UNIVERSITÉ PARIS-EST
Pôle de recherche et d'enseignement supérieur

THÈSE

Pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITÉ PARIS-EST

ÉCOLE DOCTORALE MATHEMATIQUES ET SCIENCES ET TECHNOLOGIE DE L'INFORMATION ET DE LA COMMUNICATION

DISCIPLINE : MATHÉMATIQUES

Présentée par

Bertrand CLOEZ

Comportement asymptotique de processus avec sauts et applications pour des modèles avec branchement

---

Directeur de thèse : **Djalil CHAFAÏ**

Soutenue le vendredi 14 juin 2013
Devant le jury composé de

M. Vincent BANSAYE	École Polytechnique	Examinateur
M. Michel BENAÏM	Université de Neuchâtel	Rapporteur
M. Djalil CHAFAÏ	Université Paris-Est Marne-la-Vallée	Directeur de thèse
M. Jean-Francois DELMAS	École Nationale des Ponts et Chaussées	Examinateur
M. Andreas EBERLE	Universität Bonn	Rapporteur
M. Martin HAIRER	University of Warwick	Examinateur
M. Arnaud GUILLIN	Université de Clermont-Ferrand	Examinateur
M. Florent MALRIEU	Université de Rennes 1	Examinateur

Thèse préparée au
Laboratoire LAMA CNRS UMR 8050
Université de Paris-Est Marne-la-Vallée
5, boulevard Descartes, Champs-sur-Marne
77454 Marne-la-Vallée cedex 2, France

A Julien,
qui aurait surement fait mieux si la vie avait été différente.

Remerciements

Après trois ans passés à Marne, il est temps pour moi d'écrire les pages de ma thèse qui seront peut-être les plus lues.

Bien que la démarche soit protocolaire, je voudrais tout d'abord exprimer ma profonde gratitude à mon directeur, Djalil Chafaï, pour m'avoir fait confiance et de m'avoir encadré durant ces trois ans. J'ai eu la chance de profiter de son dynamisme, de sa bonne humeur et de sa grande culture mathématique. J'ai beaucoup apprécié la liberté qu'il m'a laissée tout en étant toujours disponible lorsque j'en avais besoin. Merci Djalil.

Je tiens aussi à remercier vivement Michel Benaïm et Andreas Eberle de m'avoir fait l'honneur de rapporter ma thèse. Je leur témoigne toute ma reconnaissance pour ce grand travail de relecture. Je remercie en particulier Michel pour l'intérêt qu'il porte à mon travail et de m'avoir proposé de travailler avec lui. Je remercie aussi Andreas pour m'avoir accueilli à Bonn une semaine.

Un grand merci à Vincent Bansaye, Jean-Francois Delmas, Martin Hairer, Arnaud Guillin et Florent Malrieu d'être venu m'écouter et de me faire l'honneur de leur présence. Tout comme les trois membres de mon jury précédemment cités, leurs travaux ont été une grande source d'inspiration pour la rédaction de ce mémoire. J'aimerai en particulier remercier Martin, qui m'a accueilli un peu moins de deux mois à Warwick. En plus de ses qualités mathématiques indiscutables, j'ai vraiment apprécié ses qualités humaines et suis heureux que mon séjour ait finit sur un travail commun. Je tiens aussi à remercier particulièrement Florent, avec qui j'ai beaucoup discuté. J'ai beaucoup appris grâce à lui et la première partie de ma thèse lui doit beaucoup. Je remercie aussi Vincent pour sa bonne humeur communicative et pour diverses discussions.

Parmi les autres chercheurs que j'ai côtoyé, je dois beaucoup à Viet Chi Tran. Après avoir été un très bon enseignant en Master 2, il m'a dirigé vers Djalil et m'a énormément aidé pour mon premier travail ; la deuxième partie de ce mémoire lui doit beaucoup. Je tiens à lui exprimer ici toute ma reconnaissance. Je tiens aussi à remercier Laurence Marsalle pour m'avoir initié aux probabilités et

Stephan De Bièvre d'avoir été flexible lors de mon master.

Ayant déjà parlé de mon séjour en Angleterre, je tiens aussi à remercier Daniel Ueltschi et Xue-Mei Li pour leur gentillesse et Yvonne Collins qui m'a été d'un grand recours pour les diverses démarches administratives. Ce séjour a été financé à l'aide d'une bourse de l'école doctorale MSTIC ; un grand merci à sa gestionnaire Sylvie Cach. Sylvie est toujours de bonne humeur et c'est avec plaisir que j'allais la voir lorsque j'en avais besoin. Je remercie également Christiane Lafargue, Audrey Patout, Laurent Marciniszyn et tout le personnel administratif que j'ai côtoyé et qui ont contribué à ce travail.

Je remercie Marie Doumic, Andreas, Clément Foucart, Martin, Olivier Hénard, Aldéric Joulin, Florent, Patricia Reynaud-Bouret, Ellen Saada, Patricia Reynaud-Bouret et Denis Talay qui m'ont donné la chance d'exposer mes travaux. De même, je remercie les membres de l'ANR manège et en particulier ceux du CMAP de m'avoir intégré parmi eux et d'avoir financé plusieurs de mes déplacements. Dans une très bonne ambiance, j'ai beaucoup appris lors de ces diverses journées ou écoles d'été. Merci aussi aux organisateurs des groupes de travail ASpro et MEV.

Je remercie encore Aldéric pour diverses discussions et surtout pour me donner la chance de travailler avec lui à Toulouse. Je remercie également Alessandra Faggionato, Philippe Robert et Denis Talay d'avoir constitué des dossiers de post-doc pour moi et Arnaud, Chi, Florent, Patricia pour les renseignements qu'ils m'ont donnés sur les dossiers d'ATER.

J'ai eu le plaisir d'enseigner pendant ma thèse et cela ne m'a pas été d'une grande charge grâce à mes collègues ; merci à Thierry Berkover, Djalil, Emmanuelle Clément, Luc le Cor, Dan Goreac Thierry Jeantheau, Omar Lazar, Victor Rabiet, Francis Ribaud et Julien Roth.

La thèse, ce n'est pas que des maths ! Je tiens à remercier le LAMA de m'avoir aussi bien accueilli, que ça soit les jeunes docteurs qui m'accompagnaient à la machine à café, comme Dan, Hassan, Julien ou Ludovic mais aussi tous mes tchos : Ali, Arnaud, Harry, Houda, Marie-Noémie, Marwa, Paolo, Pierre le normand, Pierre le libanais, Rania, Sebastian, et Xavier. En particulier, je tiens à remercier la petite Marwa qui m'a beaucoup aidé pour la rédaction et SSSébastian pour les sorties latinos. Même s'ils sont tous devenus mes amis, je tiens particulièrement à remercier Xavier et Pierre(f) qui, par leur compagnie, m'ont rendu la vie à Paris beaucoup plus agréable. Une petite pensée pour les doctorants plus vieux qui m'ont aussi accompagné dans cette aventure : Ayech, Qing, Lokman, Omar, Sebastien, Sidi-Mohamed et Victor, mais aussi les cristoliens qui ont joué le jeu du séminaire des doctorants et en particulier Eduardo, David, Jean-Maxime et Johann. Au passage, je remercie Antoine qui a co-organisé le séminaire avec moi.

Mes amis m'ont aussi soutenu durant ces trois ans ; merci Amandine, Florence, Julie, Lionel, Max, Mylène, Pierre (encore un mais il se reconnaitra), Thomas. Merci à Michaël sans qui je n'aurait peut-être pas suivi ce parcours universitaire.

Il me reste à adresser un grand merci à mes parents, à Laurent, Marie, Geneviève, Fred (à qui il faudra que j'explique que la recherche mathématique ne se résume pas qu'au mélange de jeu de cartes) et le reste de ma famille et de ma belle-famille pour leurs encouragements et leur soutien constant. En particulier, je voudrais remercier Andrée et André, qui, en plus de tout ce qu'ils ont fait pour moi, ont toujours su éveiller ma curiosité scientifique à l'aide d'une bande dessinée ou d'une explication.

Enfin, merci à ma Hbiba, Sanae, pour m'avoir aidé, rassuré, supporté, soutenu et encouragé depuis le début ; du fond du cœur, merci.

Résumé

L'objectif de ce travail est d'étudier le comportement en temps long d'un modèle de particules avec une interaction de type branchement. Plus précisément, les particules se déplacent indépendamment suivant une dynamique markovienne jusqu'au temps de branchement, où elles donnent naissance à de nouvelles particules dont la position dépend de celle de leur mère et de son nombre d'enfants. Dans la première partie de ce mémoire nous omettons le branchement et nous étudions le comportement d'une seule lignée. Celle-ci est modélisée via un processus de Markov qui peut admettre des sauts, des parties diffusives ou déterministes par morceaux. Nous quantifions la convergence de ce processus hybride à l'aide de la courbure de Wasserstein, aussi nommée courbure grossière de Ricci. Cette notion de courbure, introduite récemment par Joulin, Ollivier, et Sammer correspond mieux à l'étude des processus avec sauts. Nous établissons une expression du gradient du semigroupe des processus de Markov stochastiquement monotone, qui nous permet d'expliciter facilement leur courbure. D'autres bornes fines de convergence en distance de Wasserstein et en variation totale sont aussi établies. Dans le même contexte, nous démontrons qu'un processus de Markov, qui change de dynamique suivant un processus discret, converge rapidement vers un équilibre, lorsque la moyenne des courbures des dynamiques sous-jacentes est strictement positive. Dans la deuxième partie de ce mémoire, nous étudions le comportement de toute la population de particules. Celui-ci se déduit du comportement d'une seule lignée grâce à une formule *many-to-one*, c'est-à-dire un changement de mesure de type Girsanov. Via cette transformation, nous démontrons une loi des grands nombres et établissons une limite macroscopique, pour comparer nos résultats aux résultats déjà connus en théorie des équations aux dérivées partielles. Nos résultats sont appliqués sur divers modèles ayant des applications en biologie et en informatique. Parmi ces modèles, nous étudierons le comportement en temps long de la plus grande particule dans un modèle simple de population structurée en taille.

Mots clés : distance et courbure de Wasserstein - ergodicité geometrique - Processus à valeurs mesures - $h-$transformée de Doob

Abstract

The aim of this work is to study the long time behavior of a branching particle model. More precisely, the particles move independently from each other following a Markov dynamics until the branching event. When one of these events occurs, the particle produces some random number of individuals whose position depends on the position of its mother and her number of offspring. In the first part of this thesis, we only study one particle line and we ignore the branching mechanism. So we are interested by the study of a Markov process which can jump, diffuse or be piecewise deterministic. The long time behavior of these hybrid processes is described with the notion of Wasserstein or coarse Ricci curvature. This notion of curvature, introduced by Joulin, Ollivier and Sammer, is more appropriate for the study of processes with jumps. We establish an expression of the gradient of the Markov semigroup of stochastically monotone processes which gives the curvature of these processes. Others sharp bounds of convergence, in Wasserstein distance and total variation distance, are also established. In the same way, we prove that if a Markov process evolves according to one of finitely many underlying Markovian dynamics, with a choice of dynamics that changes at the jump times of a second Markov process, then it is exponentially ergodic, under the assumption that the mean of the curvature of the underlying dynamics is positive. In the second part of the work, we study all the population. Its behaviour can be deduced to the study of the first part using a Girsavov-type transform which is called a many-to-one formula. Using this relation, we establish a law of large numbers and a macroscopic limit, in order to compare our results to the well know results on deterministic setting. Several examples, based on biology and computer science problems, illustrate our results, including the study of the largest individual in a size-structured population model.

Keywords : Wasserstein distance - Wasserstein curvature - exponential convergence - measure-valued processes - Doob $h-$transfom

Travaux

Ce court chapitre recense les articles qui composent cette thèse. Ils sont également disponibles sur ma page internet à l'adresse suivante :

http://perso-math.univ-mlv.fr/users/cloez.bertrand

Articles soumis

- B. CLOEZ, M. HAIRER (03/2013)
  Exponential ergodicity for Markov processes with random switching
  http://arxiv.org/abs/1303.6999
- B. CLOEZ (02/2012)
  Wasserstein decay of one dimensional jump-diffusions
  http://arxiv.org/abs/1202.1259
- B. CLOEZ (06/2011)
  Limit theorems for some branching measure-valued processes
  http://arxiv.org/abs/1106.0660

Articles non soumis

- B. CLOEZ (03/2013)
  Asymptotic estimates for the largest individual in a mitosis model

Table des matières

Chapitre 0

Introduction générale

Cette thèse est composée de deux parties. L'objectif de la première partie est d'étudier certains processus de sauts tandis que la deuxième repose sur l'analyse de mesures aléatoires branchantes. Chaque partie comprend un chapitre introductif en français et deux chapitres en anglais reprenant nos travaux. Les résultats de ces deux parties ne sont pas indépendants et ont été motivés par la compréhension du comportement en temps long d'un système de particules. Dans le modèle qui nous a intéressés, les particules suivent une dynamique markovienne et donnent naissance à une ou plusieurs particules. Les nouvelles particules suivent la même dynamique que leur mère, indépendamment les unes des autres. Ce modèle est semblable aux modèles de fragmentation étudiés dans [Ber04, Ber06] ou aux modèles de diffusions branchantes étudiés dans [ABBS11, EHK10]. Cependant, les dynamiques qui nous intéressent peuvent comporter des sauts, des parties diffusives ou déterministes par morceaux. Cela rend leur étude différente de ces modèles. Si dans un premier temps on omet le branchement, le comportement d'une lignée de particules peut être décrit par un processus de Markov $(X_t)_{t\geq 0}$. Un exemple type qui nous a intéressé est le processus « TCP window size ». Ce processus modélise le célèbre « Transmission Control Protocol »utilisé pour la transmission de données sur Internet. Ses trajectoires sont déterministes et linéaires par morceaux et tout l'aléatoire est contenu dans le mécanisme de saut. La figure 1 représente une de ses trajectoires. Ce type de processus est hybride, dans le sens où il comporte des sauts et est non-constant entre ses temps de sauts. Ces processus sont liés à des équations aux dérivées partielles intégro-différentielles. Bien qu'ils soient très étudiés récemment, il reste beaucoup de questions autour de l'étude de ces processus. La première partie traite de la géométrie et de la stabilité de ces processus hybrides, la deuxième prend en compte le comportement de toute la population ; c'est-à-dire que, dans la deuxième partie, nous nous inté-

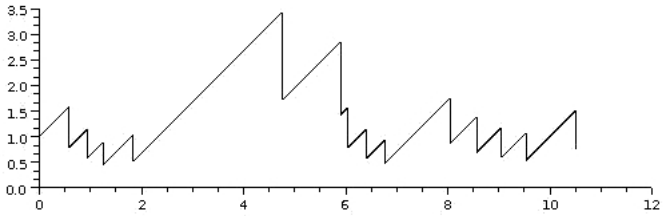

Figure 1 – Une trajectoire du processus TCP

ressons au comportement de la mesure empirique dans un modèle de de type branchement. La suite de cette introduction est aussi divisée en deux sections. La première donne les résultats principaux de la première partie et des chapitres 2 et 3 tandis que la deuxième établit les résultats principaux de la deuxième partie et des chapitres 5 et 6. Comme nous donnerons plus de détails dans les chapitres introductifs 1 et 4, nous présenterons nos résultats de la manière la plus concise possible.

0.1 Sur l'ergodicité géométrique des processus de Markov

Dans cette partie, nous étudions la convergence géométrique de quelques processus hybrides. En général, pour ce problème, on utilise des techniques de type Lyapunov [BCG08, MT93, HM11]. Ces méthodes fournissent l'existence d'une probabilité invariante et la convergence vers elle. Néanmoins les taux de convergence ne sont pas explicites ou optimaux. Dans ce travail, nous avons donc suivit une approche différente, semblable à celle de Bakry et Émery [BÉ85]. Rappelons que le critère de courbure de Bakry-Émery donne des conditions pour qu'un processus satisfasse une inégalité de Poincaré ou une inégalité logarithmique de Sobolev. Bien que leurs méthodes soient générales et très satisfaisantes pour les processus de diffusions, il est, en général, très difficile de les appliquer aux processus avec sauts. On pourra néanmoins consulter [CDPP09]. Récemment, Joulin, Ollivier et Sammer ont introduit une nouvelle notion de courbure, basée sur la distance de Wasserstein. Le chapitre 1 rappelle sa définition et ses principales propriétés. Nous avons suivit cette approche pour décrire l'ergodicité des processus qui nous ont intéressés.

0.1.1 Courbure de Wasserstein de processus stochastiquement monotone

Soit $(X_t)_{t\geq 0}$ un processus de Markov sur une partie E de $\mathbb{R}$. Si ce processus est assez régulier, c'est-à-dire qu'il vérifie la propriété de Feller, alors le générateur $\mathcal{L}$ de son semigroupe $(P_t)_{t\geq 0}$ est de

la forme suivante :

$$\mathcal{L}f(x) = \sigma(x)f''(x) + b(x)f'(x) + \int_E (f(x+y) - f(x))\eta(dy), \tag{1}$$

pour toute fonction suffisamment régulière f et tout point $x \in E$, où σ et b sont deux fonctions régulières, σ est positive et $\eta(dy)$ est une mesure qui intègre $1 \wedge y^2$. On pourra consulter [Kol11] par exemple. Nous allons supposer que ce processus est stochastiquement monotone, c'est-à-dire que pour tout $x, y \in E$ vérifiant $x \geq y$, il existe un couple $(X_t, Y_t)_{t\geq 0}$, dont chaque distribution marginale est générée par $\mathcal{L}$, vérifiant $X_0 = x, Y_0 = y$ et pour tout $t \geq 0$,

$$X_t \geq Y_t \text{ p.s.}$$

Sous cette hypothèse, nous avons établit la formule suivante :

$$\nabla P_t f(x) = \mathbb{E}\left[\nabla f(Y_t) e^{-\int_0^t V(Y_s)ds} \mid Y_0 = x\right], \tag{2}$$

pour toute fonction régulière f, $t \geq 0$ et point $x \in E$. Ici $(Y_t)_{t\geq 0}$ est un processus auxiliaire et V une fonction explicite. On en déduit immédiatement la courbure de Wasserstein de ce semigroupe. Plus précisément, rappelons que la distance de Wasserstein associée à une distance d est donnée par

$$\mathcal{W}_d(\mu_1, \mu_2) = \inf_{\nu \in \text{Marg}(\mu_1,\mu_2)} \iint_{E\times E} d(x, y)\, \nu(dx, dy),$$

pour toutes probabilités μ_1 et μ_2. Ici, $\text{Marg}(\mu_1, \mu_2)$ représente l'ensemble des lois ν, sur $E \times E$, de lois marginales μ_1 et μ_2. L'estimation précédente du gradient de notre semigroupe donne le théorème suivant :

Théorème 0.1.1 (Courbure de Wasserstein de processus stochastiquement monotone)**.** *Soit un semigroupe de Markov $(P_t)_{t\geq 0}$ généré par (1). La quantité ρ, définie par*

$$\rho = \inf_{x\in E} -g(x) - \int_E y\eta'(x, dy),$$

est la plus grande constante vérifiant

$$\forall t \geq 0,\ \mathcal{W}_d(\mu_1 P_t, \mu_2 P_t) \leq e^{-\rho t}\mathcal{W}_d(\mu_1, \mu_2),$$

pour toutes mesures de probabilité μ_1,μ_2. En particulier, il existe une unique mesure invariante π et

$$\forall t \geq 0,\ \mathcal{W}_d(\mu_1 P_t, \pi) \leq e^{-\rho t}\mathcal{W}_d(\mu_1, \pi).$$

Dans le théorème précédent, η' désigne la dérivée faible de η. La constante ρ est appelée courbure de Wasserstein. D'autres conséquences liées à la formule (2) sont développées dans le chapitre 2, tel que des bornes fines de convergence en distance de Wasserstein et en variation totale pour des cas particuliers.

0.1.2 Processus de Markov modulé (*switching Markov process*)

Considérons une dynamique différente de celle étudiée dans la sous-section précédente. Pour construire notre nouveau processus, on a besoin des ingrédients suivants :

- une chaîne de Markov à temps continu $(I_t)_{t\geq 0}$, sur un espace fini $F = \{1, ..., N\}$, possédant une mesure invariante ν ;
- une famille de N processus de Markov, sur un même espace polonais (E, d), représentés par leurs semigroupes $(P_t^{(i)})_{t\geq 0}$ et leurs générateurs $\mathcal{L}^{(i)}$, $i \in F$.

On note $\rho(i) \in \mathbb{R}$ la courbure de Wasserstein du processus généré par $\mathcal{L}^{(i)}$; c'est-à-dire que

$$\forall t \geq 0,\ \mathcal{W}_d(\mu_1 P_t^{(i)}, \mu_2 P_t^{(i)}) \leq e^{-\rho(i)t}\mathcal{W}_d(\mu_1, \mu_2),$$

pour toutes mesures de probabilité μ_1, μ_2. On considère le processus X qui se déplace de la manière suivante : tant que I ne saute pas, X se comporte comme un processus de Markov généré par $\mathcal{L}^{(I)}$, lorsque I saute, X ne saute pas mais suit cependant une nouvelle dynamique. Le processus X n'est pas un processus de Markov au contraire du couple (X, I). Ce type de processus peut avoir des comportements assez surprenants comme le montre l'exemple développé dans [BLMZ12a]. Dans cet article, le processus X tend vers l'infini, alors que chaque dynamique sous-jacente le ramène rapidement vers 0. En collaboration avec Martin Hairer, nous avons montré le résultat suivant

Théorème 0.1.2 (Convergence exponentielle des processus modulés)**.** *Si*

$$\sum_{i\in F} \rho(i)\nu(i) > 0,$$

alors le processus $(X_t, I_t)_{t\geq 0}$ admet une unique mesure invariante et converge exponentiellement vite vers elle en une distance de Wasserstein modifiée.

Ce résultat est optimal comme on le verra dans le chapitre 3. Ce résultat se démontre de la même manière que [BLMZ12c, Theorem 1.10]. Néanmoins, nous étudions aussi le cas où les sauts de I dépendent de la position de X. Dans ce cas, le processus discret I n'est plus une chaîne de Markov et l'existence d'un équilibre n'est plus garantie. L'énoncé et la démonstration de ce nouveau résultat sont plus complexes. La démonstration dépend d'une généralisation des méthodes à la Meyn-Tweedie développée par Hairer, Mattingly, et Scheutzow [HMS11].

0.2 Comportement asymptotique de populations structurées

Dans la seconde partie de ce travail, nous nous intéressons à un modèle de type branchement. Nous étudions, de façon probabiliste, l'évolution temporelle d'un trait, dans une population de particules, et faisons un lien avec la théorie des équations aux dérivées partielles. Plus précisément, on considère le modèle qui suit.

