limite v est la solution de cette ÉDP, et $\tilde{v}$ en est une certaine dérivée directionnelle. La preuve est rendue assez technique par la nécessité d'obtenir des contrôles sur les restes suffisamment uniformes pour pouvoir justifier l'approximation diffusive que nous présentons plus loin.

c) Sélection « indirecte » dans Wright–Fisher

Reprenons le modèle de Wright–Fisher d'une population de n mâles porteurs de l'allèle « blanc » ou « noir ». Pour simplifier, fixons un sex-ratio en supposant qu'il y a toujours $s \cdot n$ femelles. À chaque saison de reproduction :

1. Les f_n femelles choisissent un partenaire suivant le modèle de tirage décrit précédemment : les reproductions couronnées de succès produisent un très grand nombre M d'œufs, tous du même type que le père.
2. Parmi tous les œufs restants, on en choisit n uniformément pour former la génération suivante (de mâles).

On obtient ainsi une chaîne de Markov sur $\{0, 1/n, \dots, 1\}$, que l'on peut renormaliser comme précédemment. Notre résultat principal est que la chaîne renormalisée **converge vers une diffusion** de générateur

3. Communication personnelle.

$L = \frac{1}{2}a(x)\partial_{xx} + b(x)\partial_x$, où

$$\begin{cases} a(x) = \frac{x(1-x)}{v_s(x)}, \\ b(x) = -x(1-x) \frac{v'_s(x)}{v_s^2(x)}. \end{cases}$$

et la fonction v_s est directement reliée à la fonction v du théorème précédent par une formule simple.

Le caractère explicite de v permet de mettre en évidence les points suivants :

1. la variance est multipliée par un facteur $(1/v_s(x)) > 1$ par rapport au modèle classique de Wright-Fisher, en raison de l'étape aléatoire supplémentaire de choix des mâles.
2. La dérive effective $2b/a = -2v'_s(x)/v_s(x)$, qui caractérise les probabilités d'absorption en 0 et 1, est toujours négative, ce qui pousse le processus vers 0 — l'allèle nuisible est favorisé — mais sa force dépend de la proportion x d'individus non-nuisibles dans la population.
3. Si l'on rajoute un effet de sélection faible contre l'allèle nuisible, on obtient pour certaines valeurs des paramètres une dérive dont le signe change : par exemple, l'allèle nuisible peut être favorisé quand il est peu présent, et défavorisé quand il est très présent. Certains choix judicieux de paramètres donnent même deux changements de signe.

L'analyse de ces caractéristiques du point de vue biologique est un travail en cours avec F-X. Dechaume-Moncharmont, M. Galipaud et L. Goudenège.

Citons enfin une des pistes possibles ouvertes par ce travail.

Problème ouvert 5.2.3. *Dans le modèle élémentaire, le comportement nuisible réduit le nombre total de fécondations, certaines femelles étant « privées » de reproduction. Le fait de réguler la taille de la population dans le modèle multi-génération de type Wright-Fisher compense cet effet et empêche la population de s'éteindre.*

Ceci empêche de se poser dans ce modèle la question du « suicide évolutif », qui correspondrait ici à une invasion de la population par l'allèle nuisible, qui conduirait ensuite à l'extinction de la population en raison de la perte de fécondité totale.

Peut-on construire puis analyser un modèle mettant ce phénomène en évidence ?