- On commence avec un individu qui possède un certain trait X_0 appartenant à une partie E de $\mathbb{R}^d$, $d \in \mathbb{N}^*$.
- Durant la vie de cet individu, son trait $X = (X_t)_{t\geq 0}$ évolue de manière aléatoire selon une dynamique markovienne induite par le générateur :

$$\forall x \in E,\ Gf = b \cdot \nabla f + \sigma \Delta f,$$

  où b est une fonction régulière de $\mathbb{R}^d$ dans $\mathbb{R}^d$ et σ une constante positive. Notons que l'on pourrait considérer des dynamiques plus complexes avec les même techniques. Cependant, comme ce générateur capturait les exemples qui nous intéressaient, nous nous sommes restreints à cette dynamique par souci de clarté.
- Cet individu meurt à taux r ; c'est-à-dire à un temps τ vérifiant

$$\forall t \geq 0,\ \mathbb{P}\left(\tau > t \mid \mathcal{F}_t\right) = \exp\left(-\int_0^t r(X_s)\right),$$

  où $\mathcal{F}_t = \sigma\{X_s,\ \mid s \leq t\}$. Au temps τ, l'individu meurt et donne naissance à un nombre aléatoire d'enfants. Les traits des enfants peuvent être différents de celui de leur parent. Les nouveaux traits sont donnés de façon aléatoire suivant un noyau markovien et le trait du parent au moment de sa mort.
- Les traits des nouveaux individus évoluent ensuite, comme celui du premier individu, indépendamment les uns des autres.

Lorsque r est constant, le processus $(N_t)_{t\geq 0}$, représentant le nombre d'individus, ne dépend pas des traits. C'est un processus de branchement de Galton-Watson à temps continu [AN04, Har02]. Lorsque r n'est pas constant, on ne peut pas étudier le nombre d'individus sans étudier les traits. Une première approche pour comprendre ce modèle est de passer à la limite macroscopique. C'est-à-dire de faire tendre le nombre initial d'individus vers l'infini et de renormaliser convenablement, en temps et en espace, la mesure empirique pour faire apparaitre un nouveau processus. Plus précisément, soit $\mathbf{Z}_t$ la mesure empirique au temps $t \geq 0$; c'est-à-dire que

$$\mathbf{Z}_t = \sum_{u\in V_t} \delta_{X_t^u},$$

où V_t est l'ensemble aléatoire des individus en vie au temps t et X_t^u le trait de l'individu u. Soit $(\mathbf{Z}_0^{(n)})_{n\geq 0}$ une suite de mesures qui représentent des positions initiales possibles. Soit $(\mathbf{Z}^{(n)})_{n\geq 1}$ la suite de processus, à valeurs mesures, distribués comme $\mathbf{Z}$ qui commencent en $(\mathbf{Z}_0^{(n)})_{n\geq 0}$. On considère le processus renormalisé $\mathbf{X}^{(n)} = \frac{1}{n}\mathbf{Z}^{(n)}$. Soit $T > 0$, on a le théorème suivant

Théorème 0.2.1 (Approximation macroscopique de la distribution empirique)**.** *Si les hypothèses suivantes sont satisfaites :*

- *l'espace E est compact,*
- *le nombre de nouveaux enfants est majoré par un nombre déterministe,*
- *la suite de conditions initiales $\mathbf{X}_0^{(n)}$ converge en distribution vers une mesure $\mathbf{X}_0$, et*

$$\sup_{n\geq 0} \mathbb{E}\left[\mathbf{X}_0^{(n)}(E)\right] < +\infty,$$

alors $\mathbf{X}^{(n)}$ converge dans l'espace des processus càd-làg sur $[0,T]$, à valeurs mesures, muni de la distance de Skohorod vers un processus $\mathbf{X}$ commençant en $\mathbf{X}_0$.

Le processus $\mathbf{X}$ dans le théorème précédent vérifie une équation aux dérivées partielles décrite dans le chapitre 5. En particulier, si la condition initiale limite $\mathbf{X}_0$ est déterministe alors les trajectoire $t \mapsto \mathbf{X}_t$ le sont aussi. Ce résultat n'est pas surprenant. En effet, la propriété de branchement entraîne que les mesures $\mathbf{X}^{(n)}$ sont de la forme

$$\mathbf{X}_t^{(n)} \overset{d}{=} \frac{1}{n} \sum_{k=1}^{\mathbf{Z}_0^{(n)}(E)} \mu_{t,k,n},$$

où les mesures $(\mu_{t,k,n})_k$ sont des mesures aléatoires indépendantes deux à deux. La loi des grands nombres classique s'applique donc ponctuellement. L'hypothèse de compacité peut être omise dans

certain cas. Décrivons un exemple d'E.D.P. qui peut apparaître dans cette limite :

$$\forall t \geq 0, \forall x \geq 0,\ \partial_t n(t,x) + \partial_x n(t,x) = 4n(t,2x) - n(t,x),$$

avec $\mathbf{X}(dx) = n(t,x)dx$. Récemment, ce type d'équation a beaucoup été étudié d'un point de vue déterministe [BCnG12, DG09, LP09, Per07, PZ07]. Leur comportement en temps long dépend de l'étude spectrale de certains opérateurs. Nous avons donc suivit une approche similaire, avec un point de vue probabiliste, pour comprendre le comportement en temps long de notre système de particules.

0.2.1 Comportement en temps long de la mesure empirique

Sous la condition d'existence d'éléments propres (λ, V) d'un certain opérateur, nous avons montré la formule *many-to-one* suivante :

$$\forall t \geq 0,\ \frac{1}{\mathbb{E}\left[\sum_{u\in V_t} V(X_t^u)\right]} \mathbb{E}\left[\sum_{u\in V_t} f(X_t^u)V(X_t^u)\right] = \mathbb{E}\left[f(Y_t)\right], \tag{3}$$

pour toute fonction positive f, où V_t et X_t^u ont été définis précédemment. Dans la formule précédente $(Y_t)_{t\geq 0}$ est un processus dont le générateur s'exprime explicitement à l'aide de tous les paramètres précédents. Divers commentaires sur cette formule sont donnés dans le chapitre introductif 4. Si le processus Y est ergodique, avec une mesure invariante π alors on a

$$\lim_{t\to+\infty} \frac{1}{\mathbb{E}\left[\sum_{u\in V_t} V(X_t^u)\right]} \mathbb{E}\left[\sum_{u\in V_t} f(X_t^u)V(X_t^u)\right] = \int f d\pi,$$

pour toute fonction continue et bornée f. Nous avons démontré que la mesure empirique convergeait en distribution, en probabilité. En particulier lorsque r est constant, les fonctions constantes sont des vecteurs propres et notre théorème de convergence est donné par le théorème suivant :

Théorème 0.2.2 (Loi des grands nombres)**.** *Si le processus auxiliaire $(Y_t)_{t\geq 0}$ est ergodique, avec π comme mesure invariante, alors, pour toute fonction continue et bornée f sur E, on a*

$$\lim_{t\to+\infty} \frac{\mathbf{1}_{\{N_t>0\}}}{N_t} \sum_{u\in V_t} f(X_t^u) = \mathbf{1}_{\{N_t>0, \forall t>0\}} \int_E f d\pi \text{ en probabilité.}$$

Lorsque r est constant, ce résultat avait déjà été démontré dans [BDMT11]. Nous généralisons ce théorème lorsque r n'est pas constant dans le chapitre 5.

0.2.2 Un modèle plus simple

Considérons l'exemple simple suivant :

– le processus se déplace suivant la diffusion suivante entre les divisions

$$dX_t = (\mu + \sigma^2)X_t dt + \sqrt{2\sigma} X_t dB_t,$$

– lorsqu'un individu meurt, il donne naissance à deux enfants, dont les traits sont donnés par la moitié de celui de cet individu ;

– le taux de division r est constant.

Ce modèle est très simple et est caractérisé par les paramètres $\mu \in \mathbb{R}$, $\sigma \geq 0$ et $r > 0$. Il a l'avantage de se comparer très facilement aux modèles de marche aléatoire branchante [Shi11, Zei12] et leurs équivalents continus [ABBS11, Ber06]. Ce cas particulier nous a intéressés de par sa ressemblance avec l'exemple principal du chapitre 5 et le modèle de Bansaye-Tran [BT11]. Le théorème 0.2.2 ne s'applique pas ici mais on peut montrer qu'il n'y a que deux possibilités : les traits explosent ou s'éteignent. Cela est établit avec le théorème 6.1.1. Notre principale motivation était d'étudier le comportement du plus gros trait $\overline{X}$, définit par

$$\forall t \geq 0, \overline{X}_t = \max_{u \in V_t} X_t^u.$$

Nous avons montré le théorème suivant :

Théorème 0.2.3 (Estimée asymptotique de la particule extrémale)**.** *Quelque soit la position initiale, on a*

$$\lim_{t \to +\infty} \frac{1}{t} \ln \overline{X}_t = \inf_{\alpha \geq 0} \mu + \sigma^2 \alpha + r \frac{2e^{-\alpha \ln(2)} - 1}{\alpha}.$$

La démonstration découle de la formule 3, appliquée avec différents vecteurs propres.

0.3 Organisation de l'ouvrage

La première partie décrit les résultats que nous avons obtenus sur l'ergodicité des processus de Markov. Dans cette partie, l'espace d'état des processus de Markov, qui nous intéressent, est essentiellement fini-dimensionnel. Cette partie comprend les chapitres 1, 2 et 3. Dans le chapitre 1, nous introduisons la courbure de Wasserstein et quelques-unes de ces propriétés. En particulier, nous mettons en évidence le lien entre cette notion, la courbure de Ricci et la courbure de Bakry-Émery. Le chapitre qui suit est basé sur l'article [Clo12]. Dans celui-ci, nous explicitons la courbure des processus de Markov stochastiquement monotones, ainsi que diverses conséquences basées sur l'estimation

du gradient du semigroupe. Dans le chapitre 3, nous donnons quelques critères de convergence pour des processus de Markov modulés. Ce chapitre est basé sur l'article [CH], en collaboration avec Martin Hairer.

La deuxième partie traite du comportement en temps long d'une population structurée. Plus précisément, nous étudions un système de particules avec une interaction de type branchement ; c'est-à-dire que l'interaction n'intervient qu'à la naissance des particules. Cette partie est aussi composée de trois chapitres : les chapitres 4, 5 et 6. Le premier chapitre de cette partie rappelle rapidement ce qu'est une $h-$transformée et met en évidence son utilisation dans les divers chapitres. Le chapitre 5 est constitué de l'article [Clo11]. Nous y établissons des théorèmes limites, en grande population et en temps long, pour la mesure empirique qui décrit le système de particules. Finalement, le dernier chapitre est consacré à l'étude de la plus grande particule dans le cadre introduit dans la section 0.2.2.

Première partie

Sur l’ergodicité géométrique des processus de Markov

Chapitre 1

Sur la géométrie des processus de Markov

1.1 Introduction

Ce chapitre a pour but d'introduire brièvement la courbure de Wasserstein des semigroupes de Markov. Cette notion, commune aux deux prochains chapitres, a été introduite par Joulin, en temps continu, et par Ollivier, en temps discret. Notons que Sammer l'a aussi introduite dans sa thèse. Cette courbure apparaît comme le paramètre optimal de contraction d'un semigroupe de Markov, par rapport à la distance de Wasserstein. D'une certaine manière, elle généralise les courbures de Bakry-Émery et de Ricci pour des processus non-diffusifs et les espaces discrets. De plus, elle est naturellement reliée à la condition d'unicité de Dobrushin, à l'exposant de Chen ou au trou spectral de Wasserstein. Par la suite, nous commençons par rappeler le critère de Bakry-Émery et son lien avec la géométrie. Nous rappelons ensuite les principales propriétés de la distance de Wasserstein. Après ces rappels, nous introduisons la distance de Wasserstein et donnons quelques propriétés, incluant quelques inégalités de concentration. Nous donnons ensuite les principaux exemples. Nous évoquons quelques résultats du chapitre 2. Comme application, dans la dernière section, nous décrierons rapidement les théorèmes du chapitre 3 sur la convergence des processus de Markov modulé.

1.2 Sur la géométrie de Bakry-Émery

1.2.1 Courbure de Bakry-Émery d'un processus de Markov

Soit E un espace polonais et $(X_t)_{t\geq 0}$ un processus de Markov sur E. Pour tout $t \geq 0$, on considère l'opérateur P_t qui à toute fonction $f : E \to \mathbb{R}$ mesurable et bornée associe la fonction mesurable et bornée définie, pour tout $x \in E$, par

$$P_t f(x) = \mathbb{E}\left[f(X_t) \mid X_0 = x\right] = \int_E f(y) P_t(x, dy).$$

L'opérateur $P_0 = \mathrm{Id}$ est la fonction identité et la propriété de Markov entraîne que $(P_t)_{t\geq 0}$ est un semi-groupe :

$$\forall s, t \geq 0,\ P_{t+s} = P_{s+t} = P_t \circ P_s = P_s \circ P_t.$$

Si X_0 a pour loi de départ μ alors, pour tout $t \geq 0$, la loi de X_t est décrite par

$$\mathbb{E}\left[f(X_t)\right] = \mu P_t f = \int_E \int_E f(y) P_t(x, dy) \mu(dy)$$

pour toute fonction continue et bornée f. En particulier, une mesure de probabilité π est dite invariante, ou stationnaire, si

$$\forall t \geq 0,\ \pi P_t = P_t.$$

Si X_t converge en loi, lorsque t tend vers l'infini, vers une distribution π alors π est nécessairement invariante. Cette mesure est naturellement connectée au générateur. On dit qu'une fonction mesurable et bornée $f : E \to \mathbb{R}$ appartient au domaine du générateur $\mathcal{L}$ du processus lorsque la limite suivante existe pour tout $x \in E$:

$$\mathcal{L} f(x) = \lim_{t \to 0+} \frac{P_t f(x) - f(x)}{t} = \partial_t P_t f\,_{|t=0}(x).$$

L'opérateur $\mathcal{L}$ ainsi défini est appelé générateur infinitésimal du processus $(X_t)_{t\geq 0}$ ou du semi-groupe $(P_t)_{t\geq 0}$. La propriété de semi-groupe donne les équations de Chapman-Kolmogorov forward et backward :

$$\forall t \geq 0,\ \partial_t P_t(f) = P_t(\mathcal{L} f) = \mathcal{L}(P_t f).$$

Comme première application, on voit qu'une mesure π est invariante si

$$\pi\left(\mathcal{L} f\right) = \int_E \mathcal{L} f(x) \pi(dx) = 0,$$

pour toute fonction f appartenant au domaine de $\mathcal{L}$. Tout se passe comme si « $P_t = e^{t\mathcal{L}}$ ». En effet, la précédente expression peut être interprétée en terme d'éléments propres : la mesure π est un vecteur propre de $\mathcal{L}^*$ associé à la valeur propre 0. Notons que la fonction constante égale à 1 est un vecteur propre de $\mathcal{L}$ pour la même valeur propre. Comme pour les matrices, si ce vecteur propre est unique alors le semi-groupe converge vers elle, à une vitesse décrite par la seconde valeur propre. Plus précisément, on dit qu'une mesure π est réversible, ou symétrique, si

$$\int_E f(x) P_t g(x) \pi(dx) = \int_E g(x) P_t f(x) \pi(dx),$$

ou de manière équivalente

$$\int_E f(x) \mathcal{L} g(x) \pi(dx) = \int_E g(x) \mathcal{L} f(x) \pi(dx),$$

pour toutes fonctions f, g suffisamment régulières. C'est-à-dire que $\mathcal{L}$ et P_t sont auto-adjoints (ou symétriques) dans $L^2(E, \pi)$. On dit qu'un semigroupe symétrique, par rapport à sa mesure invariante, satisfait une inégalité de trou spectral, si le spectre de $-\mathcal{L}$ est inclus dans $\{0\} \cup [\lambda_0, \infty)$, pour un certain $\lambda_0 > 0$. Via un argument de théorie spectrale, on montre facilement que dans ce cas, on a

$$\forall t \geq 0, \ \mathrm{Var}_\pi(P_t f) \leq e^{-\lambda_0 t} \mathrm{Var}_\pi(f), \tag{1.1}$$

pour toute fonction f suffisamment régulière. Ici, on a utilisé la notation

$$\mathrm{Var}_\pi(f) = \left\| f - \int_E f d\pi \right\|_{L^2(E,\pi)} = \int_E \left(f(x) - \int_E f(y) \pi(dy) \right)^2 \pi(dx).$$

Malheureusement, la plupart des processus que l'on étudie, dans cette thèse, ne sont pas symétriques. Dans le cas de processus non-réversibles, on peut néanmoins démontrer une inégalité de Poincaré :

Lemme 1.2.1 (Inégalité de Poincaré)**.** *Supposons que le domaine de $\mathcal{L}$ est dense dans $L^2(E, \pi)$, où π est la mesure invariante de $(P_t)_{t \geq 0}$. L'inégalité du trou spectral* (1.1) *est équivalente à l'inégalité de Poincaré : pour toute fonction f dans $L^2(E, \pi)$, on a*

$$\mathrm{Var}_\pi(f) \leq -\frac{1}{\lambda_0} \int_E f(x) \mathcal{L} f(x) \pi(dx). \tag{1.2}$$

Ce lemme donne l'équivalence entre une information locale, c'est-à-dire infinitésimale, et une information globale, c'est-à-dire de convergence lorsque t tend vers l'infini. L'inégalité de Poincaré est un cas particulier de diverses inégalités fonctionnelles liées à la convexité [ABC+00, Cha04]. Notons que l'inégalité (1.1) est toujours satisfaite avec $\lambda_0 = 0$. En effet, l'inégalité de Jensen assure

que l'application $t \to \mathrm{Var}_\pi(P_t f)$ est décroissante. La quantité de droite dans l'inégalité de Poincaré fait intervenir l'énergie du processus ; c'est-à-dire la quantité

$$\mathcal{E}(f) = -\int_E f(x)\mathcal{L}f(x)\pi(dx) = \int_E \Gamma(f,f)(x)\pi(dx),$$

où l'opérateur Γ, nommé *carré du champ*, est défini par

$$\Gamma(f,g) = \frac{1}{2}\left[\mathcal{L}(fg) - g\mathcal{L}f - f\mathcal{L}g\right].$$

On peut aussi définir une famille d'opérateur $(\Gamma_n)_{n\geq 0}$ de la manière suivante :

$$\Gamma_{n+1}(f,g) = \frac{1}{2}\left[\mathcal{L}\Gamma_n(f,g) - \Gamma_n(\mathcal{L}f,g) - \Gamma_n(f,\mathcal{L}g)\right],$$

avec $\Gamma_0(f,g) = f \times g$. En particulier $\Gamma_1 = \Gamma$. On note aussi $\Gamma_n f = \Gamma_n(f,f)$. Ces divers opérateurs interviennent dans diverses inégalités fonctionnelles. Citons par exemple le critère de Bakry-Émery [ABC+00, Proposition 5.4.1] qui donne une information sur la stabilité du processus :

Théorème 1.2.2 (Critère de courbure de Bakry-Émery)**.** *Soit $\rho \in \mathbb{R}$. Sous les hypothèses du lemme précédent, les assertions suivantes sont équivalentes :*

- *pour toute fonction f dans le domaine de $\mathcal{L}$, on a*

$$\Gamma_2 f \geq \rho\Gamma f, \tag{1.3}$$

- *Pour toute fonction f dans le domaine de $\mathcal{L}$ et $t \geq 0$, on a*

$$\Gamma P_t f \leq e^{-2\rho t} P_t \Gamma f, \tag{1.4}$$

- *L'inégalité de Poincaré locale est satisfaite : pour toute fonction f dans le domaine de $\mathcal{L}$ et $t \geq 0$, on a*

$$P_t(f^2) - (P_t f)^2 \leq \frac{1-e^{-2\rho t}}{\rho} P_t \Gamma f. \tag{1.5}$$

En particulier, si $\rho > 0$ alors l'inégalité du trou spectral (1.1) est satisfaite avec $\lambda_0 = \rho$.

Le paramètre optimal ρ (i.e. le plus grand), dans le lemme précédent, est appelé courbure de Bakry-Émery car il est lié à la géométrie de l'espace comme on peut le voir avec les processus de diffusion. Notons que lorsque la mesure invariante π est réversible alors il existe aussi le « critère Γ_2 intégré » [ABC+00, Proposition 5.5.4] qui donne une condition plus faible pour que π satisfasse une inégalité du trou spectral.