Publications

- [1] Avec L. Goudenège : *A Wright-Fisher model with indirect selection*, 2013, à paraître dans Journal of Mathematical Biology.
Prépublication hal-00925744 — arXiv : 1401.4040
- [2] Avec M. Benaïm, S. Le Borgne, F. Malrieu : On the stability of planar randomly switched systems, Ann. Appl. Probab. **24**(1/2014), 292–311.
Prépublication hal-00686271 — arXiv : 1204.1921
- [3] Avec D. Chafaï, N. Gozlan : *First order global asymptotics for confined particles with singular pair repulsion*, 2013, à paraître dans Annals of Applied Probability.
Prépublication hal-00818472 — arXiv : 1304.7569
- [4] Avec M. Benaïm, S. Le Borgne, F. Malrieu : *Qualitative properties of certain piecewise deterministic Markov processes*, 2012, à paraître dans Annales de l'IHP.
Prépublication hal-00688920 — arXiv : 1204.4143
- [5] Avec R. Azaïs, J.-B. Bardet, A. Génadot, N. Krell : Piecewise deterministic Markov process — recent results, ESAIM : Proc. **44** (2014), 276-290.
Prépublication hal-00865092 — arXiv : 1309.6061
- [6] Avec J.-B. Bardet, A. Christen, A. Guillin, F. Malrieu : Total variation estimates for the TCP process, Electronic Journal of Probability **18**(10/2013), 1-21.
Prépublication hal-00655462 — arXiv : 1112.6298
- [7] Avec M. Benaïm, S. Le Borgne, F. Malrieu : Quantitative ergodicity for some switched dynamical systems, Electron. Commun. Probab. **17**(56/2012), 1–13.
Prépublication hal-00686272 — arXiv : 1204.1922
- [8] Avec H. Cardot, P. Cénac : Recursive estimation of the conditional geometric median in Hilbert spaces, Electronic Journal of Statistics **6** (2012), 2535-2562.
Prépublication hal-00687762 — arXiv : 1204.3213
- [9] Avec H. Cardot, P. Cénac : Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm., Bernoulli **19**(1), 18-43.
Prépublication hal-00558481 — arXiv : 1101.4316
- [10] Avec P. Cattiaux, A. Guillin : Poincaré inequalities and hitting times, Annales de l'IHP **49**(1),
Prépublication hal-00550125 — arXiv : 1012.5274

- [11] Super Poincaré inequalities, Orlicz norms and essential spectrum, Potential Analysis 2010, Online First.
Prépublication hal-00426367 — arXiv : 0910.4768
- [12] Functional inequalities and uniqueness of the Gibbs measure — from log-Sobolev to Poincaré, ESAIM P& S **12** (2008), 258–272.
Prépublication hal-00130824 — arXiv : math.PR/0702403
- [13] Annealing diffusions in a potential with a slow growth, Stochastic Processes and their Applications **118**(1/2008), 76–119.
Prépublication hal-00084081 — arXiv : math.PR/0607147

Bibliographie

- [ABC⁺00] Cécile Ané, Sébastien Blachère, Djalil Chafaï, Pierre Fougères, Ivan Gentil, Florent Malrieu, Cyril Roberto, and Grégory Scheffer, *Sur les inégalités de Sobolev logarithmiques*, Panoramas et synthèses, no. 10, SMF, 2000.
- [ADPY12] Marc Arnaudon, Clément Dombry, Anthony Phan, and Le Yang, *Stochastic algorithms for computing means of probability measures*, Stochastic Process. Appl. **122** (2012), no. 4, 1437–1455. MR 2914758
- [AGZ10] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni, *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010. MR 2760897 (2011m :60016)
- [AM12] Marc Arnaudon and Laurent Miclo, *Means in complete manifolds : uniqueness and approximation*, 2012.
- [AM13a] ———, *A stochastic algorithm finding generalized means on compact manifolds*, 2013.
- [AM13b] ———, *A stochastic algorithm finding p -means on the circle*, 2013.
- [AN78] K. B. Athreya and P. Ney, *A new approach to the limit theory of recurrent Markov chains*, Trans. Amer. Math. Soc. **245** (1978), 493–501. MR 511425 (80i :60092)
- [BAG97] Gérard Ben Arous and Alice Guionnet, *Large deviations for Wigner’s law and Voiculescu’s non-commutative entropy*, Probab. Theory Related Fields **108** (1997), no. 4, 517–542. MR 1465640 (98i :15026)
- [BAZ98] Gérard Ben Arous and Ofer Zeitouni, *Large deviations from the circular law*, ESAIM Probab. Statist. **2** (1998), 123–134 (electronic). MR 1660943 (2000g :60045)
- [BBCG08] Dominique Bakry, Franck Barthe, Patrick Cattiaux, and Arnaud Guillin, *A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case*, Electron. Commun. Probab. **13** (2008), 60–66. MR 2386063 (2009d :60039)
- [BBM09] M. Balde, U. Boscain, and P. Mason, *A note on stability conditions for planar switched systems*, Internat. J. Control **82** (2009), no. 10, 1882–1888. MR 2567235 (2010i :93122)
- [BCR06] Franck Barthe, Patrick Cattiaux, and Cyril Roberto, *Interpolated inequalities between exponential and gaussian, Orlicz hypercontractivity and application to isoperimetry*, Rev. Mat. Iberoamericana **22** (2006), no. 3, 993–1067.
- [BGL14] Dominique Bakry, Ivan Gentil, and Michel Ledoux, *Analysis and geometry of Markov diffusion operators*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 348, Springer, Cham, 2014. MR 3155209