1.2.2 Les processus de diffusion

Un processus de diffusion sur un espace polonais E est un processus X dont le générateur infinitésimal $\mathcal{L}$ satisfait la propriété de dérivation en chaîne suivante : pour toute famille finie $f = (f_1, \dots, f_n)$ de fonctions appartenant au domaine de $\mathcal{L}$, et toute fonction $\phi : \mathbb{R}^n \to \mathbb{R}$ régulière, on a

$$\mathcal{L}\phi(f) = \sum_{i=1}^{n} \partial_{x_i}\phi(f)\mathcal{L}f_i + \sum_{i,j} \partial_{x_i}\partial_{x_j}\phi(f)\Gamma(f_i, f_j).$$

Cette propriété est liée aux propriétés de régularité des trajectoires du processus. En effet, si $(P_t)_{t\geq 0}$ est un semigroupe de diffusion, alors les trajectoires $t \to f(X_t)$ sont continues, pour toute fonction f dans le domaine de $\mathcal{L}$. Sur $\mathbb{R}^n$, les processus de diffusions sont les solutions d'équations stochastiques dirigées par un mouvement brownien. Plus précisément,

$$\forall t \geq 0,\ dX_t = b(X_s)ds + \sigma(X_s)dB_s,$$

où $(B_t)_{t\geq 0}$ est un mouvement brownien sur $\mathbb{R}^n$, $\sigma : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ est une application mesurable tel que, pour tout $x \in \mathbb{R}^n$, $\sigma(x)$ est une matrice symétrique et positive, et $b : \mathbb{R}^n \to \mathbb{R}^n$ est une application mesurable. Le générateur est alors donné par la relation

$$\mathcal{L}f(x) = \sum_{i=1}^{n} b_i(x)\partial_{x_i}f(x) + \sum_{i,j=1}^{n} \frac{1}{2}(\sigma(x)\cdot\sigma(x)^t)_{i,j}\,\partial_{x_i}\partial_{x_j}f(x).$$

L'exemple fondamental est le processus dit de Kolmogorov-Langevin :

$$\forall t \geq 0,\ dX_t = -\nabla V(X_s)ds + \sqrt{2}dB_s, \tag{1.6}$$

où V est un potentiel régulier vérifiant $\int_{\mathbb{R}^n} e^{-V(x)}dx = 1$. Ce processus possède une unique mesure réversible et stationnaire, $\pi(dx) = e^{-V(x)}dx$, nommée mesure de Blotzmann-Gibbs. Dans ce cas,

$$\mathcal{L}f = \Delta f - \nabla V \cdot \nabla f,\ \ \Gamma(f,g) = \langle \nabla f, \nabla g\rangle,$$

et $\Gamma f = \Gamma(f,f) = \|\nabla f\|^2$, d'où la dénomination *carré du champ* (de gradient). Enfin, l'opérateur Γ_2 est donné par

$$\Gamma_2 f = \|\mathrm{Hess} f\|^2 + (\nabla f)^\top \cdot \mathrm{Hess} V \cdot \nabla f.$$

On peut donc maintenant utiliser le critère de Bakry-Émery pour avoir le résultat suivant :

Corollaire 1.2.3 (Courbure de Bakry-Émery des processus de Kolmogorov-Langevin)**.** *Si X est so-*

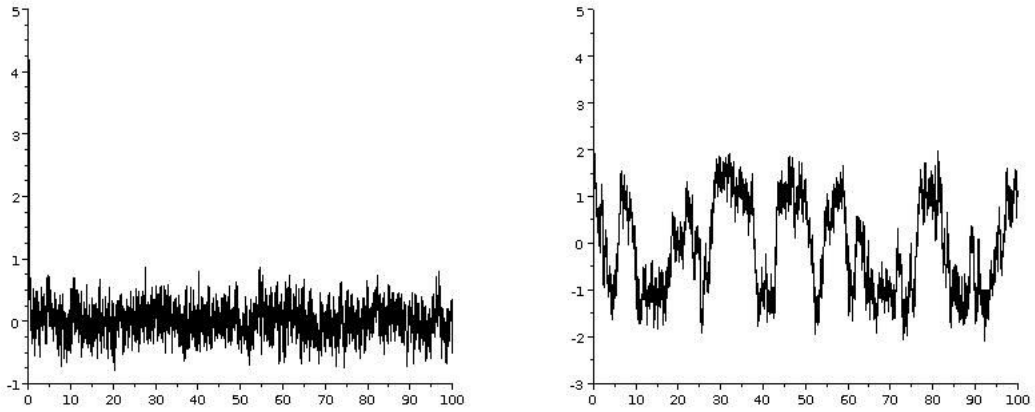

Figure 1.1 – Trajectoire de deux processus de Kolmogorov

lution de (1.6) *et* ρ *minore le spectre de la matrice hessienne de* V*, alors les assertions du Théorème 1.2.2 sont satisfaites avec* $\lambda_0 = \rho$*.*

Exemple 1.2.4 (Processus d'Ornstein-Uhlenbeck)**.** *Le processus d'Ornstein-Uhlenbeck est le processus de Kolmogorov-Langevin associé au potentiel* $V(x) = \mu\|x\|^2/2$*, pour* $\mu > 0$*. La courbure de ce processus vaut* μ *et les inégalités* (1.1)*,* (1.2)*,* (1.3) *et* (1.4) *sont des égalités. On parle parfois de courbure exacte. Le membre de gauche de la figure 1.1 représente une de ces trajectoires. On peut remarquer que le processus se stabilise très rapidement vers un équilibre concentré en zéro. Pour comparer, le membre de droite représente une diffusion de Kolmogorov associée au potentiel non-convexe* $V(x) = x^4/4 - x^2/2$*. La courbure de ce second processus est négative. Cela vient du fait que sa distribution invariante est bimodale et le processus n'atteint son équilibre qu'après avoir visité les deux « bosses »de sa probabilité stationnaire.*

Remarque 1.2.5 (Un lien avec la géométrie)**.** *Soit* M *une variété compacte de dimension* n *munie d'une métrique riemannienne* g*. Il existe un opérateur* Δ *sur* M*, dit de Laplace-Beltrami, qui généralise le laplacien sur* $\mathbb{R}^n$*. Cet opérateur génère un processus de diffusion qui admet une unique mesure invariante et réversible, qui est la mesure de volume normalisée associée à* (M, g)*. On peut montrer dans ce cas, que la courbure* ρ *de Bakry-Émery vérifie, pour tout* $k \in \mathbb{R}$*,*

$$\forall x \in M,\ \forall v \in T_xM,\ \mathbf{Ric}_x(v, v) \geq k\|v\|^2 \Rightarrow \rho \geq k.$$

Ici, Ric représente le tenseur de Ricci et T_xM *le plan tangent à* M *en* x*. Par exemple, dans le cas où* M *est une sphère de dimension* 2*,* $\rho = 1$*. Cette définition de la courbure permet donc d'unifier*

l'analyse et la géométrie dans le cadre de l'étude des processus de diffusions sur des variétés. Notons aussi que lorsque l'on considère un opérateur de la forme $\mathcal{L} = \Delta + X$, où X est un champ de vecteur, alors, comme dans le corollaire 1.2.3, la courbure de Bakry-Émery s'écrit comme la somme d'un terme lié à la courbure de Ricci et d'un terme minorant le spectre de la matrice Hessienne de X. La courbure de Bakry-Émery prend donc en compte à la fois la courbure de l'espace ambiant et la dérive du processus.

On vient de voir que la courbure de Bakry-Émery se calcule aisément dans le cadre des processus de diffusion, et que dans ce cas elle caractérise la courbure de l'espace ambiant. La situation est différente lorsque le processus admet des sauts car la règle de dérivation en chaîne ne s'applique pas et il est difficile d'expliciter l'opérateur Γ_2 , ou lorsque l'espace est discret car il n'existe pas de notion universelle équivalente à la courbure de Ricci. Ces problèmes nous poussent à étudier une autre notion de courbure.

1.3 Courbure de Wasserstein

Dans les sections précédentes, on a remarqué que les méthodes liées aux inégalités fonctionnelles ou à la théorie spectrale fonctionnaient très bien pour certains processus mais pas pour d'autres. On va donc introduire ici une autre façon d'estimer la vitesse de convergence. Précédemment, la loi au temps t était caractérisée par la fonction $x \mapsto P_t f(x)$ et la vitesse était estimée à partir d'une distance sur les fonctions (norme L^2). Maintenant, on va caractériser la loi par la mesure $\delta_x P_t$ et estimer la vitesse de convergence à l'aide d'une distance sur les mesures.

1.3.1 Distance de couplage

Nous munissons l'ensemble des mesures de probabilité sur (E, d) de la distance suivante :

Définition 1.3.1 (Distance de Wasserstein)**.** *La distance de Wasserstein est définie par*

$$\mathcal{W}_d(\mu,\nu) = \inf_{\eta \in Marg(\mu,\nu)} \int_{E\times} d(x,y)\eta(dx,dy) = \inf_{X\sim\mu, Y\sim\nu} \mathbb{E}\left[d(X,Y)\right]. \tag{1.7}$$

Où Marg est l'ensemble des lois η sur $E \times E$ tel que $\eta(\cdot, E) = \mu$ et $\eta(E, \cdot) = \nu$.

On sait par [Vil09, Theorem 1.3] que l'infimum est atteint. Cette distance est aussi connue sous les autres noms suivants : distance de transport (optimal), de couplage, de Fréchet, de Kantorovich, de Mallows (en statistique), de Monge, distance L^1 généralisée, EMD (earth mover's distance, en informatique)... Diverses propriétés de cette distance sont développées dans les livres [Che04, Rac91,

Vil09]. Retenons qu'une suite de mesures de probabilités $(\mu_n)_{n\geq 0}$ converge en Wasserstein vers μ si elle converge en loi et

$$\lim_{n\to+\infty}\int_E d(x,x_0)\mu_n(dx)=\int_E d(x,x_0)\mu(dx),$$

pour un certain $x_0\in E$, ou de manière équivalente, pour tout $x_0\in E$. On en déduit donc facilement que l'espace des mesures de probabilités, admettant un premier moment fini, muni de la distance de Wasserstein, est complet. De plus, si d est bornée sur E, la convergence en Wasserstein est équivalente à la convergence en loi.

Remarque 1.3.2 (Simulation et formule d'appariement : *matching formula*)**.** *Soit μ et $\widetilde{\mu}$ deux mesures définies par*

$$\mu=\frac{1}{n}\sum_{k=1}^{n}x_k \;\; \textit{et}\;\widetilde{\mu}=\frac{1}{n}\sum_{k=1}^{n}y_k,$$

où $x_1,\dots,x_n,y_1,\dots,y_n\in E$. On peut démontrer la formule d'appariement [Vil09, Example p.5] :

$$\mathcal{W}_d(\mu,\nu)=\inf_{\sigma\in S_n}\frac{1}{n}\sum_{k=1}^{n}d(x_k,y_{\sigma(k)}),$$

où S_n désigne le groupe des permutations d'ordre n. La démonstration est un problème de minimisation convexe. Ici, l'ensemble Marg est en bijection avec l'ensemble (convexe) des matrices bistochastiques. L'infimum est donc atteint sur l'un des points extrémaux. Or il est connu, par un théorème de Birkhoff et von Neumann que les points extrémaux correspondent aux matrices de permutations. Ceci termine la démonstration de cette formule. Comme application, on trouve un moyen simple et probabiliste de simuler la distance entre deux mesures de probabilité ν et $\widetilde{\nu}$ sur $(\mathbb{R},|\cdot|)$, possédant un premier moment. En effet, Soit $(X_i)_{i\in\mathbb{N}^*}$ et $(\widetilde{X}_i)_{i\in\mathbb{N}^*}$ deux suites de variables i.i.d. de lois ν et $\widetilde{\nu}$. La loi des grands nombres donne

$$\lim_{n\to\infty}\mathcal{W}_d\left(\frac{1}{n}\sum_{k=1}^{n}\delta_{X_k},\nu\right)=0\;\; \textit{et}\;\lim_{n\to\infty}\mathcal{W}_d\left(\frac{1}{n}\sum_{k=1}^{n}\delta_{\widetilde{X}_k},\widetilde{\nu}\right)=0.$$

En utilisant ces convergences et la formule d'appariement on a

$$\mathcal{W}_d(\nu,\widetilde{\nu})\approx\inf_{\sigma\in S_n}\frac{1}{n}\sum_{k=1}^{n}d(X_k,\widetilde{X}_{\sigma(k)}).$$

L'infimum précédent se calcule de la manière suivante : on réarrange les deux suites de façon à les

rendre croissantes i.e. $X_{(1)} \leq X_{(2)}...$ *et* $\widetilde{X}_{(1)} \leq \widetilde{X}_{(2)}...$ *Et on trouve*

$$\mathcal{W}_d(\nu, \widetilde{\nu}) \approx \frac{1}{n} \sum_{k=1}^{n} d(X_{(k)}, \widetilde{X}_{(k)}).$$

L'équation (1.7) admet la formulation équivalente suivante :

Théorème 1.3.3 (Dualité de Kantorovich-Rubinstein)**.** *Pour toutes mesures* μ *et* ν*, on a*

$$\mathcal{W}_d(\mu, \nu) = \sup_{\|f\|_{Lip} \leq 1} \int f d\mu - \int f d\nu,$$

où

$$\|f\|_{Lip} = \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x, y)}. \tag{1.8}$$

En particulier, si $E = \mathbb{R}$ et $d(x, y) = |x - y|$ alors

$$\mathcal{W}_d(\mu, \nu) = \|F_\mu - F_\nu\|_{L^1} = \|F_\mu^{-1} - F_\nu^{-1}\|_{L^1},$$

où F_μ, F_ν désignent les fonctions de répartition de μ et ν et F_μ^{-1}, F_ν^{-1} leurs inverses généralisés. On notera $\mathcal{W}$ plutôt que $\mathcal{W}_d$ pour ce choix de distance. Un autre cas intéressant est lorsque $d(x, y) = \mathbf{1}_{x \neq y}$. Dans ce cas, $\mathcal{W}_d = d_{\text{VT}}$ est la distance en variation totale. Si μ et ν ont des densités f_μ, f_ν par rapport à la même mesure η, on a alors

$$d_{\text{VT}}(\mu, \nu) = \|f_\mu - f_\nu\|_{L^1(\eta)}.$$

Généralement, si d est bornée alors la convergence en d_{VT} implique la convergence en $\mathcal{W}$. Si

$$\inf_{x \neq y} d(x, y) > 0,$$

alors la convergence en $\mathcal{W}$ implique la convergence en d_{VT}. Notons aussi que sur $\mathbb{R}$, lorsque

$$\forall x, y \in \mathbb{R},\ d(x, y) = \min\{|x - y|, 1\},$$

alors $\mathcal{W}_d$ est nommée la distance de Fortet-Mourier et engendre la topologie de la convergence en loi, tout comme la distance de Lévy-Prokhorov. Finissons cette section en évoquant la distance de Wasserstein d'ordre p. De même que (1.7), on peut de définir une nouvelle distance en minimisant la norme L^p pour $p \geq 1$:

$$\mathcal{W}_d^{(p)}(\mu, \nu) = \left(\inf_{\eta \in \text{Marg}(\mu, \nu)} \int_{E \times E} d(x, y)^p \eta(dx, dy) \right)^{1/p}.$$

Dans ce cas, une version du théorème de Kantorovich-Rubinstein donne que

$$\mathcal{W}_d^{(p)}(\mu,\nu)^p = \sup_{f,g} \int f d\mu - \int g d\nu,$$

où le supremum est pris sur les fonctions f et g vérifiant

$$f(x) - g(y) \leq d(x,y)^p.$$

Cependant il n'existe pas d'analogue direct connu du Théorème 1.3.3. On pourra cependant consulter [Rio98] pour une comparaison entre $\mathcal{W}_d^{(p)}$ et les distances de Zolotarev.

1.3.2 Définition et propriétés de la courbure de Wasserstein

On a vu dans la section précédente que les méthodes de type Bakry-Émery fonctionnaient bien avec les diffusions mais mal pour des processus avec sauts. Récemment, Aldéric Joulin, Yann Ollivier et Marcus Sammer ont introduit, indépendamment, une nouvelle notion de courbure. Celle-ci est basée sur la notion de couplage et la distance de Wasserstein.

Définition 1.3.4 (Courbure de Wasserstein)**.** *La courbure de Wasserstein, d'un semigroupe $(P_t)_{t\geq 0}$ est la plus grande constante $\rho \in \mathbb{R}$ tel que, pour tout $x, y \in E$ et $t \geq 0$, on ait*

$$\mathcal{W}_d\left(\delta_x P_t, \delta_y P_t\right) \leq e^{-\rho t} \mathcal{W}_d\left(\delta_x, \delta_y\right).$$

Cette définition dépend de la distance choisie. On peut montrer que pour un mouvement brownien se déplaçant sur une variété riemannienne, les courbures de Wasserstein et de Bakry-Émery coïncident, voir [vRS05]. Pour les espaces discrets, il existe d'autres notions de courbure basées sur le déplacement de l'entropie [EM12, OV10, GRST12]. L'inégalité précédente avait déjà été introduite par Dobrushin en 1970. Demander que ρ soit strictement positive est connu comme la *Dobrushin uniqueness condition* en mécanique statistique. Sous cette condition on a l'existence d'une mesure invariante :

Théorème 1.3.5 (Existence et unicité d'une mesure invariante)**.** *Si la courbure de Wasserstein ρ d'un semigroupe $(P_t)_{t\geq 0}$ est strictement positive alors il existe une unique mesure invariante π. De plus, pour tout $x \in E$ et $t \geq 0$, on a*

$$\mathcal{W}_d\left(\delta_x P_t, \pi\right) \leq e^{-\rho t} \mathcal{W}_d\left(\delta_x, \pi\right).$$

On peut trouver une démonstration de ce résultat dans [Che04]. Nous en donnons une différente dans l'annexe du chapitre 2. Si la mesure π est réversible alors la courbure de Wasserstein minore le trou spectral. Voir le théorème 2.4.13 pour plus de détails. D'une manière générale, une mesure de probabilité, satisfaisant une inégalité de Poincaré, satisfait aussi une inégalité de concentration exponentielle. Divers inégalités de concentration similaires ont été développées par différents auteurs lorsque la courbure est positive. Évoquons seulement le critère suivant, qui généralise [Jou07, Theorem 3.1] et [Jou09, Theorem 2.6]. Rappelons que la norme Lipschtizienne d'une fonction f est donnée en (1.8). Ajoutons aussi la notation $\mathbb{P}_x$ pour $\mathbb{P}(\cdot \mid X_0 = x)$ et $\mathbb{E}_x$ pour $\mathbb{E}[\cdot \mid X_0 = x]$. On a

Théorème 1.3.6 (Concentration de type Poisson pour la mesure invariante)**.** *Si toutes les hypothèses suivantes sont satisfaites :*

1. *la courbure de Wasserstein ρ est strictement positive,*
2. *il existe $A > 0$, tel que $\sup_{t\geq 0} d(X_{t-}, X_t) \leq A$ presque sûrement,*
3. *il existe $B > 0$, tel que pour toute fonction f, on a $\Gamma(f, f) \leq B\|f\|^2_{Lip}$,*

alors, pour tous $y > 0$, $x \in E$ et toute fonction f tel que $\|f\|_{Lip} = 1$, on a

$$\forall y > 0,\ \mathbb{P}_x\left(f(X_t) - \mathbb{E}_x[f(X_t)] \geq y\right) \leq e^{-g(y,t)},$$

où

$$g(y,t) = \frac{y}{2A}\log\left(1 + \frac{2A\rho y}{(1-e^{-2\rho t})B}\right),$$

et

$$\mathbb{P}_x\left(\frac{1}{t}\int_0^t f(X_s)ds - \int f d\pi \geq y + \frac{(1-e^{-\rho t})}{\rho t}\int d(x,y)\pi(dy)\right)$$
$$\leq \exp\left(-\frac{Bt}{A^2}\, h\left(\frac{A\rho y}{B(1-e^{-\rho t})}\right)\right),$$

où

$$h(y) = (1+y)\log(1+y) - y.$$

La fonction $y \mapsto h(y)$ se comporte comme y^2 lorsque y est proche de 0 et comme $y\log(y)$ lorsque y tend vers l'infini.

Idée de la démonstration. La preuve est basée sur l'étude de la martingale $(Z_s)_{0\leq s\leq t}$, donnée par $Z_s = P_{t-s}f(X_s) - P_t f(X_0)$, pour $s \in [0,t]$. Par définition de A et B, on a

$$\sup_{s\in[0,t]} |Z_s - Z_{s-}| \leq A\|f\|_{\text{Lip}},$$

et

$$\begin{aligned}\langle Z, Z\rangle_s &= \int_0^s \Gamma(P_{t-s}f, P_{t-s}f)(X_s)ds \\ &\leq \frac{B\|f\|_{\text{Lip}}^2}{4\rho},\end{aligned}$$

où

$$\|f\|_{\text{Lip}} = \sup_{x\neq y} \frac{|f(x)-f(y)|}{d(x,y)}.$$

Puis, comme dans [Jou07, Jou09], on construit une supermartingale à l'aide du [Kal02, Lemma 23.19]. Le reste de la démonstration suit point par point celle des théorèmes cités. □

Dans [HSV11], le trou spectral de Wasserstein d'un semigroupe de Markov est défini comme la plus grande constante $\lambda > 0$ tel qu'il existe $C > 0$ tel que, pour tous $x, y \in E$ et $t \geq 0$, on ait

$$\mathcal{W}_d(\delta_x P_t, \delta_y P_t) \leq Ce^{-\lambda t} d(x,y).$$

Si un semigroupe vérifie les hypothèses du théorème précédent avec un trou spectral de Wasserstein strictement positive, à la place d'une courbure de Wasserstein strictement positive alors la première inégalité de concentration reste valide avec d'autres constantes, mais pas la seconde.

1.3.3 Quelques exemples de courbure

Ici, nous donnons quelques exemples de courbure de Wasserstein.

1.3.3.1 Chaînes de Markov à temps continu

Supposons que E est fini et, sans perte de généralités, que $E = \{1, ..., N\}$, pour un certain $N \in \mathbb{N}^*$. L'opérateur $\mathcal{L}$ opère alors sur toutes les fonctions numériques f, définies sur E, à travers la relation

$$\forall i \in E,\ \mathcal{L}f(i) = \sum_{j=1}^{N} L_{i,j} f(j),$$

où $(L_{i,j})_{i,j\in E^2}$ est une matrice, dont les éléments non-diagonaux sont positifs, et tel que la somme des éléments d'une même colonne est nulle. L'opérateur P_t se représente de même par une matrice et la relation $P_t = e^{t\mathcal{L}}$ est l'exponentielle usuelle d'une matrice. Le *carré du champ* s'écrit alors,

$$\Gamma(f,g)(i) = \frac{1}{2}\sum_{j=1}^{N} L_{i,j}(f(j)-f(i))(g(j)-g(i)).$$

Dans ce cas le théorème de Perron-Frobenius nous donne, sous la condition d'irréductibilité, et donc de récurrence positive, l'existence d'un trou spectral. La valeur propre peut être calculée directement sur certains exemples ou numériquement si N est assez petit. Cependant, il peut être difficile de calculer explicitement cette valeur propre ou d'obtenir une borne satisfaisante. De plus, il est difficile de calculer explicitement la courbure de Bakry-Émery. Utilisant des méthodes de couplage, on montre le lemme suivant :

Lemme 1.3.7 (Convergence pour des chaînes discrètes sur un espace fini)**.** *Supposons que $(P_t)_{t\geq 0}$ est le semigroupe d'une chaîne de Markov irréductible sur un espace fini E. Il existe $\kappa > 0$ tel que pour toutes lois μ et ν, on ait*

$$\forall t \geq 0,\ d_{VT}(\mu P_t, \nu P_t) \leq Ce^{-\kappa t} d_{VT}(\mu, \nu).$$

De plus, si $L_{i,j} > 0$ pour tout $i, j \in E$ alors on a $C = 1$.

En particulier, lorsque $C = 1$, la courbure de Wasserstein, associée à la distance triviale $d : (x, y) \mapsto \mathbf{1}_{x\neq y}$, est strictement positive. La preuve est basée sur le couplage de Doeblin, introduit par lui-même en 1938, dans un article intitulé *Exposé de la théorie des chaînes simple constantes de Markov à un nombre fini d'états*.

Démonstration du Lemme 1.3.7 . On considère deux processus indépendants X et $\widetilde{X}$, générés par $\mathcal{L}$, jusqu'au temps aléatoire T où ils coalescent, puis on les considère égaux. La propriété de Markov donne que chaque coordonnée suit la bonne dynamique et la propriété de récurrence du couple $(X, \widetilde{X})$ donne que le temps de coalescence vérifie

$$\mathbb{P}(T > t) \leq Ce^{-\kappa t},$$

pour un certain $\kappa > 0$. □

Lorsque l'espace d'état est dénombrable mais pas fini, la situation est différente. On pourra notamment lire [CJ10] qui décrit la courbure de Wasserstein, associée à diverses distances, des processus de naissance et mort. Leurs démonstrations sont basées sur une relation de commutation qui a inspiré les résultat du chapitre 2.

1.3.3.2 Processus de diffusion de Kolmogorov-Langevin

Soit $(P_t)_{t\geq 0}$ le semigroupe d'un processus de diffusion de Kolmogorov-Langevin $(X_t)_{t\geq 0}$, solution de l'E.D.S. suivante

$$dX_t = \sqrt{2} dB_t - \nabla V(X_t) dt.$$

Rappelons que $\Gamma f = |\nabla f|^2$. Soit κ un nombre réel et d la distance usuelle. Les assertions suivantes sont équivalentes :

1. pour toute fonction régulière f, on a $\Gamma_2 f \geq \kappa \Gamma f$,
2. pour toute fonction régulière f et $t \geq 0$, on a $\Gamma P_t f \leq e^{-2\kappa t} P_t \Gamma f$,
3. pour toute fonction régulière f et $t \geq 0$, on a $\sqrt{\Gamma P_t f} \leq e^{-\kappa t} P_t \sqrt{\Gamma f}$,
4. le semigroupe $(P_t)_{t\geq 0}$ satisfait l'inégalité de Poincaré locale (1.5) avec $\lambda_0 = \kappa$,
5. la constante κ minore le spectre de la hessienne de V,
6. pour tout $x, y \in E$ et $t \geq 0$, on a $\mathcal{W}_d(\delta_x P_t, \delta_y P_t) \leq e^{-\kappa t} d(x, y)$.