- [BH99] Thierry Bodineau and Bernard Helffer, *Correlations, spectral gaps and log-Sobolev inequalities for unbounded spins systems*, Differential equations and mathematical physics (Birmingham), International Press, 1999, pp. 27–42.
- [BH12] Y. Bakhtin and T. Hurth, *Invariant densities for dynamical systems with random switching*, Nonlinearity **25** (2012), no. 10, 2937–2952. MR 2979976
- [BKKP05] O. Boxma, H. Kaspi, O. Kella, and D. Perry, *On/Off Storage Systems with State-Dependent Input, Output and Switching Rates*, Probability en the Engineering and Informational Siences **19** (2005), 1–14.
- [BM02] Thierry Bodineau and F. Martinelli, *Some new results on the kinetic ising model in a pure phase*, J. Statist. Phys. **109** (2002), no. 1-2, 207–235.
- [Bor08] Vivek S. Borkar, *Stochastic approximation*, Cambridge University Press, Cambridge ; Hindustan Book Agency, New Delhi, 2008, A dynamical systems viewpoint. MR 2442439 (2009j :60001)
- [BR03] Franck Barthe and Cyril Roberto, *Sobolev inequalities for probability measures on the real line*, Studia Math. **159** (2003), no. 3, 481–497, Dedicated to Professor Aleksander Pełczyński on the occasion of his 70th birthday (Polish). MR MR2052235 (2006c :60019)
- [BR11] E. Buckwar and M. G. Riedler, *An exact stochastic hybrid model of excitable membranes including spatio-temporal evolution*, J. Math. Biol. **63** (2011), no. 6, 1051–1093. MR 2855804 (2012h :92075)
- [CCM10] Hervé Cardot, Peggy Cénac, and Jean-Marie Monnez, *Fast clustering of large datasets with sequential k-medians : a stochastic gradient approach*, Tech. report, Institut de Mathématiques de Bourgogne, 2010.
- [CD08] O. L. V. Costa and F. Dufour, *Stability and ergodicity of piecewise deterministic Markov processes*, SIAM J. Control Optim. **47** (2008), no. 2, 1053–1077.
- [CG09] Patrick Cattiaux and Arnaud Guillin, *Functional inequalities via lyapunov conditions*, Proceedings of the summer school on Optimal Transport, Grenoble, To appear Cambridge Univ. Press, 2009, available at <http://perso.math.univ-toulouse.fr/cattiaux/files/2013/11/cg-procgrenoble-revised.pdf>.
- [CGGR10] Patrick Cattiaux, Nathael Gozlan, Arnaud Guillin, and Cyril Roberto, *Functional inequalities for heavy tailed distributions and application to isoperimetry*, Electron. J. Probab. **15** (2010), no. 13, 346–385. MR 2609591 (2011f :60036)
- [CGWW09] Patrick Cattiaux, Arnaud Guillin, Feng-Yu Wang, and Liming Wu, *Lyapunov conditions for super Poincaré inequalities*, J. Funct. Anal. **256** (2009), no. 6, 1821–1841. MR 2498560 (2011g :46059)
- [CK83] René Carmona and Abel Klein, *Exponential moments for hitting times of uniformly ergodic Markov processes*, Ann. Probab. **11** (1983), no. 3, 648–655. MR 704551 (85c :60107)
- [Clo11] Bertrand Cloez, *Limit theorems for some branching measure-valued processes*, preprint, arXiv :1106.0660, 2011.
- [CMP10] Djalil Chafaï, Florent Malrieu, and Katy Paroux, *On the long time behavior of the TCP window size process*, Stochastic Process. Appl. **120** (2010), no. 8, 1518–1534. MR 2653264