L'équivalence $1 \Leftrightarrow 2 \Leftrightarrow 4$ vient du théorème 1.2.2 de Bakry-Émery. Via les expressions de Γ et Γ_2, on voit facilement que $1 \Leftrightarrow 5$. L'inégalité de Jensen donne $3 \Rightarrow 2$. L'inégalité des accroissement fini donne $3 \Leftrightarrow 6$. Remarquons aussi que l'on peut déduire $5 \Rightarrow 6$ en couplant deux processus avec le même mouvement brownien.

Ces équivalences donnent en particulier que la courbure de Wasserstein et de Bakry-Émery coïncident pour les processus de diffusion de Kolmogorov-Langevin. C'est l'un des points qui a motivé cette définition. Finissons cet exemple en évoquant un résultat récent du à Andreas Eberle [Ebe11] : s'il existe une constante strictement positive qui minore le spectre de la Hessienne de V en dehors d'un ensemble compact, alors il existe une distance d_f tel que la courbure de Wasserstein, associée à d_f, est strictement positive. La preuve est basée sur un couplage différent que celui décrit pour $5 \Rightarrow 7$, nommé couplage par réflexion.

Remarque 1.3.8 (Processus d'Ornstein-Uhlenbeck et de Kolmogorov-Langevin à double puits)**.** *Si on reprend les deux exemples de la figure 1.1, on peut se faire une idée de leurs courbures à l'aide de l'algorithme indiqué dans la remarque 1.3.2. En effet, la figure 1.2 donne une approximation de*

$$t \mapsto \mathcal{W}_d\left(\delta_{-1} P_t, \delta_1 P_t\right),$$

où d est la distance en valeur absolue et $(P_t)_{t\geq 0}$ désigne, sur la figure de gauche, le semigroupe d'un processus d'Ornstein-Uhlenbeck, et, sur la figure de droite, celui du processus avec le potentiel non-convexe introduit dans l'exemple (1.2.4).

Remarque 1.3.9 (Sur l'équivalence $2 \Leftrightarrow 7$ et la distance intrinsèque)**.** *Soit $(P_t)_{t\geq 0}$ un semigroupe de Markov quelconque sur un espace polonais E. Si Γ désigne son opérateur* carré du champ *alors on peut définir la distance intrinsèque d_Γ sur E de ce semigroupe par*

$$\forall x, y \in E, d_\Gamma(x, y) = \sup_f \left\{ |f(x) - f(y)| \mid \|\Gamma f\|_\infty \leq 1 \right\}.$$

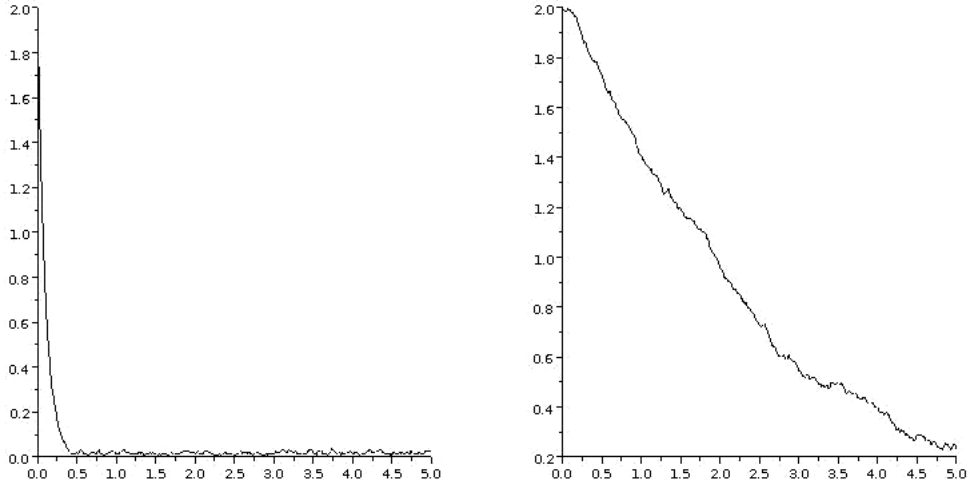

Figure 1.2 – Convergence vers l'équilibre de deux processus de Kolmogorov

L'inégalité (1.3) *du théorème 1.2.2 entraine donc que la courbure de Bakry-Émery coïncide avec la courbure de Wasserstein associée à la distance intrinsèque pour tout processus de Markov. Dans le cas particulier des processus de diffusion, la distance intrinsèque correspond à la distance usuelle.*

1.3.3.3 Processus stochastiquement monotone

Dans le chapitre 2, nous allons décrire la courbure de Wasserstein des processus unidimensionnelles et stochastiquement monotones. Plus précisément, on considérera un processus de Markov $(X_t)_{t\geq 0}$, défini sur une partie E de $\mathbb{R}$, qui vérifiera : pour tout $x, y \in E$, $x \geq y$, et $t \geq 0$, il existe deux copies de X, X^x et X^y, tel que $X_0^x = x$, $X_0^y = y$ et $X_t^x \geq X_t^y$. Pour ce type de processus, nous montrerons que nous avons

$$\partial_x \mathbb{E}_x \left[f(X_t) \right] = \mathbb{E}_x \left[f'(Y_t) e^{-\int_0^t V(Y_s) ds} \right],$$

pour toute fonction régulière f, $x \in E$ et $t \geq 0$. Ici $(Y_t)_{t\geq 0}$ est un certain processus auxiliaire et V une fonction explicite. La démonstration est basée sur une relation de commutation entre le gradient et le générateur. Avec cette formule, on voit facilement que la courbure est donnée par

$$\rho = \inf_{x \in E} V(x).$$

D'autres applications de cette formule seront donnés dans ce chapitre. Par exemple, en utilisant le fait que

$$\mathbb{E}_x\left[e^{-\int_0^t V(Y_s)ds}\right] \approx e^{-\lambda t},$$

où λ est la première valeur propre, de l'opérateur de type Schrödinger associé à cette formule de type Feynman-Kac, on montrera que la vitesse de convergence, en distance de Wasserstein, d'un processus de Kolmogorov-Langevin, vers son équilibre, est décrite par ce λ qui correspond aussi au trou spectral.

1.4 Application pour les Processus de Markov modulé

Ici on s'intéresse au comportement en temps long des processus de Markov modulés. Un processus de Markov modulé est construit à partir des ingrédients suivants :

- un espace polonais E et un espace fini F,
- une famille de processus de Markov $(Z^{(i)})_{i\in F}$,
- une famille de fonction positive $a(\cdot,i,j)_{i,j\in F^2}$.

On considère le processus $\mathbf{X}_t = (X_t, I_t)$ généré par

$$\mathbf{L}f(x,i) = \mathcal{L}^{(i)}f(x,i) + \sum_{j\in F} a(x,i,j)(f(x,j) - f(x,i))$$

pour toute fonction régulière f et pour tout $(x,i) \in E \times F$. Ici, $\mathcal{L}^{(i)}$ représente le générateur de $Z^{(i)}$. Le processus X n'est pas markovien au contraire du couple (X,I). Le processus I l'est si et seulement si a ne dépend pas de sa première composante.

1.4.1 Quelques exemples au comportement particulier

Avant de décrire nos résultats principaux, nous allons développer quelques exemples pour exposer les difficultés liées à ce type de modèle.

1.4.1.1 Explosion alors que chaque dynamique sous-jacente converge

L'exemple présenté ici vient de [BLMZ12c]. On considère le processus $\mathbf{X}$, sur $\mathbb{R}^2 \times \{-1,1\}$, généré par

$$\mathbf{L}f(x,i) = A_i\partial_x f(x,i) + a\left(f(x,-i) - f(x,i)\right),$$

où $a > 0$ et

$$A_{-1} = \begin{pmatrix} -1 & 3 \\ -1/3 & -1 \end{pmatrix} \text{ and } A_1 = \begin{pmatrix} -1 & -1/3 \\ 3 & -1 \end{pmatrix}.$$

Avec les notations précédentes, on a $\mathcal{L}^{(i)} = A_i \nabla_x$. Il est facile de voir que $Z^{(-1)}$ et $Z^{(1)}$ sont déterministes et convergent, exponentiellement vite, vers zero. Pourtant si a est suffisamment grand alors on a

$$\lim_{t \to +\infty} X_t = +\infty.$$

1.4.1.2 Convergence mais pas en variation totale

Ici, on suppose que $(I_t)_{t \geq 0}$ est une chaîne de Markov irréductible, à temps continu, sur un espace fini, avec mesure invariante ν. On suppose que X est continue sur $\mathbb{R}$ et est, entre les sauts de I, solution de

$$\partial_t X_t = -a(I_t) X_t,$$

Dans l'expression précédente $(a(i))_{i \in F}$ désigne une suite quelconque sur $\mathbb{R}$. On montre facilement que

$$X_t = e^{-\int_0^t a(I_s) ds} X_0.$$

Le théorème ergodique donne donc

$$\lim_{t \to +\infty} X_t = \begin{cases} 0 \text{ si } \sum_{i \in F} a(i)\nu(i) > 0 \\ +\infty \text{ si } \sum_{i \in F} a(i)\nu(i) < 0 \end{cases} \tag{1.9}$$

Ce qui signifie que si, en moyenne, le processus se rapproche de l'origine, alors il converge, alors que si, en moyenne, il s'éloigne, alors il diverge. Nos théorèmes principaux vont généraliser le premier point. Notons que pour cet exemple, nous avons que si $X_0 \neq 0$ alors le processus ne touche jamais zéro. On en déduit que la distance en variation totale entre la loi de X au temps t et son équilibre δ_0 est toujours égale à 1. Les techniques de type Lyapunov, ne fonctionnent donc pas ici.

Exemple 1.4.1 (Cas limite)**.** *Si l'espace d'état de I est $\{-1, 1\}$, si son état change à taux constant égal à 1 et si l'application a est l'identité, c'est-à-dire que $a : i \mapsto i$, alors la figure 1.3 représente une de ces trajectoires.*

1.4.2 Résultat principal

Supposons dans un premier temps que a ne dépend pas de sa première composante. C'est-à-dire que $a(\cdot, i, j) = a(i, j)$ pour tout $i, j \in F$. Dans ce cas I est une chaîne de Markov à temps continu. Si le couple (X, I) est ergodique alors I l'est aussi. Nous allons donc supposer que I est une chaîne

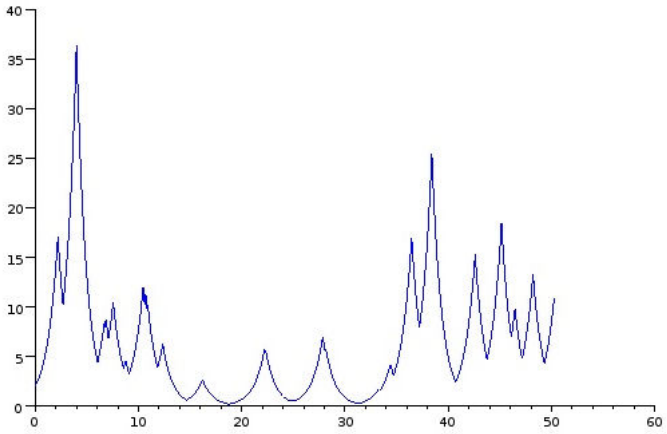

Figure 1.3 – Une trajectoire d'un processus de Markov modulé

ergodique, ou de manière équivalente qu'elle est irréductible. Notons $\rho(i)$ la courbure de Wasserstein du processus généré par $\mathcal{L}^{(i)}$, pour chaque $i \in F$. On a

Théorème 1.4.2 (Convergence exponentielle)**.** *Si a ne dépend pas de sa première composante, la chaîne I est irréductible, avec ν comme mesure invariante, et*

$$\sum_{i \in F} \rho(i)\nu(i) > 0,$$

alors le processus (X, I) converge, exponentiellement vite, en une distance de Wasserstein, vers une unique probabilité invariante.

Ce théorème nous dit que si en moyenne, on a tendance à converger, alors on converge. Ce théorème n'est pas notre résultat principal sur les processus modulés. En effet, nous établissons dans le chapitre 3, un théorème dans le cas où a dépend de sa première composante. Le théorème ci-dessus peut se démontrer, directement, avec l'approche de [BLMZ12c]. Cependant, cette approche échoue lorsque a n'est pas constant en sa première composante. Nous avons donc suivit une autre approche, utilisant une version faible des théorèmes de type Meyn-Tweedie [HMS11], qui a déjà eu diverses applications, comme par exemple l'étude de l'équation de Navier-Stockes bruité. Dans le chapitre 3, nous démontrerons le théorème 1.4.2 avec l'approche de [BLMZ12c] qui est plus simple et plus concise. Dans la sous-section qui suit, nous donnons une preuve similaire à celle de notre théorème plus compliqué dans le cadre plus simple du théorème 1.4.2.

1.4.3 Démonstration didactique du théorème 1.4.2

Nous allons démontrer le théorème 1.4.2 à l'aide d'une version plus faible du théorème d'Harris classique [HMS11]. Avant de détailler ce que l'on doit démontrer, ajoutons quelques notations. Soit $\mathbf{E} = E \times F$ et $\mathbf{P}$ le semigroupe du couple $\mathbf{X} = (X, I)$. Pour démontrer le théorème 1.4.2, il nous suffit de montrer les trois points qui suivent.

i) L'existence d'une fonction positive V, et de constantes $C, K, \lambda > 0$, tel que

$$\forall t \geq 0,\ \mathbb{E}\left[V(X_t)\right] \leq Ce^{-\lambda t}\mathbb{E}\left[V(X_0)\right] + K,$$

ii) L'existence d'une distance $\widetilde{d} : E \times E \to [0,1]$ tel que pour tout $A > 0$ il existe $\varepsilon_A > 0$ et $t_A > 0$ vérifiant, pour tout $t \geq t_A$,

$$\forall \mathbf{x}, \mathbf{y} \in \{V \leq A\},\ \mathcal{W}_{\widetilde{d}}(\delta_{\mathbf{x}}\mathbf{P}_t, \delta_{\mathbf{y}}\mathbf{P}_t) \leq 1 - \varepsilon_A$$

iii) L'existence d'une constante $\alpha \in (0,1)$, tel que pour tous $\mathbf{x}, \mathbf{y} \in \mathbf{E}$,

$$\widetilde{d}\,(\mathbf{x}, \mathbf{y}) < 1 \ \Rightarrow\ \forall t \geq 0, \mathcal{W}_{\widetilde{d}}(\delta_{\mathbf{x}}\mathbf{P}_t, \delta_{\mathbf{y}}\mathbf{P}_t) \leq \alpha\widetilde{d}\,(\mathbf{x}, \mathbf{y})\,.$$

Comme dans [BLMZ12c], la preuve de ces trois points découle du lemme suivant

Lemme 1.4.3 (Exponentielle des fonctionnelles additives des processus de Markov ergodiques)**.** *Si $(I_t)_{t\geq 0}$ est une chaîne de Markov, à temps continu, irréductible, sur un espace d'état fini F, avec comme mesure invariante ν, et α est une fonction tel que*

$$\sum_{i \in F} \nu(i)\alpha(i) > 0,$$

alors il existe $p \in (0,1]$ et $C, c, \eta > 0$ tel que pour tout $t \geq 0$, on ait

$$ce^{-\eta t} \leq \mathbb{E}\left[e^{-\int_0^t p\alpha(I_s)ds}\right] \leq Ce^{-\eta t}.$$

Nous allons supposer par la suite que la fonction $x \mapsto d(x, x_0)$ appartient au domaine de chaque générateur. cette hypothèse ne sera pas nécessaire dans les démonstrations du chapitre 3.

1.4.3.1 Démonstration du point i)

Pour tout $x_0 \in E$, on pose $V_{x_0} : x \mapsto d(x, x_0)$. Soit $i \in F$ et $P^{(i)}$ les semigroupes de $Z^{(i)}$. Pour tout $t \geq 0$, on a

$$\begin{aligned} P_t^{(i)} V_{x_0}(x_0) &= \mathcal{W}_d\left(\delta_{x_0} P_t^{(i)}, \delta_{x_0}\right) \\ &\leq \frac{1}{n} \sum_{k=0}^{n-1} \mathcal{W}_d\left(\delta_{x_0} P_{(k+1)t/n}^{(i)}, \delta_{x_0} P_{kt/n}^{(i)}\right) \\ &\leq \frac{e^{-\rho(i)t} - 1}{e^{-\rho(i)t/n-1} - 1} P_{k/n}^{(i)} V_{x_0}(x_0). \end{aligned}$$

Passant à la limite $n \to +\infty$ et en utilisant l'inégalité triangulaire, on trouve, pour tout $x \in E$,

$$P_t^{(i)} V_{x_0}(x) \leq e^{-\rho(i)t} V_{x_0}(x) + \frac{e^{-\rho(i)t} - 1}{-\rho} \mathcal{L}^{(i)} V_{x_0}(x_0).$$

Ici $\mathcal{L}^{(i)}$ est le générateur de $P_t^{(i)}$. En dérivant, on obtient

$$\mathcal{L}^{(i)} V_{x_0}(x) \leq -\rho(i) V_{x_0}(x) + K,$$

où $K = \max_{i \in F} \mathcal{L}^{(i)} V_{x_0}(x_0)$. Le lemme de Gronwall donne ensuite

$$\mathbf{P}_t V_{x_0}(x) \leq K \int_0^t \mathbb{E}\left[e^{-\int_s^t \alpha(I_u) du}\right] ds + V_{x_0}(x) \mathbb{E}\left[e^{-\int_0^t \alpha(I_s) ds}\right].$$

Finalement, en utilisant le lemme 1.4.3, l'inégalité de Jensen et la concavité de $x \mapsto x^p$, on prouve qu'il existe $p \in (0, 1]$ tel que $V = V_{x_0}^p$ est une fonction de Lyapunov, c'est-à-dire que le point i) est vérifié avec V.

1.4.3.2 Démonstration du point ii)

On choisit la distance $\widetilde{d}$, définie pour tout $(x, i), (y, j) \in \mathbf{E}$, par

$$\widetilde{d}((x, i), (y, j)) = \mathbf{1}_{i \neq j} + \mathbf{1}_{i=j} d(x, y)^p \wedge 1.$$

On fixe $\mathbf{x} = (x, i)$ et $\mathbf{y} = (y, j)$ et on va construire un couplage entre deux processus (X, I) et (Y, J) commençant en ces points et généré par $\mathbf{L}$. Premièrement, on considère que I et J sont indépendant pour tout temps. Puis, on fixe un temps $t_0 > 0$ et $i_0 \in F$ tel que $\rho(i_0) > 0$. La probabilité que $I = J = i_0$ sur $[t_0, t]$ est non nul et conditionné à cette événement, on peut coupler X et Y pour que

la distance $d(X,Y)^p$ décroit en moyenne. Cela vient de l'inégalité de Jensen et du fait que $p \leq 1$.On en déduit que l'inégalité voulue est satisfaite pour ce choix de $\mathbf{x}$ et $\mathbf{y}$. Il suffit de passer à l'infimum sur l'ensemble borné $\{V \leq A\}$ pour conclure.

1.4.3.3 Démonstration du point iii)

Si $\widetilde{d}((x,i),(y,j)) < 1$ alors $i = j$ et on peut donc coupler les composantes discrètes de tel façon à qu'elles restent égales pour tout temps. En utilisant ce couplage pour la composante discrète et le couplage optimal pour les autres composantes, on trouve

$$\forall t \geq 0, \mathbb{E}\left[\widetilde{d}((x,i),(y,j))\right] \leq \mathbb{E}\left[e^{-\int_0^t p\alpha(I_s)ds}\right] \widetilde{d}((x,i),(y,j)).$$

Le lemme 1.4.3 finit donc la démonstration de ce point.

1.4.3.4 Conclusion

En utilisant le [HSV11, Theorem 1.7], voir aussi le théorème 3.3.6, on trouve que, pour tout $\mathbf{x}, \mathbf{y} \in \mathbf{E}$ et $t \geq t_0$, on a

$$\mathcal{W}_{\mathbf{d}}\left(\delta_{\mathbf{x}}\mathbf{P}_t, \delta_{\mathbf{y}}\mathbf{P}_t\right) \leq Ce^{-\lambda t}\mathbf{d}(\mathbf{x},\mathbf{y}),$$

pour certaines constances $t_0, C, \lambda > 0$, et, où $\mathbf{d}$ est définie, avec les notations précédentes, par

$$\mathbf{d}(\mathbf{x},\mathbf{y}) = \sqrt{(\mathbf{1}_{i\neq j} + \mathbf{1}_{i=j}d(x,y)^p \wedge 1)(1 + d(x,x_0)^p + d(y,x_0)^p)},$$

pour un certain $x_0 \in E$.

Remark 1.4.4 (Sur les généralisations du théorème de Harris)**.** *Les théorèmes de type Harris reposent essentiellement sur deux hypothèses. La première est une fonction de Lyapunov, qui prouve que le processus passe la plupart de son temps dans un compact, et la deuxième hypothèse de type récurrence consiste, en général, à prouver que les compacts sont « small », c'est-à-dire que les lois au temps $t \geq 0$ ont toutes une partie à densité par rapport à la même mesure. Plusieurs travaux ont affaibli ces hypothèses. Pour l'hypothèse d'existence d'une fonction de Lyapunov, on pourra par exemple lire [DFG09], qui montre que si l'on peut trouver une fonction positive V, dont les ensembles de niveau sont compacts, qui vérifie*

$$\mathcal{L}V \leq -\Phi(V) + K,$$

où $\mathcal{L}$ est le générateur du processus dont on veut montrer la convergence, Φ une application concave et $K > 0$ une constante, alors, sous condition que les compacts sont « small », le processus converge

en variation totale à une vitesse sous-exponentielle. La deuxième hypothèse peut être affaiblie en utilisant la notion d'ensemble « $d-$small »qui prend plus en compte la topologie de l'espace [HMS11]. L'article [But12] synthétise ces deux approches.

Chapter 2

Gradient estimate and quantitative bounds for stochastically monotone jump-diffusions

2.1 Introduction

We are interested by a process which moves continuously for some random time and then jumps. It can represent some natural phenomena that can be observed at a great variety of scales. To give just few examples, let us simply mention the modeling of the Transmission Control Protocol (TCP) used for data transmission over the Internet [BCG+11, CMP10, GK10, GRZ04, LvL08, vLLO09], parasite evolution or size of cell in biology [BDMT11, BT11, Clo11, LP09], reliability and queuing [CD08, Dav93, Las04, RT00]. More precisely, this process $X = (X_t)_{t\geq 0}$ has an open interval $E \subset \mathbb{R}$ as state space and its infinitesimal generator is given, for any C^2 function $f : E \mapsto \mathbb{R}$ with compact support and $x \in E$ by

$$\begin{aligned} \mathcal{L}f(x) &= \sigma(x)f''(x) + g(x)f'(x) \\ &\quad + \int_E (f(x+y) - f(x))\eta(x, dy), \end{aligned} \tag{2.1}$$

where σ, g are two smooth functions and η is a, non-necessarily finite, positive measure. We assume that $\mathcal{L}$ generates a unique and non-explosive Feller process. Conditions are given in [Bec07, Dav93, Kol11] with full details on the domain of $\mathcal{L}$. The associated semigroup $(P_t)_{t\geq 0}$ is defined, for any

bounded function, by

$$P_t f(x) = \mathbb{E}[f(X_t)|X_0 = x] = \int_E f(y) P_t(x, dy).$$

By the Itô-Dynkin Theorem, it satisfies

$$P_t f(x) = f(x) + \int_0^t P_s \mathcal{L} f(x) ds = f(x) + \int_0^t \mathcal{L} P_s f(x) ds, \tag{2.2}$$

for all f in the domain of $\mathcal{L}$. If η is finite then we can write $\eta = rK$, where r is a non-negative function and K is a Markov kernel. In this case, X will not run through an infinite number of jumps in finite time and the dynamics is as follows. Between the jumps, this process evolves like a diffusion which satisfies the following stochastic differential equation:

$$dX_t = g(X_t)dt + \sqrt{2\sigma(X_t)} dB_t,$$

where $(B_t)_{t\geq 0}$ is a standard Brownian motion. At a random and inhomogeneous time T satisfying

$$\mathbb{P}\left(T > t \mid X_s,\ s \leq t\right) = \exp\left(-\int_0^t r(X_s) ds\right),$$

it jumps. Conditionally to X_{T-}, X_T is a random variable of law $K(X_{T-}, \cdot)$. Then, this process repeats these steps again. This class of processes is called hybrid processes in [Bec07]. When $\eta = 0$, it is a diffusion and when $\sigma = 0$, it is a piecewise deterministic Markov process (PDMP) [Dav93]. When η is not finite, this class of processes include all one-dimensional Lévy processes.

Several properties of this process were established in the literature. Our aim in this paper is to get quantitative estimates for the convergence to equilibrium of X. Using Lyapunov techniques, [CD08, Las04, RT00] give some conditions to have a geometric convergence. Nevertheless, this process is, in general, irreversible and it has infinite support. This makes Lyapunov techniques less efficient for the derivation of quantitative exponential ergodicity. Furthermore, another main difficulty is that entropy methods fails. In general, the invariant measure of the process does not satisfy a Poincaré or log-Sobolev inequality (see remark 2.4.5 and [Wu10]). In this work, we establish an intertwining relation of the form $\nabla P_t = S_t \nabla$, where S is a Feynman-Kac semigroup. As application, we obtain some explicit rates of convergence.

Overview of the main results. Before expressing our main results, let us introduce the assumptions holding in all the article. We assume that $(X_t)_{t\geq 0}$ is a Feller process and is stochastically monotone.

The second assumption means that for any $x^{(1)}, x^{(2)} \in E$, if $x^{(1)} \geq x^{(2)}$ then there exists a coupling $(X_t^{(1)}, X_t^{(2)})_{t\geq 0}$, starting from $(x^{(1)}, x^{(2)})$, whose each component is generated by $\mathcal{L}$, such that $X_t^{(1)} \geq X_t^{(2)}$ almost surely for all $t \geq 0$. Actually it is equivalent to

$$x \mapsto \int_a^\infty \eta(x, dy)$$

is non-decreasing and

$$x \mapsto \int_{-\infty}^a \eta(x, dy)$$

is non-increasing. With this assumption and some regularity assumptions, we are able to prove that for any C^2 function f and $x \geq 0$, we have

$$(P_t f)'(x) = \mathbb{E}\left[f'(Y_t)e^{-\int_0^t V(Y_s)ds} \mid Y_0 = x\right], \tag{2.3}$$

Where $(Y_t)_{t\geq 0}$ is an auxiliary Markov process and V is a potential. The generator $\mathcal{A}$ of $(Y_t)_{t\geq 0}$ will be introduced in the next section by the formula (2.7). In all the paper, we consider the following regularity assumptions:

- the functions g and σ belong to $C^\infty(E)$;
- we have $\eta(x, \{y \in E \mid x + y \in E\}) = 1$ and η is weak differentiable on E; that means that, for every $x \in E$, there exists a measure $\eta'(x, \cdot)$ verifying

  $$\int_E f(y)\eta'(x, dy) = \partial_x \int_E f(y)\eta(x, dy),$$

  for a sufficiently large class of function f.
- the set C_b^2 is included in the domains of $\mathcal{L}$ and $\mathcal{A}$ and is invariant over the semigoups generated by these two operators.
- there exists a unique solution to the martingale problem for $\mathcal{L}$ (resp. $\mathcal{A}$) on the set $C_c^2(E)$ of C^2 functions.

Sufficient conditions can be found in [Kol11, Theorem 5.9.4] and [Str03, Corollary 4.2.6]. We give some examples which easily verify theses assumptions in section 2.4. The proof of the formula (2.3) is based on a remarkable and simple intertwining relation between $\mathcal{L}$, ∇ and $\mathcal{A}$. It is a continuous time generalisation of [CJ10, Theorem 2.1]. Various functional inequalities are provided in this paper. Here, we provide others applications based on the long time behaviour of $(X_t)_{t\geq 0}$, including some quantitative bounds in Wasserstein distance. Let us recall that the Wasserstein distance between two

probability measures μ_1, μ_2 is defined by

$$\mathcal{W}(\mu_1, \mu_2) = \inf_{\nu \in \text{Marg}(\mu_1, \mu_2)} \iint_{E \times E} |x - y| \, \nu(dx, dy),$$

where $\text{Marg}(\mu_1, \mu_2)$ is the set of probability measures on E^2 with marginal distributions are μ_1 and μ_2, respectively. This infimum is attained [Vil09, Theorem 1.3]. The Kantorovich-Rubinstein duality [Vil09, Theorem 5.10] gives the following representation:

$$\mathcal{W}(\mu_1, \mu_2) = \sup_{g \in \text{Lip}_1} \int_E g \, d\mu_1 - \int_E g \, d\mu_2, \tag{2.4}$$

where Lip_1 is the set of Lipschitz function g satisfying $|g(x) - g(y)| \leq |x - y|$ for any $x, y \in E$. Our first result is

Theorem 2.1.1 (Wasserstein convergence for stochastically monotone jump-diffusion)**.** *If*

$$\rho = \inf_{x \in E} \left(-g'(x) - \int_E y \, \eta'(x, dy) \right) \geq 0, \tag{2.5}$$

then ρ is the largest constant which satisfies

$$\forall t \geq 0, \ \mathcal{W}\left(\mu P_t, \nu P_t\right) \leq e^{-\rho t} \mathcal{W}\left(\mu, \nu\right), \tag{2.6}$$

for all probability measures μ, ν. In particular, if $\rho > 0$ then X admits a unique invariant probability measure π and

$$\mathcal{W}\left(\mu P_t, \pi\right) \leq e^{-\rho t} \mathcal{W}\left(\mu, \pi\right).$$

The optimal constant in (2.6) is called the Wasserstein curvature in [Jou07]. We give several properties of this curvature in an appendix, including the proof of a Poincaré inequality under a reversibility assumption. With a totally different approach, [Wu10, Theorem 2.2] gives a different bound for a similar class of processes. With our approach, we obtain the optimality of the constant ρ in the inequality (2.6). We give others bounds in the next sections when $\rho \leq 0$. In particular, we obtain the following bound in the special case of Kolmogorov-Langevin processes:

Theorem 2.1.2 (Wasserstein exponential ergodicity for Kolmogorov-Langevin processes)**.** *Assume that $\eta = 0$, σ is constant and*

$$\lim_{|x| \to +\infty} g'(x) = -\infty,$$

then there exist a unique invariant probability measure π and $\lambda > 0$, such that for all $x \in \mathbb{R}$ there

exists $C_x > 0$ *such that*

$$\forall t \geq 0, \mathcal{W}(\pi, \delta_x P_t) \leq C_x e^{-\lambda t}.$$

The proof of this bound depends to the precedent expression of the gradient and a spectral analysis. Contrary to the usual proof, based on a Lyapunov-type argument, λ is explicit in term of eigenvalue; it correspond to the first non-null eigenvalue to $\mathcal{L}$ This theorem is a direct consequence of theorem 2.3.8 below. Finally, using theorem 2.1.1, we deduce a convergence in total variation in the special case of PDMP. Let us recall that, for any probability measure μ_1, μ_2, the total variation distance is defined by

$$d_{TV}(\mu_1, \mu_2) = \inf \mathbb{P}(X_1 \neq X_2),$$

where the infimun is taken over all couple (X_1, X_2) such that X_1, X_2 are respectively distributed according to μ_1, μ_2.

Theorem 2.1.3 (Total variation convergence for stochastically PDMP)**.** *If the following assumptions hold:*

i) the constant ρ*, introduced in* (2.5)*, satisfies* $\rho \geq \kappa > 0$*, for some* κ*,*

ii) there exist a non-negative function r *and a Markov kernel* K *satisfying* $\eta = rK$*,* $\sigma = 0$*,*

iii) we have $\inf_{x \in E} r(x) > 0 \wedge \sup_{y \in E} g'(y)$*,*

iv) there exists $C > 0$ *such that*

$$\forall x, y \in E,\ d_{TV}\left(K(x, \cdot), K(y, \cdot)\right) \leq C|x - y|,$$

then there exists a constant $\theta > 0$ *and a constant* $C_{\mu,\nu}$*, which only depends to* μ *and* ν*, such that*

$$d_{TV}\left(\mu P_t, \nu P_t\right) \leq C_{\mu,\nu} e^{-\theta t}.$$

In particular, there exists a probability measure π *which satisfies*

$$d_{TV}\left(\mu P_t, \pi\right) \leq C_{\mu,\pi} e^{-\theta t}.$$

All constants of this theorem are explicit as can be seen in section 2.3.3. Note that, if X have a diffuse part then its density is regular and it is easier to obtain a bound in total variation. To our knowledge, these theorems and their proofs are new, nevertheless there are some related results in the literature.

Outline. In Section 2.2, we explicit our gradient estimate, which will be an important and simple ingredient to prove our main theorems. The rest of paper is organised as follow. In Section 2.3,

we give the proofs and the applications of our main results. In section 2.4.1, we give some direct applications. Finally, at the end of the paper, we give an appendix on the Wasserstein curvature. This appendix completes the article [Jou07] and gives some others applications of our main results.

2.2 Gradient estimate via a Feynman-Kac semigroup

We begin by associating an auxiliary process Y to the Markov process X. This auxiliary process is a Markov process with generator $\mathcal{A}$ given by

$$\begin{aligned}\mathcal{A}f(x) &= \sigma(x)f''(x) + (\sigma'(x) + g(x))\, f'(x) \\ &+ \int_E (f(x+y) - f(x))\eta(x, dy) \\ &+ \int_E \int_0^y (f(x+u) - f(x))du\ \eta'(x, dy),\end{aligned} \tag{2.7}$$

for every $f \in C_c^2(E)$. The processes X and Y are connected via the following commutation relation:

$$\nabla \mathcal{L} = (\mathcal{A} - V)\nabla, \tag{2.8}$$

where

$$V(x) = -g'(x) - \int_E y\eta(x, dy).$$

The intertwining relation (2.8) is the infinitesimal version at time $t = 0$ of the formula (2.3). Let us state our first lemma which will be fundamental in the proof of our main results.

Lemma 2.2.1 (Intertwining relation and gradient estimate)**.** *If there exists $c \in \mathbb{R}$ such that*

$$\forall x \in E,\ V(x) \geq c,$$

then we have

$$\forall x \geq 0,\ (P_t f)'(x) = \mathbb{E}\left[f'(Y_t)e^{-\int_0^t V(Y_s)ds} \mid Y_0 = x\right], \tag{2.9}$$

for every $f \in C_c^2(E)$.

The proof of this lemma, which follows, is remarkably simple but it seems to be new in this general context.

Proof. Let $(S_t)_{t\geq 0}$ be the semigroup defined, for all $t \geq 0, x \in E$ and $g \in C_c^2$, by

$$S_t g(x) = \mathbb{E}\left[g(Y_t)e^{-\int_0^t V(Y_s)ds}\right].$$

Is is known that

$$S_t g(x) = g(x) + \int_0^t (\mathcal{A} - V)\, S_s g(x) ds. \tag{2.10}$$

Now let $x_0 \in E$ and $\mathcal{I}_{x_0}$ be the operator defined by

$$\mathcal{I}_{x_0} g(x) = \int_{x_0}^x g(y) dy.$$

The relation (2.8) is equivalent to the following one:

$$\mathcal{L}\mathcal{I}_{x_0} g(x) - \mathcal{L}\mathcal{I}_{x_0} g(x_0) = \mathcal{I}_{x_0}(\mathcal{A} - V) g(x_0),$$

for every $x \in E$ and $g \in C_c^2$. Now using (2.10) and the previous expression, we have

$$\mathcal{I}_{x_0} S_t g(x) = \mathcal{I}_{x_0} g(x) + \int_0^t \mathcal{L}\mathcal{I}_{x_0}(S_s g)(x) ds - \int_0^t \mathcal{L}\mathcal{I}_{x_0}(S_s g)(x_0) ds.$$

Now we set

$$\varphi_g(t) = \int_0^t \mathcal{L}\mathcal{I}_{x_0}(S_s g)(x_0) ds,$$

and we introduce the semigroup $(T_t)_{\geq 0}$ defined, for every $f \in C_c^2$, by

$$T_t f(x) = \mathcal{I}_{x_0} S_t f'(x) + f(x_0) + \varphi(t).$$

We see that $T_0 f = f$ and for every $t \geq 0$,

$$T_t f = T_0 f + \int_0^t \mathcal{L} T_s f ds.$$

And thus, by uniqueness $T = P$. Finally S is continuous and then $P = \mathcal{I}S + \varphi$ is derivable and we have the announced result. □

Remark 2.2.2 (Propagation of monotonicity)**.** *Let Q be a Markov semigroup of a real Markov process. If, similarly to (2.9), its gradient is equal to a Feynman-Kac semigroup then it satisfies*

$$f' \geq 0 \Rightarrow Q_t f' \geq 0,$$

for every $f \in C_c^2(E)$. It is known to be equivalent to the stochastic monotonicity of Q. We deduce that our commutation and intertwining relation is equivalent to the stochastically monotonicity.

Remark 2.2.3 (Others gradients)**.** *Lemma 2.2.1 possesses a natural analogue for others gradients.*

Indeed, for any smooth and increasing function a, we have

$$\nabla_a \mathcal{L} = (\mathcal{A}_a - V_a)\nabla_a,$$

where, for every $f \in C_c^2(E)$ and $x \in E$, the previous operators are defined by $\nabla_a f = af'$,

$$\mathcal{A}_a f(x) = \sigma(x)f''(x) + g_a f'(x) + \int_E (f(x+y) - f(x))\eta_a(x, dy),$$

$$g_a(x) = \sigma'(x) + g(x) - 2\sigma(x)\frac{a'(x)}{a(x)},$$

$$\int_E f(y)\ \eta_a(x, dy) = \int_E \frac{f(y)}{a(x+y)}\ \eta(x, dy) + \int_E \int_0^u \frac{f(y)}{a(x+y)} dy\ \eta'(x, du)$$

and

$$V_a(x) = \sigma(x)\frac{a''(x)}{a(x)} - g' + \frac{a'}{a}g_a - \int_E y\eta(x, dy) + \frac{1}{a(x)}\int_E (a(x+y) - a(x))\eta_a(x, dy).$$

Thus we have

$$(P_t f)'(x) = \frac{1}{a(x)}\mathbb{E}\left[a(Y_t^{(a)})f'(Y_t^{(a)})e^{-\int_0^t V_a(Y_s^{(a)})ds} \mid Y_0^{(a)} = x\right],$$

where $(Y_t^{(a)})_{t\geq 0}$ is a Markov process generated by $\mathcal{A}_a$.

2.3 Proof of the main results

In this section, we prove and complete the results announced in the introduction.

2.3.1 Wasserstein convergence

We begin to prove Theorem 2.1.1 and then we give some corollaries.

2.3.1.1 Proof of Theorem 2.1.1

Let us fix $f \in \mathrm{Lip}_1 \cap C_c^2(E)$ and $x, y \geq 0$. By Lemma 2.2.1, we have

$$\begin{aligned}
|P_t f(x) - P_t f(y)| &\leq \sup_{z\geq 0} |(P_t f)'(z)||x-y| \\
&\leq \sup_{z\geq 0} \mathbb{E}\left[|f'(Y_t)| e^{-\int_0^t V(Y_s)ds} | Y_0 = z\right] |x-y| \\
&\leq \sup_{z\geq 0} \mathbb{E}\left[e^{-\int_0^t V(Y_s)ds} | Y_0 = z\right] |x-y|.
\end{aligned}$$

So that dividing by $|x-y|$ and taking suprema entail the following inequality

$$\sup_{\substack{x,y\in E \\ x\neq y}} \frac{|P_t f(x) - P_t f(y)|}{|x-y|} \leq \sup_{z\geq 0} \mathbb{E}\left[\exp\left(-\int_0^t V(Y_s)\, ds\right) \mid Y_0 = z\right].$$

Now, taking $f(x) = x$, we show that

$$\sup_{x\neq y} \frac{\mathcal{W}(\delta_x P_t, \delta_y P_t)}{|x-y|} = \sup_{z\geq 0} \mathbb{E}\left[e^{-\int_0^t V(Y_s)ds} | Y_0 = z\right].$$

Let ρ be the optimal constant verifying (2.6). Using the last expression, we have $\rho \geq \inf_{z\in E} V(z)$. On the other hand, it also gives that, for every $z \in E$,

$$\rho \leq -\frac{1}{t} \ln \mathbb{E}\left[e^{-\int_0^t V(Y_s)ds} | Y_0 = z\right].$$

Taking the limit $t \to 0$ and the supremum over $z \in E$, it achieves the proof.

Remark 2.3.1 ($h-$transform and first eigenvalue)**.** *Assume that $\mathcal{A} - V$ has a first eigenvalue $\lambda > 0$ such that its eigenvector ψ is positive. Using an $h-$transform with the space-time harmonic function $h = e^{-\lambda t}\psi$, we get for any function $f \in$ Lip$_1 \cap C_c^2(E)$,*

$$\mathbb{E}\left[f'(Y_t) e^{-\int_0^t V(Y_s)ds}\right] = e^{-\lambda t}\psi(x)\mathbb{E}\left[\frac{f'(Z_t)}{\psi(Z_t)}\right] \leq e^{-\lambda t}\mathbb{E}\left[\frac{\psi(Z_0)}{\psi(Z_t)}\right],$$

where $(Z_t)_{t\geq 0}$ is another Markov process. Then, if ψ is smooth enough, the Wasserstein decay is in order to $e^{-\lambda t}$. Section 2.3.2 gives an application for Kolmogorov-Langevin processes with non convex potential.

Remark 2.3.2 (A proof by coupling)**.** *By stochastically monotonicity, for any $x < y$, there exists a*

coupling (X^x, X^y), with marginals generated by $\mathcal{L}$, which start from (x, y), and

$$\forall t \geq 0,\ X_t^x \leq X_t^y \quad a.s.$$

Using this coupling (X^x, X^y), we have

$$\begin{aligned}\mathcal{W}(\delta_x P_t, \delta_y P_t) &\leq \mathbb{E}[|X_t^x - X_t^y|] \\ &\leq \mathbb{E}[X_t^y] - \mathbb{E}[X_t^x] \\ &\leq P_t(id)(y) - P_t(id)(x) \\ &\leq \sup_{z\in E} \left(P_t(id)\right)'(z)|x-y|.\end{aligned}$$

The bound for $\sup_{z\in E} \left(P_t(id)\right)'(z)$ can be found using the generator. Nevertheless, this proof did not give any information about the optimality. Our approach confirms the optimality of this coupling. As remarked in [CMP10], this one favours the simultaneous jumps.

Remark 2.3.3 (Another bound when Y is ergodic)**.** *Assume that Y is ergodic with invariant probability measure π_A. The previous calculation gives*

$$\mathcal{W}(\delta_x P_t, \delta_y P_t) \leq \sup_{z\in E} \mathbb{E}\left[e^{-\int_0^t V(Y_s)ds} \mid Y_0 = z\right] |x-y|.$$

Using Jensen's inequality, we have

$$\mathbb{E}\left[e^{-\int_0^t V(Y_s)ds}\right] \geq \exp\left(-t \times \mathbb{E}\left[\frac{1}{t}\int_0^t V(Y_s)ds\right]\right),$$

and the ergodic theorem gives

$$\mathbb{E}\left[\frac{1}{t}\int_0^t V(Y_s)ds\right] \approx \int_E V d\pi_A.$$

We deduce that the rate of convergence is, at most, $\int_E V d\pi_A$.

Remark 2.3.4 (Kolmogorov distance and duality process)**.** *In [Kol11], some theorems, like [Kol11, Theorem 5.9.1], give the existence of a dual process. More precisely, they prove that if $(X_t)_{t\geq 0}$ is stochastically monotone then there exists a process $(\widetilde{X}_t)_{t\geq 0}$ satisfying*

$$\mathbb{P}\left(\widetilde{X}_t \leq x \mid \widetilde{X}_0 = \widetilde{x}\right) = \mathbb{P}\left(X_t \geq \widetilde{x} \mid X_0 = x\right),$$

for every, $x, \widetilde{x} \in E$ and $t \geq 0$. This give a bound for the Kolmogorov distance. Indeed, we have

$$\begin{aligned} d_{Kol}(\delta_x P_t, \delta_y P_t) &= \sup_{\widetilde{x} \in E} |\mathbb{P}(X_t \geq \widetilde{x} \mid X_0 = x) - \mathbb{P}(X_t \geq \widetilde{x} \mid X_0 = x)| \\ &= \sup_{\widetilde{x} \in E} \mathbb{P}\left(\widetilde{X}_t \in (y, x] \mid \widetilde{X}_0 = \widetilde{x}\right), \end{aligned}$$

for any $x \geq y$.

2.3.1.2 Wasserstein convergence when $\rho = 0$

Let us state a criterion for an exponential convergence when $\rho = 0$.

Theorem 2.3.5 (Exponential decay when the curvature is null)**.** *If $V \geq 0$, Y is irreducible, there exist a compact set $K = [a, b]$ and a constant $\varepsilon > 0$ such that*

$$\forall x \notin K, V(x) \geq \varepsilon,$$

then there exist $\tilde{t} \geq 0$ and $\kappa > 0$ such that

$$\forall t \geq \tilde{t},\ \mathcal{W}(\mu P_t, \nu P_t) \leq e^{-\kappa t}\, \mathcal{W}(\mu, \nu),$$

for any probability measure μ, ν.

Notice that, if there exist $x \in E$ such that $V(x) = 0$ then the Wasserstein curvature is null.