- [Dav84] M. H. A. Davis, *Piecewise-deterministic Markov processes : a general class of nondiffusion stochastic models*, J. Roy. Statist. Soc. Ser. B **46** (1984), no. 3, 353–388, With discussion. MR MR790622 (87g :60062)
- [Dav93] ———, *Markov models and optimization*, Monographs on Statistics and Applied Probability, vol. 49, Chapman & Hall, London, 1993.
- [DB06] David Damiens and Guy Boivin, *Why do sperm-depleted parasitoid males continue to mate?*, Behavioral Ecology **17** (2006), no. 1, 138–143.
- [DC99] F. Dufour and O. L. V. Costa, *Stability of piecewise-deterministic Markov processes*, SIAM J. Control Optim. **37** (1999), no. 5, 1483–1502 (electronic).
- [DD02] F. Dufour and Y. Dutuit, *Dynamic reliability : A new model*, Proceedings of ESREL 2002 Lambda-Mu 13 Conference, 2002, pp. 350–353.
- [DF90] Persi Diaconis and James Allen Fill, *Strong stationary times via a new form of duality*, Ann. Probab. **18** (1990), no. 4, 1483–1522. MR 1071805 (91m :60127)
- [DF99] P. Diaconis and D. Freedman, *Iterated random functions*, SIAM Rev. **41** (1999), no. 1, 45–76. MR 1669737 (2000c :60102)
- [DGR02] V. Dumas, F. Guillemin, and Ph. Robert, *A Markovian analysis of additive-increase multiplicative-decrease algorithms*, Adv. in Appl. Probab. **34** (2002), no. 1, 85–111. MR MR1895332 (2003f :60168)
- [DHKR12] Marie Doumic, Marc Hoffmann, Nathalie Krell, and Lydia Robert, *Statistical estimation of a growth-fragmentation model observed on a genealogical tree*, preprint, arXiv :1210.3240, to appear in Bernoulli, 2012.
- [Dia88] Persi Diaconis, *Group representations in probability and statistics*, Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11, Institute of Mathematical Statistics, Hayward, CA, 1988. MR 964069 (90a :60001)
- [DMT95] D. Down, S. P. Meyn, and R. L. Tweedie, *Exponential and uniform ergodicity of Markov processes*, Ann. Probab. **23** (1995), no. 4, 1671–1691. MR 1379163 (97c :60181)
- [DS07] P. D. Dragnev and E. B. Saff, *Riesz spherical potentials with external fields and minimal energy points separation*, Potential Anal. **26** (2007), no. 2, 139–162. MR 2276529 (2007k :31004)
- [Duf97] Marie Duflo, *Random iterative models*, Applications of Mathematics (New York), vol. 34, Springer-Verlag, Berlin, 1997, Translated from the 1990 French original by Stephen S. Wilson and revised by the author. MR MR1485774 (98m :62239)
- [Dur08] Richard Durrett, *Probability models for DNA sequence evolution*, second ed., Probability and its Applications (New York), Springer, New York, 2008. MR 2439767 (2009k :60001)
- [DZ10] Amir Dembo and Ofer Zeitouni, *Large deviations techniques and applications*, Stochastic Modelling and Applied Probability, vol. 38, Springer-Verlag, Berlin, 2010, Corrected reprint of the second (1998) edition. MR 2571413 (2011b :60094)
- [EO05] R. Erban and H. G. Othmer, *From individual to collective behavior in bacterial chemotaxis*, SIAM J. Appl. Math. **65** (2004/05), no. 2, 361–391 (electronic). MR 2123062 (2005j :35220)

- [Eth11] Alison Etheridge, *Some mathematical models from population genetics*, Lecture Notes in Mathematics, vol. 2012, Springer, Heidelberg, 2011, Lectures from the 39th Probability Summer School held in Saint-Flour, 2009. MR 2759587 (2011j :92043)
- [Fro35] Otto Frostman, *Potentiel d'Équilibre et Capacité des Ensembles*, Ph.D. thesis, Faculté des sciences de Lund, 1935.
- [GL10] N. Gozlan and C. Léonard, *Transport inequalities. A survey*, Markov Process. Related Fields **16** (2010), no. 4, 635–736. MR 2895086
- [GM13] Sébastien Gadat and Laurent Miclo, *Spectral decompositions and L^2 -operator norms of toy hypocoercive semi-groups*, Kinet. Relat. Models **6** (2013), no. 2, 317–372. MR 3030715
- [GR09] C. Graham and Ph. Robert, *Interacting multi-class transmissions in large stochastic networks*, Ann. Appl. Probab. **19** (2009), no. 6, 2334–2361.
- [GR11] ———, *Self-adaptive congestion control for multiclass intermittent connections in a communication network*, Queueing Syst. **69** (2011), no. 3-4, 237–257. MR 2886470
- [Gri78] D. Griffeath, *Coupling methods for Markov processes*, Studies in probability and ergodic theory, Adv. in Math. Suppl. Stud., vol. 2, Academic Press, New York-London, 1978, pp. 1–43. MR 517252 (80k :60080)
- [GRZ04] F. Guillemin, Ph. Robert, and B. Zwart, *AIMD algorithms and exponential functionals*, Ann. Appl. Probab. **14** (2004), no. 1, 90–117. MR MR2023017 (2004m :60216)
- [GW02] Fu-Zhou Gong and Feng-Yu Wang, *Functional inequalities for uniformly integrable semigroups and application to essential spectrums*, Forum Math. **14** (2002), no. 2, 293–313. MR MR1880915 (2003a :47097)
- [Hal48] J. B. S. Haldane, *Note on the median of a multivariate distribution*, Biometrika **35** (1948), no. 3-4, 414–417.
- [Har12] Adrien Hardy, *A note on large deviations for 2D Coulomb gas with weakly confining potential*, Electron. Commun. Probab. **17** (2012), no. 19, 12. MR 2926763
- [HCS87] Chii-Ruey Hwang, Tzue-Shuh Chiang, and Shuenn-Jyi Sheu, *Diffusion for global optimization in $\mathbb{R}^n$* , Siam Journal on Control and Optimization **25** (1987), 737–753.
- [HKPV06] J. Ben Hough, Manjunath Krishnapur, Yuval Peres, and Bálint Virág, *Determinantal processes and independence*, Probab. Surv. **3** (2006), 206–229. MR 2216966 (2006m :60068)
- [HKS89] Richard A. Holley, Shigeo Kusuoka, and Daniel W. Stroock, *Asymptotics of the spectral gap with applications to the theory of simulated annealing*, Journal of Functional Analysis **83** (1989), no. 2, 333–347. MR MR995752 (92d :60081)
- [HS88] Richard A. Holley and Daniel W. Stroock, *Simulated annealing via Sobolev inequalities*, Communications in Mathematical Physics **115** (1988), no. 4, 553–569. MR MR933455 (90g :60091)
- [HS04] D. P. Hardin and E. B. Saff, *Discretizing manifolds via minimum energy points*, Notices Amer. Math. Soc. **51** (2004), no. 10, 1186–1194. MR 2104914 (2006a :41049)
- [Jak88] Adam Jakubowski, *Tightness criteria for random measures with application to the principle of conditioning in Hilbert spaces*, Probab. Math. Statist. **9** (1988), no. 1, 95–114. MR 945679 (89h :60003)