Proof. The proof is adapted to [MT06, Section 5.1]. Let

$$D(t, x) = \mathbb{E}_x \left[\exp\left(- \int_0^t V(Y_s) ds \right) \mid Y_0 = x \right],$$

and $\bar{D}(t) = \sup_{x \in E} D(t, x)$. It is easy to see that for all $x \in E$, $D(\cdot, x)$ and $\bar{D}$ are non increasing. Furthermore, if $D(t, x) = 1$ and $Y_0 = x$ then $V(Y_s) = 0$ almost surely, for all $s \leq t$. Then, there exists $t_0 > 0$ such that we have

$$\forall x \in E,\ \forall t > t_0, D(t, x) < 1. \tag{2.11}$$

Now, we begin to prove the existence of $t_1 \geq 0$ such that

$$\forall t \geq t_1,\ \bar{D}(t) < 1.$$

If $x \notin K$ then we set $\tau = \inf\{t \geq 0 \mid Y_t \in K\}$. We have,

$$\begin{aligned} D(t,x) &= \mathbb{E}_x\left[\mathbf{1}_{\tau<t}\exp\left(-\int_0^t V(Y_s)ds\right)\right] + \mathbb{E}_x\left[\mathbf{1}_{\tau\geq t}\exp\left(-\int_0^t V(Y_s)ds\right)\right] \\ &\leq \mathbb{E}_x\left[\mathbf{1}_{\tau<t}\exp\left(-\int_0^t V(Y_s)ds\right)\right] + e^{-\varepsilon t}. \end{aligned}$$

And

$$\begin{aligned} \mathbb{E}_x\left[\mathbf{1}_{\tau<t}\exp\left(-\int_0^t V(Y_s)ds\right)\right] &\leq \mathbb{E}_x\left[\mathbf{1}_{\tau<t}e^{-\varepsilon\tau}\mathbb{E}_x\left[\exp\left(-\int_\tau^t V(Y_s)ds\right) \mid \mathcal{F}_\tau\right]\right] \\ &\leq \mathbb{E}_x\left[\mathbf{1}_{\tau<t}e^{-\varepsilon\tau}\max_{c\in\{a,b\}} D(t-\tau,c)\right] \\ &\leq \mathbb{E}_x\left[\mathbf{1}_{\tau<t/2}e^{-\varepsilon\tau}\max_{c\in\{a,b\}} D(t-\tau,c)\right] \\ &+ \mathbb{E}_x\left[\mathbf{1}_{t/2\leq\tau<t}e^{-\varepsilon\tau}\max_{c\in\{a,b\}} D(t-\tau,c)\right] \\ &\leq \max_{c\in\{a,b\}} D(t/2,c) + e^{-\varepsilon t/2}, \end{aligned}$$

where $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$ is the natural filtration associated to Y. Thus,

$$\sup_{x\notin K} D(t,x) \leq \max_{c\in\{a,b\}} D(t/2,c) + e^{-\varepsilon t} + e^{-\varepsilon t/2}.$$

We deduce that $\limsup_{t\to+\infty}\sup_{x\notin K} D(t,x) < 1$ and the existence of t_1 such that for all $t \geq t_1$,

$$\sup_{x\notin K} D(t,x) < 1.$$

The Feynman-Kac semigroup is continuous on K. So, we deduce that $\bar{D}(t) < 1$ for all $t \geq \tilde{t} = \max(t_0,t_1)$. Lemma 2.4.12, in the appendix, ends the proof.

Remark 2.3.6 (A link with the quasi-stationary distributions (QSD))**.** *If $V \geq 0$ then we have another representation of the gradient of $(P_t)_{t\geq 0}$. Indeed, we have*

$$\partial_x P_t f(x) = \mathbb{E}[f'(Y_t)\mathbf{1}_{\{t<\tau\}} \mid Y_0 = x],$$

where Y is a Markov process generated by $\mathcal{A}$ and τ is a random variable satisfying

$$\mathbb{P}(\tau > t \mid Y_s, s \leq t) = \exp\left(-\int_0^t V(Y_s)ds\right).$$

If Y admits a Yaglom limit, that is, there exists μ such that

$$\lim_{t\to+\infty} \mathbb{E}[f'(Y_t) \mid Y_0 = x, t < \tau] = \int f d\mu,$$

then we have $\mathbb{P}(\tau > t) \sim e^{-\theta t}$, for some $\theta > 0$. And thus, if π is the invariant distribution of X then

$$\begin{aligned}\left|P_t f(x) - \int_E f d\pi\right| &= \left|\int_E \int_{[x,y]} \mathbb{E}[f'(Y_t) \mid Y_0 = u] du \, d\pi(y)\right| \\ &\leq Ce^{-\theta t}.\end{aligned}$$

Remark 2.3.7 (On the usage of another gradient)**.** *Another way to prove an exponential decay in the case of non positive curvature is to change the distance. Let A be a smooth, positive and increasing function, the following mapping defines a distance:*

$$\forall x, y \in E, d_A(x, y) = |A(x) - A(y)| = \int_{x\wedge y}^{x\vee y} A'(u) du.$$

Our proofs are generalisable for this distance. Indeed, if $x \geq y$ and f is smooth enough then we have

$$|P_t f(x) - P_t f(y)| \leq \int_y^x A'(u) \nabla_{1/A'} P_t f(u) du.$$

This method was efficient for the $M/M/1$ in [CJ10] and it can improve the rate of convergence. Nevertheless, it does not generalise the field of concerned processes; that are the stochastically monotone processes. Another way to interpret this method is to consider $(A(X_t))_{t\geq 0}$, instead of $(X_t)_{t\geq 0}$. Indeed, as A is an increasing bijection, $(A(X_t))_{t\geq 0}$ is also a stochastically monotone Markov process.

□

2.3.2 The special case of Kolmogorov-Langevin diffusions

In this subsection, we consider, in detail, the simple case of Kolmogorov-Langevin diffusions. More precisely, here, we consider the process belonging to $E = \mathbb{R}$ and satisfying

$$\forall t \geq 0, \; dX_t = \sqrt{2} dB_t - q'(X_s) ds,$$

where q is C^∞ and $(B_t)_{t\geq 0}$ is a standard Brownian motion. We have changed the notation to be closer to the usual notations. It is already known that, under suitable assumptions, this process converges to the Boltzmann-Gibbs measure $\pi(du) = e^{-q(u)} du/\mathcal{Z}$, where $\mathcal{Z}$ is a renormalizing constant. Theorem

2.1.1 shows that X converges exponentially with rate $\rho = \inf_{z\in\mathbb{R}} q''(z)$. The special case where $\rho = 0$ was studied in [MT06, Section 5]. when $\rho < 0$, we have

Theorem 2.3.8 (Wasserstein exponential ergodicity)**.** *Assume that*

$$\lim_{|x|\to+\infty} q''(x) = +\infty,$$

then there exists $\lambda > 0$ *such that*

$$\forall x, y \in \mathbb{R},\ \exists C_{x,y} > 0,\ \forall t \geq 0,\ \mathcal{W}\left(\delta_x P_t, \delta_y P_t\right) \leq C_{x,y} e^{-\lambda t}.$$

Furthermore for all $x \in \mathbb{R}$ *there exists* C_x *such that*

$$\forall t \geq 0, \mathcal{W}\left(\pi, \delta_x P_t\right) \leq C_x e^{-\lambda t}.$$

Notice that λ is "explicit": it is the first, non-null, eigenvalue to the operator

$$f \mapsto -f'' + f(q''/2 + q'^2/4).$$

Or equivalently it is also the first, non-null, eigenvalue to $\mathcal{L}f \mapsto: f'' - q'f'$.

Proof. By Lemma 2.2.1, we have, for all $f \in C_c^2(\mathbb{R})$,

$$\partial_x P_t f = S_t f',$$

where $(S_t)_{t\geq 0}$ is the Feynman-Kac semigroup generated by $\mathcal{A} - V$; recall that

$$(\mathcal{A} - V)f = f'' - q'(x)f' - q''f.$$

As said in Remark 2.3.1, we will do an $h-$transform. Let H be the closure on $L^2(\pi)$ of the operator defined by

$$Hf = -f'' + f\left(\frac{q''}{2} + \frac{q'^2}{4}\right),$$

for every $f \in C_c^2(\mathbb{R})$. There exist a unique positive function $\varphi \in L^2 \cap C^\infty$ and a real number $\lambda > 0$ such that $H\varphi = \lambda\varphi$. See [BS83, Theorem 3.1 p. 57] and [BS83, chapter 2] for further details. Now, let $(Q_t)_{t\geq 0}$ be defined for all continuous and bounded function f by

$$Q_t(f) = \frac{e^{\lambda t}}{\varphi} e^{-q/2} S_t(f\varphi e^{q/2}). \tag{2.12}$$

for every $f \in C_c^2(\mathbb{R})$, we have

$$\begin{aligned}\partial_t Q_t(f) &= \lambda \frac{e^{\lambda t}}{\varphi} e^{-q/2} S_t(f\varphi e^{q/2}) + \frac{e^{\lambda t}}{\varphi} e^{-q/2} S_t(\mathcal{L}_{\mathbb{S}}(f\varphi e^{q/2})) \\ &= \frac{e^{\lambda t}}{\varphi} e^{-q/2} S_t(f e^{q/2} H\varphi) + \frac{e^{\lambda t}}{\varphi} e^{-q/2} S_t(-e^{q/2} H(f\varphi)) \\ &= \frac{e^{\lambda t}}{\varphi} e^{-q/2} S_t(e^{q/2}\varphi G f) \\ &= Q_t(Gf),\end{aligned}$$

where

$$Gf = f'' + 2\frac{\varphi'}{\varphi} f'.$$

The relation (2.12) gives

$$\partial_x P_t f(x) = e^{-\lambda t} e^{q(x)/2} \varphi(x) \mathbb{E}\left[\frac{f'(Y_t)}{\varphi(Y_t)} e^{-q(Y_t)/2} \mid Y_0 = x \right],$$

for every $f \in C_c^2(\mathbb{R})$, where Y is a Kolmogorov-Langevin process generated by G. Thus, if $f \in \mathrm{Lip}_1 \cap C_c^2(\mathbb{R})$ then,

$$|\partial_x P_t f(x)| \leq e^{-\inf_{z\in\mathbb{R}} q(z)/2} e^{-\lambda t} e^{q(x)/2} \varphi(x) Q_t\left(\frac{1}{\varphi}\right)(x).$$

But

$$G\frac{1}{\varphi} = \frac{-\varphi''}{\varphi^2} = \frac{1}{\varphi}\left(\lambda - \frac{q''}{2} - \frac{q'^2}{4}\right) \leq C - B\frac{1}{\varphi},$$

for some $B, C > 0$. We deduce that $Q_t(1/\varphi)$ is bounded. Hence, the first inequality of the theorem holds with

$$C(x,y) = D \times \sup_{z\in[x,y]} e^{q(z)/2} \left(1 + \varphi(z)\right) |x - y|,$$

for some $D > 0$. Furthermore the Cauchy-Schwarz inequality gives that for all $x \in \mathbb{R}$, we have $\int_{\mathbb{R}} C(x,y)\pi(dy) < +\infty$. □

Remark 2.3.9 ($h-$transform and Schödinger equation)**.** *The transformation* (2.12) *is usual in the study of schrödinger equation [Pin95]. It has many applications in the study of processes with killing [MV11] and branching measures [Clo11].*

Remark 2.3.10 (Ornstein-Uhlenbeck process)**.** *If $q(x) = \mu x^2/2$ then the assumptions of the theorem do not hold. But we can follow the proof step by step. The mapping $\varphi : x \mapsto e^{-\mu x/2}$ is an eigenvector*

of H with respect to the eigenvalue μ. So, we find that $G = \mathcal{L}$ and

$$\mathcal{W}\left(\delta_x P_t, \delta_y P_t\right) \leq e^{-\mu t}|x-y|.$$

In this case, $\mathcal{L}$ is the generator of the Ornstein-Uhlenbeck process and the previous expression is an equality. This example points the sharpness of our method.

2.3.3 Total variation convergence for PDMP

Let X be a piecewise deterministic Markov process (PDMP) on $E \subset \mathbb{R}$. More precisely, we consider that X is generated by

$$\mathcal{L}f(x) = g(x)f'(x) + r(x)\int_E (f(y) - f(x))K(x, dy),$$

where K is a Markov kernel and g, r are two smooth functions such that r is non-negative. We assume again that X is stochastically monotone. Theorem 2.1.1 gives a convergence with rate ρ given by

$$\rho = \inf_{x\in E}\left(-g'(x) + r(x)\partial_x \int_E (x-y)K(x, dy) + r'(x)\int_E (y-x)K(y, dx)\right).$$

If ρ is positive and K "contracts in total variation" then we are able to prove an exponential decay in total variation distance. Hereafter, we will use the following notations:

$$\bar{g'} = \sup_{z\in E} g'(z) \text{ and } \underline{r} = \inf_{z\in E} r(z).$$

Theorem 2.3.11 (Total variation convergence for monotone PDMP)**.** *If the following assumptions hold:*

i) there exists $\kappa > 0$ such that $\rho \geq \kappa$,
ii) $\underline{r} > 0 \wedge \bar{g'}$,
iii) there exists $C > 0$ such that

$$\forall x, y \in E,\ d_{TV}\left(K(x, \cdot), K(y, \cdot)\right) \leq C|x-y|,$$

then

$$d_{TV}\left(\mu P_t, \nu P_t\right) \leq K_{\mu,\nu} e^{-\theta t},$$

where $\theta = \frac{\kappa \underline{r}}{\kappa + \underline{r}}$ *and*

$$K_{\mu,\nu} = \left(\kappa\left(C + \frac{\bar{r'}}{\underline{r}}\right)\frac{\underline{r}}{\underline{r} - \bar{g'}}\mathcal{W}(\mu,\nu)\right)^{\frac{\underline{r}}{\underline{r}+\kappa}} + \kappa^{\frac{\kappa}{\underline{r}+\kappa}}\left(\left(C + \frac{\bar{r'}}{\underline{r}}\right)\frac{\underline{r}}{\underline{r} - \bar{g'}}\mathcal{W}(\mu,\nu)\right)^{\frac{\underline{r}+2\kappa}{\underline{r}+\kappa}}.$$

Remark 2.3.12 (Optimality)**.** *Let* $(X_t)_{t\geq 0}$ *be the Markov process, on* $E = \mathbb{R}_+^*$*, generated by* L*, be defined for every* $f \in C_c^2$ *and* $x \in E$ *by*

$$Lf(x) = -\mathbf{g}xf'(x) + r(x)\int_{\mathbb{R}_+}(f(x+u) - f(x))\lambda e^{-\lambda u}du.$$

Here $\mathbf{g}, \lambda > 0, r > 0$*. It is a simple model of storage: the current stock decreases exponentially, and increases at inhomogeneous random times by a random amount (distributed following an exponential variable). We deduce directly, from our main theorem, that if* r *is increasing then for any* $x, y \geq 0$*,*

$$\forall t \geq 0, \mathcal{W}(\delta_x P_t, \delta_y P_t) \leq e^{-(\mathbf{g} + \frac{1}{\lambda}\inf_{x\geq 0} r'(x))t}|x - y|.$$

On this example, the constants of the previous theorem are

$$K_{\mu,\nu} = \left(\rho\left(\frac{1}{\lambda} + \frac{\bar{r'}}{r(0)}\right)\frac{r(0)}{r(0) - \mathbf{g}}\mathcal{W}(\mu,\nu)\right)^{\frac{r(0)}{r(0)+\rho}} + \rho^{\frac{\rho}{r(0)+\rho}}\left(\left(\frac{1}{\lambda} + \frac{\bar{r'}}{r(0)}\right)\frac{r(0)}{r(0) - \mathbf{g}}\mathcal{W}(\mu,\nu)\right)^{\frac{r(0)+2\rho}{r(0)+\rho}}.$$

and

$$\theta = \frac{\rho r(0)}{\rho + r(0)}.$$

In particular, if r *is constant then we have*

$$d_{TV}(\mu P_t, \nu P_t) \leq e^{-t\frac{\mathbf{g}r}{\mathbf{g}+r}}K_{\mu,\nu}.$$

This rate is not optimal [BCG+11]. Our approach is similar to [BCG+11], we build a coupling such that the components are closer on $[0, s]$ *and we change the coupling to stick the components on* $[t - s, t]$*. In [BCG+11], the time* s *is random while in our proof, it is deterministic. Here,* s *is not random because when* r *is not constant the countable process associated to the jumps is not a Poisson process.*

The proof is based on the following lemma, proved via coupling argument, whose proof postponed at the end of proof of Theorem 2.3.11

Lemma 2.3.13 (Local total variation estimate)**.** *Let $x > y$ and $t \geq 0$, under the same assumptions of Theorem 2.3.11, we have*

$$d_{TV}(\delta_x P_t, \delta_y P_t) \leq e^{-t\underline{r}} + |x-y| \left(C + \frac{\bar{r'}}{\underline{r}} \right) \frac{\underline{r}}{\underline{r} - \bar{g'}}.$$

Here $\bar{r'} = \sup_{z \in E} r'(z)$.

Proof of Theorem 2.3.11. The previous expression is equivalent to

$$d_{TV}\left(\mu P_t, \nu P_t\right) \leq e^{-t\underline{r}} + \left(C + \frac{\bar{r'}}{\underline{r}} \right) \frac{\underline{r}}{\underline{r} - \bar{g'}} \, \mathcal{W}(\mu, \nu),$$

for any μ and ν which have a first moment. As $\rho > 0$, we deduce that, for all $s \leq t$,

$$\begin{aligned} d_{TV}\left(\mu P_t, \nu P_t\right) &= d_{TV}\left((\mu P_s) P_{t-s}, (\nu P_s) P_{t-s}\right) \\ &\leq e^{-(t-s)\underline{r}} + e^{-\rho s} \left(C + \frac{\bar{r'}}{\underline{r}} \right) \frac{\underline{r}}{\underline{r} - \bar{g'}} \mathcal{W}\left(\mu, \nu\right). \end{aligned}$$

And thus, $d_{TV}\left(\mu P_t, \nu P_t\right) \leq K_{\mu,\nu} e^{-\theta t}$ □

Proof of Lemma 2.3.13. Let us consider the coupling (X, Y), starting from (x, y), and generated by G, be defined for every $f \in C_c^2(\mathbb{R}_+^* \times \mathbb{R}_+^*)$ and $(x, y) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$ by

$$\begin{aligned} Gf(x,y) &= g(x)\partial_x f(x,y) + g(y)\partial_y f(x,y) \\ &\quad + (r(x) \wedge r(y)) \int_{E \times E} (f(u,v) - f(x,y)) \mathbb{K}((x,y), d(u,v)) \\ &\quad + (r(x) - r(y))^+ \int_E (f(u,y) - f(x,y)) K(x, du) \\ &\quad + (r(y) - r(x))^+ \int_E (f(x,u) - f(x,y)) K(y, du). \end{aligned}$$

In the last expression, $\mathbb{K}$ is the Markov kernel, on $E \times E$, which satisfies

$$\int_{E \times E} \mathbf{1}_{u \neq v} \mathbb{K}((x,y), d(u,v)) = d_{TV}\left(K(x, \cdot), K(y, \cdot)\right).$$

The dynamics of this coupling is as follows.

– It start from $(X_0, Y_0) = (x, y)$ and for all $t < T$, $X_t' = g(X_t)$ and $Y_t' = g(Y_t)$.

– The law of the time T is defined by

$$\mathbb{P}(T > t) = \exp\left(-\int_0^t r(x_s) \vee r(y_s) ds\right),$$

where $x_0 = x,\ y_0 = y$ and $x'_t = g(x_t),\ y'_t = g(y_t)$.

– At time T, we toss a coin B such that

$$\mathbb{P}(B = 0 \mid T) = \frac{r(x_T) \wedge r(y_T)}{r(x_T) \vee r(y_T)} \text{ and } \mathbb{P}(B = 1 \mid T) = \frac{|r(x_T) - r(y_T)|}{r(x_T) \vee r(y_T)}.$$

If $B = 0$ then the two trajectories jump simultaneously and if $B = 1$ only one component jumps.

– If the two trajectories jump in the same time then we stick them.

– We repeat these steps starting from (X_T, Y_T).

Since to stick them before is impossible, we would like to stick the trajectories at the first jump, and so maximise the quantity $\mathbb{P}(X_t = Y_t)$. More precisely, we have

$$\mathbb{P}(X_t = Y_t) \geq \mathbb{P}(X_T = Y_T, t \geq T, B = 0).$$

And as we have

$$|x_t - y_t| = x_t - y_t \leq e^{T \sup_{z \in E} g'(z)} |x - y|,$$

we deduce,

$$\begin{aligned}
\mathbb{P}(X_T = Y_T, t \geq T, B = 0) \geq & \mathbb{E}\left[\mathbf{1}_{t \geq T, B=0} \mathbb{P}(X_T = Y_T \mid (T, B))\right] \\
\geq & \mathbb{E}\left[\mathbf{1}_{t \geq T, B=0} \left(1 - C|x - y|e^{T\bar{g'}}\right)\right] \\
\geq & \mathbb{E}\left[\mathbf{1}_{t \geq T} \left(1 - C|x - y|e^{T\bar{g'}}\right) \mathbb{P}(B = 0 \mid T)\right] \\
\geq & \mathbb{E}\left[\mathbf{1}_{t \geq T} \left(1 - C|x - y|e^{T\bar{g'}}\right) \left(1 - \frac{\bar{r'}}{\underline{r}}|x - y|e^{T\bar{g'}}\right)\right].
\end{aligned}$$

Finally, as we can upper bounded T by an exponential variable E with parameter $\underline{r}$, we conclude that

$$\begin{aligned} d_{TV}(\delta_x P_t, \delta_y P_t) &\leq \mathbb{P}(X_t \neq Y_t) \\ &\leq \mathbb{E}\left[\mathbf{1}_{t<T} + \mathbf{1}_{t\geq T}|x-y|e^{T\bar{g'}}\left(C + \frac{\bar{r'}}{\underline{r}}\right)\right] \\ &\leq \mathbb{P}(t < E) + |x-y|\left(C + \frac{\bar{r'}}{\underline{r}}\right)\mathbb{E}\left[\mathbf{1}_{t\geq T}e^{T\bar{g'}}\right] \\ &\leq e^{-t\underline{r}} + |x-y|\left(C + \frac{\bar{r'}}{\underline{r}}\right)\frac{\underline{r}}{\underline{r}-\bar{g'}}. \end{aligned}$$

□

We have a similar result when the curvature is null:

Corollary 2.3.14 (Total variation convergence for PDMP with weaker assumptions)**.** *Assume that ii) and iii) hold, $\rho = 0$ and there exist $[a,b] \subset E$ and $\varepsilon > 0$ such that for all $x \notin [a,b]$*

$$\left(-g'(x) + r(x)\partial_x \int_E (x-y)K(x,dy)) + r'(x)\int_E (y-x)K(y,dx)\right) \geq \varepsilon.$$

Then there exist $M > 0$ and $\theta > 0$ such that

$$d_{TV}(\mu P_t, \nu P_t) \leq M(1 + \mathcal{W}(\mu,\nu))e^{-\theta t}.$$

For any starting distribution μ, ν.

Proof. It is a direct application of Theorem 2.3.5. □

Remark 2.3.15 (Total variation decay of diffusion processes with jumps)**.** *We can prove a similar result if the process have a diffusive part. Nevertheless, if X diffuses then the convergence will be faster.*

Remark 2.3.16 (General Markov kernel)**.** *The main assumption of this theorem is a contraction of the kernel in total variation (point iii)). This assumption is natural but it does not capture others examples such that the TCP process [BCG+11].*

2.4 Examples and applications

In this section we develop several examples. In Subsection 2.4.1, our main results are applied to some models with biological applications. In the subsection that follows, we give several remarks for

the example of the TCP process. We finish by studying the long time behaviour of the integral of a Lévy process driven by a Brownian motion.