- [Kem87] J. H. B. Kemperman, *The median of a finite measure on a Banach space*, Statistical data analysis based on the L_1 -norm and related methods (Neuchâtel, 1987), North-Holland, Amsterdam, 1987, pp. 217–230. MR 949228 (89h :46032)
- [KGV83] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, *Optimization by simulated annealing*, Science **220** (1983), no. 4598, 671–680. MR MR702485 (85f :90091)
- [KW52] J. Kiefer and J. Wolfowitz, *Stochastic estimation of the maximum of a regression function*, Ann. Math. Statistics **23** (1952), 462–466. MR 0050243 (14,299e)
- [KY03] Harold J. Kushner and G. George Yin, *Stochastic approximation and recursive algorithms and applications*, second ed., Applications of Mathematics (New York), vol. 35, Springer-Verlag, New York, 2003, Stochastic Modelling and Applied Probability. MR MR1993642 (2004e :62005)
- [Led01] Michel Ledoux, *Logarithmic Sobolev inequalities for unbounded spin systems revisited*, Séminaire de Probabilités, XXXV, Lecture Notes in Math., vol. 1755, Springer, Berlin, 2001, pp. 167–194. MR 2002g :82001
- [LG10] A. López García, *Greedy energy points with external fields*, Recent trends in orthogonal polynomials and approximation theory, Contemp. Math., vol. 507, Amer. Math. Soc., Providence, RI, 2010, pp. 189–207. MR 2647570 (2011i :31005)
- [Lin92] T. Lindvall, *Lectures on the coupling method*, Wiley Series in Probability and Mathematical Statistics : Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1992, A Wiley-Interscience Publication. MR 1180522 (94c :60002)
- [LLL11] Eva Löcherbach, Dasha Loukianova, and Oleg Loukianov, *Polynomial bounds in the ergodic theorem for one-dimensional diffusions and integrability of hitting times*, Ann. Inst. Henri Poincaré Probab. Stat. **47** (2011), no. 2, 425–449. MR 2814417 (2012d :60099)
- [LLS11] Oleg Loukianov, Dasha Loukianova, and Shiqi Song, *Spectral gaps and exponential integrability of hitting times for linear diffusions*, Ann. Inst. Henri Poincaré Probab. Stat. **47** (2011), no. 3, 679–698. MR 2841071 (2012g :60238)
- [LMR14] Sean D. Lawley, Jonathan C. Mattingly, and Michael C. Reed, *Sensitivity to switching rates in stochastically switched ODEs*, Commun. Math. Sci. **12** (2014), no. 7, 1343–1352. MR 3210750
- [LO00] Rafał Łatała and Krzysztof Oleszkiewicz, *Between Sobolev and Poincaré*, Geometric aspects of functional analysis, Lecture Notes in Math., vol. 1745, Springer, Berlin, 2000, pp. 147–168. MR MR1796718 (2002b :60025)
- [LP09] P. Laurençot and B. Perthame, *Exponential decay for the growth-fragmentation/cell-division equation*, Commun. Math. Sci. **7** (2009), no. 2, 503–510. MR 2536450 (2010e :45010)
- [LPW09] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov chains and mixing times*, American Mathematical Society, Providence, RI, 2009, With a chapter by James G. Propp and David B. Wilson. MR 2466937 (2010c :60209)
- [Mic92] Laurent Miclo, *Recuit simulé sur $\mathbb{R}^n$. Étude de l'évolution de l'énergie libre*, Annales de l'Institut Henri Poincaré **28** (1992), no. 2, 235–266.
- [Mic10] ———, *On absorption times and Dirichlet eigenvalues*, ESAIM Probab. Stat. **14** (2010), 117–150. MR 2654550 (2011i :60130)

- [MM13] L. Miclo and P. Monmarché, *Étude spectrale minutieuse de processus moins indécis que les autres*, Séminaire de Probabilités, XLV, Lecture Notes in Math., vol. 2078, Springer, Berlin, 2013, pp. 459–481.
- [Mon14] Pierre Monmarché, *On χ^1 and entropic convergence for contractive pdmp*, 2014.
- [MT93] S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*, Communications and Control Engineering Series, Springer-Verlag London Ltd., London, 1993.
- [Neal11] Radford M. Neal, *MCMC using Hamiltonian dynamics*, Handbook of Markov chain Monte Carlo, Chapman & Hall/CRC Handb. Mod. Stat. Methods, CRC Press, Boca Raton, FL, 2011, pp. 113–162. MR 2858447
- [Num78] E. Nummelin, *A splitting technique for Harris recurrent Markov chains*, Z. Wahrsch. Verw. Gebiete **43** (1978), no. 4, 309–318. MR 0501353 (58 #18732)
- [Num84] ———, *General irreducible Markov chains and nonnegative operators*, Cambridge Tracts in Mathematics, vol. 83, Cambridge University Press, Cambridge, 1984. MR 776608 (87a :60074)
- [Nur95] Kari J. Nurmela, *Constructing spherical codes by global optimization methods*, Research Report A32, Helsinki University of Technology, Department of Computer Science and Engineering, Digital Systems Laboratory, Espoo, Finland, February 1995.
- [ODA88] H. G. Othmer, S. R. Dunbar, and W. Alt, *Models of dispersal in biological systems*, J. Math. Biol. **26** (1988), no. 3, 263–298. MR 949094 (90a :92064)
- [OKM96] T. J. Ott, J. H. B. Kemperman, and M. Mathis, *The stationary behavior of ideal TCP congestion avoidance*, unpublished manuscript available at <http://www.teunisott.com/>, 1996.
- [Pel00] Mariane Pelletier, *Asymptotic almost sure efficiency of averaged stochastic algorithms*, SIAM J. Control Optim. **39** (2000), no. 1, 49–72 (electronic). MR MR1780908 (2001i :62088)
- [PH98] Dénes Petz and Fumio Hiai, *Logarithmic energy as an entropy functional*, Advances in differential equations and mathematical physics (Atlanta, GA, 1997), Contemp. Math., vol. 217, Amer. Math. Soc., Providence, RI, 1998, pp. 205–221. MR 1606719 (99b :46101)
- [PJ92] B.T. Polyak and A.B. Juditsky, *Acceleration of stochastic approximation.*, SIAM J. Control and Optimization **30** (1992), 838–855.
- [PR05] B. Perthame and L. Ryzhik, *Exponential decay for the fragmentation or cell-division equation*, J. Differential Equations **210** (2005), no. 1, 155–177. MR 2114128 (2006b :35328)
- [PTW10] K. Pakdaman, M. Thieullen, and G. Wainrib, *Fluid limit theorems for stochastic hybrid systems with application to neuron models*, Adv. in Appl. Probab. **42** (2010), no. 3, 761–794. MR 2779558 (2011m :60070)
- [RM51] Herbert Robbins and Sutton Monro, *A stochastic approximation method*, Ann. Math. Statistics **22** (1951), 400–407. MR 0042668 (13,144j)
- [RMC07] O. Radulescu, A. Muller, and A. Crudu, *Théorèmes limites pour des processus de Markov à sauts. Synthèse des résultats et applications en biologie moléculaire*, Technique et Science Informatiques **26** (2007), no. 3-4, 443–469.
- [Ros02] J. S. Rosenthal, *Quantitative convergence rates of Markov chains : a simple account*, Electron. Comm. Probab. **7** (2002), 123–128 (electronic). MR 1917546 (2003m :60188)