2.4.1 Stochastic models for population dynamics

2.4.1.1 Feller diffusion with multiplicative jumps: the Bansaye-Tran process

Let us consider the process studied in [BT11]. It belongs to $E = \mathbb{R}_+$ and evolves according to a Feller diffusion; namely

$$\forall t \geq 0,\ dX_t = \mathbf{g} X_t dt + \sqrt{2\mathbf{s} X_t} dB_t,$$

where $\mathbf{g}$ and $\mathbf{s}$ are two positive numbers. When it jumps from x, this new state is Hx, where the random variable H is distributed according to a probability measure $\mathcal{H}$ on $[0,1]$. We assume that $H \overset{d}{=} 1 - H$. This process models the rate of parasite in a cell population. The number of parasite grows in each cell and the cells divide. These two phenomena do not unwind in the same time scale. The parasites born and die faster than the cells divide. Thus, the rate of parasite is modelled by a Feller diffusion. This one can be understood as the limit of birth and death process. The jumps model the cell division. In this setting, we have

Corollary 2.4.1 (Exponentially decreasing to 0 when r is decreasing)**.** *If r is decreasing and*

$$\rho = \mathbb{E}[H]\left(\inf_{x\geq 0} r(x) - x\bar{r'}\right) - \mathbf{g} > 0,$$

where $\bar{r'} = \sup_{x\geq 0} r'(x)$, then for any $t \geq 0$,

$$\mathbb{E}[|X_t|] \leq e^{-t\rho}\mathbb{E}[|X_0|].$$

Proof. Using the invariance of δ_0 and Theorem 2.1.1, we have

$$\mathcal{W}(\delta_x P_t, \delta_0) \leq e^{-\rho t}\mathcal{W}(\delta_x, \delta_0)$$

Furthermore, we have $\mathcal{W}(\delta_x P_t, \delta_0) = P_t Id(x) = \mathbb{E}[X_t \mid X_0 = x]$. □

Remark 2.4.2 (On the open set assumption)**.** *Notice that this process does not satisfy exactly our main assumptions but it is easy to see the conclusion is also true.*

In [BT11], it is proved that if r is monotone then X convergences, almost surely, to zero. They do not give an explicit bound for the convergence. Our corollary gives a new bound for the L^1-convergence. To compare, [BT11, Proposition 3.1] says

Theorem 2.4.3 (Extinction criterion when r is constant)**.** *We have the following duality.*
(i) If $\mathbf{g} \leq -r\int_0^1 \log(h)\mathcal{H}(dh)$*, then* $\mathbb{P}(\exists t > 0,\ Y_t = 0) = 1$.
Moreover if $\mathbf{g} < -r\int_0^1 \log(h)\mathcal{H}(dh)$*,*

$$\exists \alpha > 0,\ \forall x_0 \geq 0,\ \exists c > 0, \forall t \geq 0, \quad \mathbb{P}(X_t > 0 \mid X_0 = x_0) \leq ce^{-\alpha t}. \tag{2.13}$$

(ii) If $\mathbf{g} > -r\int_0^1 \log(h)\mathcal{H}(dh)$*, then* $\mathbb{P}(\forall t \geq 0,\ X_t > 0) > 0$.
Furthermore, for every $0 \leq \alpha < \mathbf{g} + r\int \log(h)\mathcal{H}(dh)$*,*

$$\mathbb{P}\left(\lim_{t\to+\infty} e^{-\alpha t}Y_t = \infty\right) = \{\forall t,\ X_t > 0\} \quad \textit{a.s.}$$

The point (2.13) can be written as

$$d_{TV}\left(\delta_{x_0}P_t, \delta_0\right) \leq ce^{-\alpha t}.$$

And the second point (ii) implies that

$$\underline{\lim}_{t\to+\infty} d_{TV}\left(\delta_{x_0}P_t, \delta_0\right) > 0.$$

But when r is constant and the curvature ρ is positive, we have $\mathbf{g} < -r\int_0^1 \log(h)\mathcal{H}(dh)$. Thus X converges almost surely to 0, and

$$\mathbb{E}[|X_t|] \leq e^{-\rho t}\mathbb{E}[|X_0|].$$

More precisely, a rapidly calculation gives

$$\mathbb{E}[X_t] = e^{-\rho t}\mathbb{E}[X_0].$$

So, if

$$\int_0^1 -\log(h)\mathcal{H}(dh) > \frac{\mathbf{g}}{r} > \kappa = 1 - \int_0^1 h\mathcal{H}(dh),$$

then

$$\lim_{t\to+\infty} X_t = 0 \text{ a.s. but } \lim_{t\to+\infty} \mathbb{E}[X_t] = +\infty.$$

2.4.1.2 Rate of convergence for branching measure-valued processes

Let us consider a model of structured population. We observe a Markov process indexed by a supercritical continuous time Galton Watson tree. Along the branches of the tree, the process evolves as a diffusion. The branching event is nonlocal; namely the positions of the offspring are described

by a random vector $(F_{j,K}(x,\Theta))_{j\leq K}$. They depend on the position x of the mother just before the branching event and on the number K of offspring. The randomness of these positions is modelled via a uniform variable Θ on $[0,1]$. This process can be described with the following empirical measure:

$$\mathbf{Z}_t = \sum_{u\in V_t} \delta_{X_t^u},$$

where X_t^u belongs to the branch u at time t and V_t is the set of branches at time t. It was proved in [BDMT11] and [Clo11] that, for every continuous and bounded function f,

$$\frac{1}{\mathbb{E}[\mathbf{Z}_t(E)]}\mathbb{E}\left[\int_E f(x)\mathbf{Z}_t(ds)\right] = \frac{1}{\mathbb{E}[\text{card}(V_t)]}\mathbb{E}\left[\sum_{u\in V_t} f(X_t^u)\right] = \mathbb{E}\left[f(Y_t)\right],$$

where Y is generated by (2.1), with biased parameter. With this formula, we can deduce the long time behavior and the contraction properties of the mean measure. A similar formula holds when r is not constant [Clo11]. It will be interesting to capture the speed of convergence to $\mathbf{Z}$ instead of $\mathbb{E}[\mathbf{Z}]$. A first approach is given with [Clo11, Theorem 1.2].

2.4.2 TCP window size process

2.4.2.1 The continuous time process

Now, we consider a process which represents the TCP congestion. This Markov process is generated, for any $f \in C_c^2(\mathbb{R}_+^*$ and $x > 0$, by

$$\mathcal{L}f(x) = f'(x) + r(x)\int_0^1 (f(hx) - f(x))\mathcal{H}(dh). \tag{2.14}$$

Firstly, our main result gives:

Corollary 2.4.4 (Wasserstein curvature)**.** *Assume that r is non increasing and C^1. If*

$$\rho = \int_0^1 (1-h)\mathcal{H}(dh)\inf_{x\geq 0}\left(r(x) - xr'(x)\right) > 0, \tag{2.15}$$

then the semigroup $(P_t)_{t\geq 0}$ generated by $\mathcal{L}$ admits a unique probability invariant π and

$$\forall t \geq 0,\ \mathcal{W}(\mu P_t, \pi) \leq e^{-\rho t}\mathcal{W}(\mu P_t, \pi),$$

for every probability measure μ.

By [CMP10], there is also an exponential convergence when $r(x) = x + a$ with $\rho = \frac{a}{2}$. A recent work [BCG+11] prove that, nevertheless the curvature is null when $r(x) = x$, this process converges exponentially to its invariant distribution in Wasserstein and total variation distance. If r is non-increasing and $\rho = 0$, Theorem 2.3.5 gives an exponential convergence in Wasserstein distance.

Remark 2.4.5 (Poincaré or log-Sobolev inequality and the Bakry-Emery criterion)**.** *Our approach is based on a commutation formula as well as the Bakry-Emery calculus [BÉ85]. Nevertheless, in general, our processes do not satisfy a Poincaré or a log-Sobolev inequality. That is, for the first one,*

$$\lambda Var_\pi f \leq \int \Gamma f d\pi,$$

for a sufficiently large class of function f, where $\Gamma f = \frac{1}{2}\mathcal{L}(f^2) - f\mathcal{L}f$ and $\lambda > 0$. Indeed, in the case of the TCP window size, we have

$$\Gamma f(x) = r(x) \int_0^1 (f(hx) - f(x))^2 \mathcal{H}(dh).$$

And if $\mathcal{H} = \delta_{1/2}$, we easily construct a lot of functions f such that $\Gamma f = 0$ (see [LP09]). Thus, we have an example where the process has a positive Wasserstein curvature and which does not satisfy a Poincaré inequality.

2.4.2.2 The embedded chain

Let $(\hat{X}_n)_{n\geq 0}$ be the embedded chain of the TCP process. That is defined by

$$\hat{X}_n = X_{T_n} \text{ where } T_n = \inf\{t > T_{n-1} \mid X_{t+} \neq X_{t-}\} \text{ for } n \geq 1 \text{ and } T_0 = 0. \tag{2.16}$$

This Markov chain is often easier to study than the continuous time process. For instance, if $r(x) = ax^\alpha$, it is easy to see that

$$R(\hat{X}_{T_{n+1}}) = R(H_n(R^{-1}(E_n + R(\hat{X}_{T_n})))) = H_n^{\alpha+1}(E_n + R(\hat{X}_{T_n})),$$

where R is the antiderivative of r and $H_n = X_{T_n}/X_{T_n-}$. This autoregressive relation gives the ergodicity. Furthermore, the limiting random variable $\hat{X}_\infty$ satisfies

$$R(\hat{X}_{T_\infty}) \stackrel{d}{=} H_1^{\alpha+1}(E_1 + R(\hat{X}_\infty)).$$

Now, using [GRZ04, Proposition 5], we deduce that $\hat{X}_\infty$ have a density given by

$$x \mapsto \frac{1}{\prod_{n\geq 1}(1-\mathbf{h}^{(\alpha+1)n})} \sum_{n\geq 0} \prod_{k=1}^{n} \frac{\mathbf{h}^{-(\alpha+1)(n+1)}}{1-\mathbf{h}^{-(\alpha+1)k}} a x^\alpha e^{-\mathbf{h}^{-(\alpha+1)(n+1)} a(\alpha+1)^{-1} x^{(\alpha+1)}}.$$

Now, applying [Dav93, Theorem 34.31], we can deduce the invariant law of the continuous time process. This result generalises [GRZ04], but it is already known via others techniques [vLLO09]. For this Markov chain, we arrive to bound all Wasserstein distance. Recall that for every $p \geq 1$, the $\mathcal{W}^{(p)}$ Wasserstein distance, between two laws μ_1 and μ_2 on E with finite p^{th} moment, is defined by

$$\mathcal{W}^{(p)}(\mu_1,\mu_2) = \inf_{\{X\sim\mu_1, Y\sim\mu_2\}} \left(\mathbb{E}\left[|X-Y|^p\right]\right)^{1/p},$$

where the infimum runs over all coupling of μ_1 and μ_2. We have

Theorem 2.4.6 (Wasserstein exponential ergodicity for the embedded chain)**.** *Assume that $\mathcal{L}(X_0)$ and $\mathcal{L}(Y_0)$ have finite p^{th} moment for some real $p \geq 1$ and r belongs to $C^1(\mathbb{R}_+^*)$ and is increasing. Let $\hat{X}$ and $\hat{Y}$ be the embedded chains of X and Y. Then, for any $n \geq 0$, with a random variable $H \sim \mathcal{H}$,*

$$\mathcal{W}^{(p)}(\mathcal{L}(\hat{X}_n), \mathcal{L}(\hat{Y}_n)) \leq \mathbb{E}(H^p)^{n/p} \mathcal{W}_p(\mathcal{L}(X_0), \mathcal{L}(Y_0)).$$

In particular, if $\hat{\pi}$ is the invariant law of $\hat{X}$ then

$$\mathcal{W}^{(p)}(\mathcal{L}(\hat{X}_n), \hat{\pi}) \leq \mathbb{E}(H^p)^{n/p} \mathcal{W}_p(\mathcal{L}(X_0), \hat{\pi}).$$

This result generalises [CMP10, Theorem 2.1] but the proof is exactly the same and we give it for sake of completeness.

Proof. It is sufficient to provide a good coupling. Let $x \geq 0$ and $y \geq 0$ be two non-negative real numbers, and let $(E_n)_{n\geq 1}$ and $(H_n)_{n\geq 1}$ be two independent sequences of i.i.d. random variables with respective laws the exponential law of unit mean and the law $\mathcal{H}$. Let $\hat{X}$ and $\hat{Y}$ be the discrete time Markov chains on $[0,\infty)$ defined by

$$\begin{aligned} \hat{X}_0 = x \quad &\text{and} \quad \hat{X}_{n+1} = H_{n+1} R^{-1}(R(\hat{X}_n) + E_{n+1}) \quad \text{for any } n \geq 0 \\ \hat{Y}_0 = y \quad &\text{and} \quad \hat{Y}_{n+1} = H_{n+1} R^{-1}(R(\hat{Y}_n) + E_{n+1}) \quad \text{for any } n \geq 0. \end{aligned}$$

The law of $\hat{X}$ (respectively $\hat{Y}$) is the law of the embedded chain of a process generated by L and starting from x (respectively y). Now, let a be a non-negative number, if $\varphi_a : x \mapsto R^{-1}(a + R(x))$

then

$$\varphi_a'(x) = \frac{r(x)}{r\left(R^{-1}(a+R(x))\right)} \leq 1 \Rightarrow |\varphi_a(x) - \varphi_a(y)| \leq |x-y|. \tag{2.17}$$

And we get

$$\begin{aligned}\forall p \geq 1,\ \mathbb{E}[|\hat{X}_{n+1} - \hat{Y}_{n+1}|^p] &= \mathbb{E}[H_{n+1}^p |\varphi_{E_{n+1}}(\hat{X}_n) - \varphi_{E_{n+1}}(\hat{Y}_n)|^p] \\ &\leq \mathbb{E}[H_{n+1}^p |\hat{X}_n - \hat{Y}_n|^p] = \mathbb{E}[H_{n+1}^p]\mathbb{E}[|\hat{X}_n - \hat{Y}_n|^p].\end{aligned}$$

A straightforward recurrence leads to

$$\mathbb{E}[|\hat{X}_n - \hat{Y}_n|^p] \leq \mathbb{E}[H_1^p]^n |x-y|^p.$$

It gives the desired inequality when the initial laws are Dirac masses. The general case follows by integrating this inequality with respect to couplings of the initial laws. □

This theorem gives a bound of the coarse Ricci curvature [Oll10], which is the discrete time equivalent of the Wasserstein curvature. We can compare the curvature of this Markov chain and its continuous time equivalent.

2.4.3 Integral of Lévy processes with respect to a Brownian motion

Let us consider a fragmentation process i.e. $\int_0^1 f(x+y)d\eta = \int f(hx)\mathcal{H}(dh)$, for every continuous and bounded function, where $\mathcal{H}$ is a probability measure on $[0,1]$. The evolution of X can be described in term of stochastic differential equation. Let $(B_s)_{s\geq 0}$ be a standard Brownian motion, and let $Q(ds,du,d\theta)$ be a Poisson point process on $\mathbb{R}_+ \times \mathbb{R}_+ \times [0,1]$, of intensity $ds\ du\ d\theta$, independent from the Brownian motion. Here $ds, du, d\theta$ are the Lebesgue measures on $\mathbb{R}_+, \mathbb{R}_+$ and $[0,1]$. We have,

$$\begin{aligned}X_t = X_0 &+ \int_0^t g(X_s)ds + \int_0^t \sigma(X_s)dB_s \\ &- \int_0^t \int_{E\ \times[0,1]} \mathbf{1}_{\{u\leq r\}}\theta X_{s-} Q(ds,du,d\theta).\end{aligned}$$

The jump times $(T_k)_{k\geq 1}$ of X are distributed following a Poisson process $(\mathcal{N}_t)_{t\geq 0}$. Between these times, the process evolves like a diffusion. At these times, we have $X_{T_j} = H_j X_{T_j-}$, where (H_j) is a i.i.d. sequence of law $\mathcal{H}$. we assume that $H_1 \in (0,1)$ almost surely. If you take the logarithm of X then the multiplicative jumps become additive jumps. Then, as the jump times are Poissonian, we can obtain a continuous process by renormalising our process with a Lévy process. Formally, let $(L_t)_{t\geq 0}$

be the Lévy process defined by

$$\forall t \geq 0,\ L_t = -\int_{\mathbb{R}_+ \times [0,1]} \mathbf{1}_{\{u \leq r\}} \ln(h) Q(ds, du, dh) = -\ln\left(\prod_{j=1}^{\mathcal{N}_t} H_j\right),$$

and let $(\bar{X}_t)_{t\geq 0}$ be the continuous process defined by $\bar{X}_t = X_t e^{L_t}$. We have

Lemma 2.4.7 (Stochastic differential equation for $\bar{X}$ and X)**.** *For any $t \geq 0$,*

$$\bar{X}_t = X_0 + \int_0^t e^{L_s} g(X_s) ds + \int_0^t e^{L_s} \sigma(X_s) dB_s$$

and,

$$X_t = X_0 e^{-L_t} + \int_0^t e^{-L_s} g(X_{t-s}) ds + \int_0^t e^{-L_s} \sigma(X_{t-s}) dB_s.$$

Proof. All the stochastic integrals that we write are well defined as local martingales. Using Itô's formula with jumps [IW89, Theorem 5.1, p.67] (see also [JS03, Theorem 4.57, p.57]) and since $\bar{X}$ is continuous, we get

$$\begin{aligned}
\bar{X}_t &= X_0 + \int_0^t e^{L_s} \left(g(X_s) ds + \sigma(X_s) dB_s\right) \\
&\quad + \int_0^t X_{s-} e^{L_{s-}} - X_s e^{L_s} \mathbf{1}_{\{u \leq r\}} Q(ds, du, dh) \\
&= X_0 + \int_0^t e^{L_s} \left(g(X_s) ds + \sigma(X_s) dB_s\right).
\end{aligned}$$

Then, we deduce,

$$\begin{aligned}
X_t &= X_0 e^{-L_t} + \int_0^t e^{L_s - L_t} \left(g(X_s) ds + \sigma(X_s) dB_s\right) \\
&= X_0 e^{-L_t} + \int_0^t e^{-L_{t-s}} \left(g(X_s) ds + \sigma(X_s) dB_s\right) \\
&= X_0 e^{-L_t} + \int_0^t e^{-L_s} g(X_{t-s}) ds + \int_0^t e^{-L_s} \sigma(X_{t-s}) dB_s.
\end{aligned}$$

□

This lemma is a generalisation of the relation of [BT11, Lemma 3.2] and [LvL08, Section 6]. In [BT11], it is a preliminary for the proof of Theorem 2.4.3. In [LvL08], they deduce that when g is constant, and $\sigma = 0$, we have

$$\lim_{t \to +\infty} X_t = g \int_0^{+\infty} e^{-L_s} ds.$$

The behaviour of the right hand side was studied in [BY05] and [CPY01] for general Lévy processes.

We give another application. Let Y be defined, for all $t \geq 0$, by

$$Y_t = \int_0^t e^{-L_s} dB_s,$$

where $L_t = \sum_{k=1}^{N_t} \ln(H_j)$ is independent from the Brownian motion B, and $(H_j)_{j\geq 0}$ are i.i.d. and distributed according to $\mathcal{H}$. Let us define for all $n \in \mathbb{N}$, $\kappa_n = 1 - \int_0^1 h\mathcal{H}(dh)$, we have

Theorem 2.4.8 (Long time behaviour of integral of compound Poisson process with respect to an independent Brownian motion)**.** *The process $(Y_t)_{t\geq 0}$ converges in law to a measure π satisfying*

$$\int x^n \pi(dx) = \frac{n!}{r^{n/2} \prod_{k=1}^{n/2} \kappa_{2k}}$$

for all $n \in 2\mathbb{Z}$, and

$$\int x^n \pi(dx) = \frac{n!}{r^{(n+1)/2} \prod_{k=1}^{(n+1)/2} \kappa_{2k-1}}$$

otherwise. Furthermore all its moments converges and for all $t \geq 0$,

$$\mathcal{W}\left(\mathcal{L}(Y_t), \pi\right) \leq e^{-\kappa_1 t} \mathcal{W}(\delta_0, \pi).$$

Proof. By the previous lemma, we know that Y is generated by (2.1), where, $g = 0, \sigma = 1$, r is constant and $Y_0 = 0$. This process is positively curved thus it admits a unique invariant probability measure and converges exponentially to it. Furthermore, applying the generator on the functions $\alpha_n : x \mapsto x^n$ gives the moments of π and we can use the Carleman criterion to prove that Y converges also to a measure with this moment.

□

Appendix: properties of the Wasserstein curvature

Here, we present briefly some definitions and properties about Lipschitz contraction of general Markov semigroups. Let (E, d) be a Polish space and $(P_t)_{t\geq 0}$ be any Markov semigroup. We denote by $(X_t)_{t\geq 0}$ the Markov process with semigroup $(P_t)_{t\geq 0}$.

Definition 2.4.9 (Wasserstein curvature)**.** *The Wasserstein curvature of $(P_t)_{t\geq 0}$ is the optimal (largest) constant ρ in the following contraction inequality:*

$$\sup_{\substack{x,y\in E \\ x\neq y}} \frac{|P_t g(x) - P_t g(y)|}{d(x,y)} \leq e^{-\rho t}, \tag{2.18}$$

for all $g \in Lip_1(d)$ *and* $t \geq 0$.

It is actually equivalent to be optimal in the following expression:

$$\mathcal{W}(\mu_1 P_t, \mu_2 P_t) \leq e^{-\rho t}\, \mathcal{W}(\mu_1, \mu_2),$$

for any probability measures μ_1, μ_2 and $t \geq 0$. Here $\mu_1 P_t$ stands for the law of $(X_t)_{t\geq 0}$ starting from a random variable distributed according to μ_1. That is,

$$\mu_1 P_t f(x) = \int_E P_t f(x) \mu_1(dx).$$

And $\mathcal{W}$ is the Wasserstein distance associated to d. This notion of curvature was introduced by Joulin [Jou07] and Ollivier [Oll10]. It is naturally connected to the notion of Ricci curvature on Riemannian manifolds [vRS05], the Dobrushin's uniqueness criterion, the Chen exponent [Che96] or the Wasserstein spectral gap [HSV11]. A first simple result is its calculation by random scaling time:

Lemma 2.4.10 (Markov processes indexed by a subordinator)**.** *Let* $(X_t)_{t\geq 0}$ *be a Markov process with curvature* ρ *and let* $(\tau(t))_{t\geq 0}$ *be a subordinator independent of* X. *If* $(Q_t)_{t\geq 0}$ *is the semigroup be defined, for every continuous and bounded function* f, $t \geq 0$ *and* $x \in E$ *by*

$$Q_t f(x) = \mathbb{E}[f(X_{\tau(t)}) \mid X_0 = x],$$

then its Wasserstein curvature ρ_Q *satisfies*

$$\rho_Q = b\rho + \int_0^\infty (1 - e^{-\rho z})\nu(dz) = \psi(\rho),$$

where ψ *is the Laplace exponent of* τ, b *its drift term and* ν *its Lévy measure.*

The proof is straightforward. A more important result is

Theorem 2.4.11 (Exponential convergence)**.** *If* $\rho > 0$ *then there exists a unique invariant probability measure* π. *furthermore*

$$\forall t \geq 0,\ \mathcal{W}(\mu P_t, \pi) \leq e^{-\rho t}\mathcal{W}(\mu, \pi),$$

for every probability measure μ.

This result is a direct consequence of [Che04, Theorem 5.23] but we will give another proof (with an additional assumption). In general, $(X_t)_{t\geq 0}$ is not ergodic when $\rho \leq 0$ as can be easily checked with Brownian motion.