- [Roy89] Gilles Royer, *A remark on simulated annealing of diffusion processes*, SIAM Journal on Control and Optimization **27** (1989), no. 6, 1403–1408. MR MR1022435 (91f:60102)
- [Roy99] ———, *Une initiation aux inégalités de Sobolev logarithmiques*, Cours spécialisés, no. 5, SME, 1999.
- [RR96] Gareth O. Roberts and Jeffrey S. Rosenthal, *Quantitative bounds for convergence rates of continuous time Markov processes*, Electron. J. Probab. **1** (1996), no. 9, approx. 21 pp. (electronic). MR 1423462 (97k:60198)
- [RR01] ———, *Optimal scaling for various Metropolis-Hastings algorithms*, Statist. Sci. **16** (2001), no. 4, 351–367. MR 1888450 (2003a:65006)
- [RW01] Michael Röckner and Feng-Yu Wang, *Weak Poincaré inequalities and L^2 convergence rates of Markov semigroups*, Journal of Functional Analysis **185** (2001), 564–603.
- [SHR08] Sven Steiner, Nina Henrich, and Joachim Ruther, *Mating with sperm-depleted males does not increase female mating frequency in the parasitoid *Lariophagus distinguendus**, Entomologia Experimentalis et Applicata **126** (2008), no. 2, 131–137.
- [SK97] E. B. Saff and A. B. J. Kuijlaars, *Distributing many points on a sphere*, Math. Intelligencer **19** (1997), no. 1, 5–11. MR 1439152 (98h:70011)
- [SZT09] Antonello Scardicchio, Chase Zachary, and Salvatore Torquato, *Statistical properties of determinantal point processes in high-dimensional Euclidean spaces*, Phys. Rev. E **79** (2009), no. 4, Article 041108.
- [Vil03] Cédric Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003. MR MR1964483 (2004e:90003)
- [Vil09] Cédric Villani, *Hypocoercivity*, Mem. Amer. Math. Soc. **202** (2009), no. 950, iv+141. MR 2562709 (2011e:35381)
- [VZ00] Yehuda Vardi and Cun-Hui Zhang, *The multivariate L_1 -median and associated data depth*, Proc. Natl. Acad. Sci. USA **97** (2000), no. 4, 1423–1426 (electronic). MR MR1740461 (2000i:62066)
- [Wan00a] Feng-Yu Wang, *Functional inequalities for empty essential spectrum*, J. Funct. Anal. **170** (2000), no. 1, 219–245. MR MR1736202 (2001a:58043)
- [Wan00b] ———, *Functional inequalities, semigroup properties and spectrum estimates*, Infin. Dimens. Anal. Quantum Probab. Relat. Top. **3** (2000), no. 2, 263–295. MR MR1812701 (2002b:47083)
- [Yos99] Nobuo Yoshida, *The log-Sobolev inequality for weakly coupled lattice fields*, Probability Theory and Related Fields (1999), no. 115, 1–40.
- [Yos01] ———, *The equivalence of the log-Sobolev inequality and a mixing condition for unbounded spin systems on the lattice*, Annales de l’Institut H. Poincaré **37** (2001), no. 2, 223–243.
- [YZ10] G. G. Yin and C. Zhu, *Hybrid switching diffusions*, Stochastic Modelling and Applied Probability, vol. 63, Springer, New York, 2010, Properties and applications. MR 2559912 (2010i:60226)
- [Zeg96] Bogusław Zegarliński, *The strong decay to equilibrium for the stochastic dynamics of unbounded spin systems on a lattice*, Comm. Math. Phys. **175** (1996), no. 2, 401–432. MR MR1370101 (97m:82009)