Proof. Let us assume the existence of $x_0 \in E$ such that the mapping $f_d : x \mapsto d(x, x_0)$ is in the domain of the generator of P. Then we have

$$\begin{aligned} P_t f_d(x_0) &= \mathcal{W}(\delta_{x_0} P_t, \delta_{x_0}) \\ &\leq \sum_{k=1}^{n} \mathcal{W}(\delta_{x_0} P_{kt/n}, \delta_{x_0} P_{(k-1)t/n}) \\ &\leq \sum_{k=1}^{n} e^{-\rho t(k-1)/n} P_{t/n} f_d(x_0). \end{aligned}$$

Taking the limit $n \to +\infty$, we obtain that

$$P_t f_d(x_0) \leq \rho^{-1}(1 - e^{-\rho t}) \mathcal{L} f_d(x_0).$$

We deduce that if $f \in \text{Lip}_1$ then $P_t f$ is bounded by a constant C which does not depend to f. Let μ be a probability measure, for every $s, t \geq 0$, we have

$$\mathcal{W}(\mu P_{t+s}, \mu P_t) \leq C e^{-\rho t}.$$

Finally, the sequence $(\mu P_t)_{t \geq 0}$ is a Cauchy sequence, in a complete space [Che04], thus it converges to a probability measure π which is trivially the unique invariant measure. □

Note that $(P_t)_{t \geq 0}$ can be geometrically ergodic even if its curvature is not positive:

Lemma 2.4.12 (Exponential decay when $\rho = 0$)**.** *If $\rho = 0$ and there exists $t_0 > 0$ such that*

$$K_{t_0} = \sup_{x_0, y_0 \in E} \frac{\mathcal{W}(\delta_{x_0} P_{t_0}, \delta_{y_0} P_{t_0})}{d(x_0, y_0)} < 1,$$

then there exists $\kappa > 0$ such that, for all $x_0, y_0 \in E$, we have

$$\forall t \geq t_0, \mathcal{W}(\delta_{x_0} P_t, \delta_{y_0} P_t) \leq e^{-t\kappa} |x_0 - y_0|.$$

Furthermore we can choose

$$\kappa = -\frac{\ln(K_{t_0})}{t_0}.$$

Lemma 2.4.12 implies the existence of an invariant distribution and the convergence to it. Theorem 2.3.5 and its corollaries give some applications of this lemma. The conditions of Lemma 2.4.12 are not always satisfied. For instance, if $(L_t)_{t \geq 0}$ is a Lévy process on $\mathbb{R}$, then its semigroup $(P_t)_{t \geq 0}$

satisfies

$$\forall t \geq 0,\ \mathcal{W}(\delta_x P_t, \delta_y P_t) = |x - y|.$$

Now, let ω and $\bar{\omega}$ be defined, for every $x \neq y$ and $t \geq 0$, by

$$\omega(t, x, y) = \frac{\mathcal{W}(\delta_x P_t, \delta_y P_t)}{d(x, y)},$$

and

$$\bar{\omega}(t) = \sup_{x,y \in E} \omega(t, x, y).$$

The proof of the previous lemma is based on the fact that $\ln(\bar{\omega})$ is sub-additive.

Proof of Lemma 2.4.12. First, we have

$$\mathcal{W}(\mu P_t, \nu P_t) \leq \bar{\omega}(t)\mathcal{W}(\mu, \nu).$$

Then, for any $x, y \in E$ and $t \geq 0$, the Markov property gives

$$\begin{aligned}\mathcal{W}(\delta_x P_{t+s}, \delta_y P_{t+s}) &= \mathcal{W}((\delta_x P_t)P_s, (\delta_y P_t)P_s) \\ &\leq \bar{\omega}(t)\mathcal{W}(\delta_x P_s, \delta_y P_s).\end{aligned}$$

We deduce that

$$\omega(t + s, x, y) \leq \bar{\omega}(t)\omega(s, x, y) \ \Rightarrow\ \bar{\omega}(t + s) \leq \bar{\omega}(t)\bar{\omega}(s).$$

Now, the curvature is non-negative, thus

$$\forall t > 0,\ \bar{\omega}(t) \leq 1.$$

So $\bar{\omega}$ is decreasing. But, as there exists $t_0 > 0$ such that $\bar{\omega}(t_0) < 1$, we have

$$\forall t \geq t_0,\ \bar{\omega}(t) < 1.$$

Finally, for all $t \geq t_0$, there exists $n \in \mathbb{N}$ such that $t \geq nt_0$, and then

$$\begin{aligned}\mathcal{W}(\delta_x P_t, \delta_y P_t) &\leq \bar{\omega}(t)\, \mathcal{W}(\delta_x, \delta_y) \leq \bar{\omega}(nt_0)\, \mathcal{W}(\delta_x, \delta_y) \\ &\leq \bar{\omega}(t_0)^n\, \mathcal{W}(\delta_x, \delta_y) \leq \exp\left(t\, \frac{\ln(\bar{\omega}(t_0))}{t_0}\right)\, \mathcal{W}(\delta_x, \delta_y).\end{aligned}$$

□

The Wasserstein curvature is a local characteristic. It depends to the behavior on all the space and at any time. It corresponds to the worst possible decay. This notion of curvature is connected with the notion of L^2-spectral gap:

Theorem 2.4.13 (Wasserstein contraction implies L^2-spectral gap for reversible semigroup)**.** *Let $(P_t)_{t\geq 0}$ be the semigroup of a Markov process with invariant distribution π. If the following assumptions hold*

- *the Wasserstein curvature ρ is positive;*
- *the semigroup is reversible;*
- $\lim_{t\to 0} \|P_t f - f\|_{L^2(\pi)} = 0$ *for all* $f \in L^2(\pi)$*;*
- $Lip_1 \cap L^\infty(\pi) \cap L^2(\pi)$ *is dense in* $L^2(\pi)$*;*

then

$$Var_\pi(P_t f) \leq e^{-2\rho t} Var_\pi(f),$$

where $Var_\pi(f) = \int_E (f - \int_E f d\pi)^2 d\pi = \|f - \int f d\pi\|_{L^2(\pi)}$.

The proof relies crucially on reversibility. Nevertheless, for most of our examples, this condition is never satisfied (while in contrast all one dimensional diffusions are reversible). This theorem is just the continuous time adaptation of [HSV11, Proposition 2.8] and is close to [Che04, Theorem 5.23], [Wan03, Theorem 2.1 (2)], [Vey12, Theorem 2], and [Oll10, Proposition 30].

Proof. Let $f \in \mathrm{Lip}_1$ be a non-negative and bounded function such that $\int_E f d\pi = 1$. Using the reversibility and the invariance of π, we have,

$$\begin{aligned}\mathrm{Var}_\pi(P_t f) &= \int_E f P_{2t} f d\pi - \left(\int_E f d\pi\right)^2 \\ &\leq \mathcal{W}\left(P_{2t} f d\pi, \pi\right).\end{aligned}$$

The measure $P_{2t}f\,d\pi$ satisfies, for every continuous and bounded function φ, the following expression

$$\begin{aligned}\int_E \varphi P_{2t} f\, d\pi &= \int_E \int_E \varphi(x) f(y) P_{2t}(x, dy)\pi(dx) \\ &= \int_E \int_E \varphi(x) f(y) P_{2t}(y, dx)\pi(dy).\end{aligned}$$

Thus,

$$\begin{aligned}\mathrm{Var}_\pi(P_t f) &\leq \mathcal{W}\left((f\pi)P_{2t}, \pi\right) \\ &\leq C_f e^{-2\rho t}, \end{aligned} \tag{2.19}$$

for some constant $C_f > 0$ which depend to f. By translation and dilatation, the last inequality holds for all bounded function f which belongs to Lip_1. Now, let $f \in L^2(\pi)$ be a Lipschitz and bounded function such that,

$$\int_E f(x)\pi(dx) = 0 \ \text{ and } \ \int_E f(x)^2\pi(dx) = 1.$$

Applying spectral Theorem, Jensen inequality and (2.19), we find

$$\begin{aligned}\mathrm{Var}_\pi(P_t f) = \int P_t f^2 d\pi &= \int_E \int_0^\infty e^{-\lambda t} dE_\lambda(f) d\pi \\ &\leq \left(\int_E \int_0^\infty e^{-\lambda(t+s)} dE_\lambda(f) d\pi\right)^{\frac{t}{t+s}} \\ &\leq C_f^{\frac{t}{t+s}} e^{-2\rho t}.\end{aligned}$$

Taking the limit $s \to +\infty$, we conclude the proof. □

This result can not be generalised in the non reversible case as can be viewed in the remark 2.4.5.

Remark 2.4.14 (Another approach)**.** *We can give an alternative proof, using Inequality* (2.19) *and [CGZ10, Lemma 2.12]. This lemma is based on the convexity of the mapping* $t \mapsto \ln \|P_t f\|_{L^2(\pi)}$ *which is also a consequence of the reversibility.*

Chapter 3

Exponential ergodicity for Markov processes with random switching *

3.1 Introduction

Markov processes with switching are intensively used for modelling purposes in applied subjects like biology [CMMS12, CDMR11, FGM10], storage modelling [BKKP05], neuronal activity [PTW12, GT11]. This class of Markov processes is reminiscent of the so-called iterated random functions [DF99] or branching processes in random environment [Smi68] in the discrete time setting. Several recent works [BH12, BGM10, BLMZ12b, BLMZ12c, CD08, dSY05, GIY04, GG96] deal with their long time behaviour (existence of an invariant probability measure, Harris recurrence, exponential ergodicity, hypoellipticity...). In particular, in [BH12, BLMZ12b], the authors provide a kind of hypoellipticity criterion with Hörmander-like bracket conditions. Under these conditions, they deduce the uniqueness and absolute continuity of the invariant measure, provided that a suitable tightness condition is satisfied. They also obtain geometric convergence in the total variation distance. Nevertheless, there are many simple processes with switching which do not verify any hypoellipticity condition. To illustrate this fact, let us consider the simple example of [BLMZ12c]. Let (X, I) be the Markov process on $\mathbb{R}^2 \times \{-1, 1\}$ generated by

$$Af(x, i) = -(x - (i, 0)) \cdot \nabla_x f(x, i) + (f(x, -i) - f(x, i)) . \tag{3.1}$$

∗. In collaboration with Martin Hairer

This process is ergodic and the first marginal π of its invariant measure is supported on $\mathbb{R}\times\{0\}$. Thus, in general, it does not converge in the total variation distance. However, it is proved in [BLMZ12c] that it converges in a certain Wasserstein distance. Let us recall that the p^{th} Wasserstein distance $\mathcal{W}^{(p)}$, with $p \geq 1$, on a Polish space (E, d) is defined by

$$\mathcal{W}_d^{(p)}(\mu_1, \mu_2) = \inf_{X_1, X_2} \mathbb{E}\left[d(X_1, X_2)^p\right]^{1/p},$$

for every probability measure μ_1, μ_2 on E, where the infimum is taken over all pairs of random variables X_1, X_2 with respective laws μ_1, μ_2. When $p = 1$, we set $\mathcal{W}_d = \mathcal{W}_d^{(1)}$. The Kantorovich-Rubinstein duality [Vil09, Theorem 5.10] shows that one also has

$$\mathcal{W}_d(\mu_1, \mu_2) = \sup_{f \in \mathrm{Lip}_1} \int_E f d\mu_1 - \int_E f d\mu_2,$$

where $f\colon E \mapsto \mathbb{R}$ is in Lip_1 if and only if it is a 1-Lipschitz function, namely

$$\forall x, y \in E, \qquad |f(x) - f(y)| \leq d(x, y).$$

The total variation distance d_{TV} can be viewed as the Wasserstein distance associated to the trivial distance function, namely

$$d_{\mathrm{TV}}(\mu_1, \mu_2) = \inf_{X_1, X_2} \mathbb{P}(X_1 \neq X_2) = \frac{1}{2} \sup_{\|f\|_\infty \leq 1} \int_E f d\mu_1 - \int_E f d\mu_2,$$

where the infimum is again taken over all random variables X_1, X_2 with respective distributions μ_1, μ_2. In the present article, we will give convergence criteria for a general class of switching Markov processes. These processes are built from the following ingredients:

- a Polish space (E, d) and a finite set F;
- a family $(Z^{(n)})_{n\in F}$ of E-valued strong Markov processes represented by their semigroups $(P^{(n)})_{n\in F}$, or equivalently by their generators $(\mathcal{L}^{(n)})_{n\in F}$ with domains $(\mathcal{D}^{(n)})_{n\in F}$;
- a family $(a(\cdot, i, j))_{i,j\in F}$ of non-negative functions on E.

We are interested by the process $(\mathbf{X}_t)_{t\geq 0} = (X_t, I_t)_{t\geq 0}$, defined on $\mathbf{E} = E \times F$, which jumps between these dynamics. Roughly speaking, X_t behaves like $Z_t^{(I_t)}$ as long as I does not jump. The process I is discrete and jumps at a rate given by a. More precisely, the dynamics of $(\mathbf{X}_t)_{t\geq 0}$ is as follows:

- Given a starting point $(x, i) \in E\times F$, we take for $Z^{(i)}$ an instance as above with initial condition $Z_0^{(i)} = x$. The initial conditions for $Z^{(j)}$ with $j \neq i$ are irrelevant.

– The discrete component I is constant and equal to i until the time $T = \min_{j\in F} T_j$, where $(T_j)_{j\geq 0}$ is a family of random variables that are conditionally independent given $Z^{(i)}$ and that verify
$$\forall j \in F,\ \mathbb{P}\left(T_j > t \mid \mathcal{F}_t\right) = \exp\left(-\int_0^t a\left(Z_s^{(i)}, i, j\right) ds\right),$$
where $\mathcal{F}_t = \sigma\{Z_s^{(i)} \mid s \leq t\}$.
– For all $t \in [0, T)$, we then set $X_t = Z_t^{(i)}$ and $I_t = i$.
– At time T, there exists a unique $j \in F$ such that $T = T_j$ and we set $I_T = j$ and $X_T = X_{T-}$.
– We take (X_T, I_T) as a new starting point at time T.

Let us make a few remarks about this construction. First, this algorithm guarantees the existence of our process under the condition that there is no explosion in the switching rate. In other words, our construction is global as long as I only switches value finitely many time in any finite time interval. Assumption 3.1.1 below will be sufficient to guarantee this non-explosion. Also note that, in general, X and I are not Markov processes by themselves, contrary to $\mathbf{X}$. Nevertheless, we have that I is a Markov process if a does not depend on its first component. The construction given above shows that, provided that there is no explosion, the infinitesimal generator of $\mathbf{X}$ is given by

$$\mathbf{L}f(x,i) = \mathcal{L}^{(i)} f(x,i) + \sum_{j\in F} a(x,i,j)\left(f(x,j) - f(x,i)\right), \tag{3.2}$$

for any bounded function f such that $f(\cdot, i)$ belongs to $\mathcal{D}^{(i)}$ for every $i \in F$. We will denote by $(\mathbf{P}_t)_{t\geq 0}$ the semigroup of $\mathbf{X}$. To guarantee the existence of our process, we will consider the following natural assumption:

Assumption 3.1.1 (Regularity of the jumps rates)**.** *The following boundedness condition is verified:*

$$\bar{a} = \sup_{x\in E} \sup_{i\in F} \sum_{j\in F} a(x,i,j) < +\infty,$$

and the following Lipschitz condition is also verified:

$$\sup_{i\in F} \sum_{j\in F} |a(x,i,j) - a(x,i,j)| \leq \kappa d(x,y),$$

for some $\kappa > 0$.

We will also assume the following hypothesis to guarantee the recurrence of I:

Assumption 3.1.2 (Recurrence assumption)**.** *The matrix* $(\underline{a}(i,j))_{i,j\in F}$ *defined by*

$$\underline{a}(i,j) = \inf_{x\in E} a(x,i,j),$$

yields the transition rates of an irreducible and positive recurrent Markov chain.

With these two assumptions, we are able to get exponential stability in two situations. The first situation is one where each underlying dynamics does on average yield a contraction in some Wasserstein distance, but no regularising assumption is made. The second situation is the opposite, where we replace the contraction by a suitable regularising property.

3.1.1 Two criteria without hypoellipticity assumption

In this section, we assume that we have some information on the Lipschitz contraction (or expansion) of our underlying processes:

Assumption 3.1.3 (Lipschitz contraction)**.** *For each* $i \in F$*, there exists* $\rho(i) \in \mathbb{R}$ *such that*

$$\forall t \geq 0, \mathcal{W}_d\left(\mu P_t^{(i)}, \nu P_t^{(i)}\right) \leq e^{-\rho(i)t}\mathcal{W}_d\left(\mu,\nu\right), \tag{3.3}$$

for any two probability measures μ, ν*. Furthermore there exist* $x_0 \in E$ *and* $t_{x_0} > 0$ *such that if* $V_{x_0} : x \mapsto d(x,x_0)$ *then*

$$\sup_{t\in[0,t_{x_0}]} P_t V_{x_0}(x_0) < +\infty.$$

In the previous assumption, given a semigroup $(P_t)_{t\geq 0}$, we used the notation μP_t to denote the measure defined by

$$(\mu P_t) f = \int P_t f d\mu.$$

if $\mu = \delta_x$, for some x, then in this work, we also use the notation $\delta_x P_t(dy) = P_t(x,dy)$.

To verify equation (3.3) is not much of a restriction because we do not assume that $\rho(i) > 0$. The best constant in this inequality is called the Wasserstein curvature in [Jou07, Jou09] and the coarse Ricci curvature in [Oll09, Oll10], since it is heavily related to the geometry of the underlying space as illustrated in [**?** , Theorem 2]. If $\rho(i) > 0$, then we can deduce some properties like geometric ergodicity, a Poincaré inequality or some concentration inequalities [Clo12, Jou07, Jou09, HSV11, Oll10]. A trivial bound on $\rho(i)$ is given in the special case of diffusion processes in Section 3.4.1.

The bound (3.3) is quite stringent since, if $\rho(i) > 0$, it implies that there is some Wasserstein contraction for *every* $t > 0$ and not just for sufficiently long times. This is essentially equivalent to

the existence of a Markovian coupling between two instances X_t and Y_t of the Markov process with generator $\mathcal{L}^{(i)}$ such that $\mathbb{E}d(X_t, Y_t) \leq e^{-\rho t} d(X_0, Y_0)$.

In principle, this condition could be slightly relaxed by the addition of a proportionality constant C_i, provided that one assumes that the switching rate of the process is sufficiently slow. This ensures that, most of the time, it spends a sufficiently long time in any one state for this proportionality constant not to play a large role.

One could also imagine allowing for jumps of the component in E at the switching times, and this would lead to a similar difficulty.

In the same way, the distance d appearing in Assumption 3.1.3 is the *same* for every i and that it does not allow for a constant prefactor in the right hand side of (3.3). This may seem like a very strong assumption since usual convergence theorems, like Harris' theorem, do not give this kind of bound. We will see however in Section 3.5 an example which illustrates that there is no obvious way in general to weaken this condition. The intuitive reason why this is so is that if the process switches rapidly, then it is crucial to have some local information (small times) and not only global information (large times) on the behaviour of each underlying dynamics.

We now have presented all the assumptions that are necessary to state our main results. The first one describes the simplest situation, that is when a does not depend on its first component:

Theorem 3.1.4 (Wasserstein exponential ergodicity in the constant case)**.** *Under assumptions 3.1.1, 3.1.2 and 3.1.3, if $a(x, i, j)$ does not depend on x and the Markov process I has an invariant probability measure ν verifying*

$$\sum_{i \in F} \nu(i)\rho(i) > 0,$$

then there exist a probability measure $\boldsymbol{\pi}$, some constants $C, \lambda > 0$ and $q \in (0, 1]$ such that

$$\forall t \geq 0,\ \mathcal{W}_{\mathbf{d}}\left(\delta_{\mathbf{y}_0}\mathbf{P}_t, \boldsymbol{\pi}\right) \leq Ce^{-\lambda t}\left(1 + \sum_{i \in F} \int_E d(y_0, x)^q \boldsymbol{\pi}(dx, i)\right),$$

for every $\mathbf{y}_0 = (y_0, j_0) \in \mathbf{E}$, where the distance $\mathbf{d}$, on $\mathbf{E}$, is defined by

$$\mathbf{d}(\mathbf{x}, \mathbf{y}) = \mathbf{1}_{i \neq j} + \mathbf{1}_{i=j}(1 \wedge d^q(x, y)), \tag{3.4}$$

for every $\mathbf{x} = (x, i)$, $\mathbf{y} = (y, j)$ belonging to $\mathbf{E}$.

This statement is not surprising: it states that if the process contracts in mean, then it converges exponentially to an invariant distribution. The conditions are rather sharp as will be illustrated in Section 3.5. In particular, we recover [BLMZ12c, Theorem 1.10] and this (slight) generalisation

could be deduced from the argument given there. Using Hölder's inequality, we can also deduce convergence in the p^{th} Wasserstein distance $\mathcal{W}_{\mathbf{d}}^{(p)}$ with $p \geq 1$ provided that $\mathbf{X}$ satisfies a moment condition. We give the previous theorem and its proof for sake of completeness and for a better understanding of the more complicated case, where a is allowed to depend on its first argument. That is

Theorem 3.1.5 (Wasserstein exponential ergodicity with an on-off type criterion)**.** *Let us suppose that Assumptions 3.1.1, 3.1.2, and 3.1.3 hold. We set*

$$F_0 = \{i \in F \mid \rho(i) > 0\} \text{ and } F_1 = \{i \in F \mid \rho(i) \leq 0\},$$

$$\rho_0 = \min_{i \in F_0} \rho(i) > 0 \text{ and } \rho_1 = \min_{i \in F_1} \rho(i) \leq 0,$$

$$a_0 = \max_{i \in F_0} \sup_{x \in E} \sum_{j \in F_1} a(x,i,j) \text{ and } a_1 = \min_{i \in F_1} \inf_{x \in E} \sum_{j \in F_0} a(x,i,j).$$

If

$$\rho_0 a_1 + \rho_1 a_0 > 0,$$

then there exist a probability measure $\boldsymbol{\pi}$, some constants $C, \lambda > 0$ and $q \in (0,1]$ such that

$$\forall t \geq 0,\ \mathcal{W}_{\mathbf{d}}\left(\delta_{\mathbf{y}_0}\mathbf{P}_t, \boldsymbol{\pi}\right) \leq Ce^{-\lambda t}\left(1 + \sum_{i \in F} \int_E d(y_0,x)^q \boldsymbol{\pi}(dx,i)\right),$$

for every $\mathbf{y}_0 = (y_0, j_0) \in \mathbf{E}$, where the distance $\mathbf{d}$, on $\mathbf{E}$, is defined by

$$\mathbf{d}(\mathbf{x},\mathbf{y}) = \mathbf{1}_{i \neq j} + \mathbf{1}_{i=j}(1 \wedge d^q(x,y)),$$

for every $\mathbf{x} = (x,i)$, $\mathbf{y} = (y,j)$ belonging to $\mathbf{E}$.

With this result, we not only recover [BLMZ12c, Theorem 1.15], but we extend it significantly. In our case, the underlying dynamics are not necessarily deterministic and do not need to be strictly contracting in a Wasserstein distance. One drawback is that the constants λ and C are much less explicit. This theorem is a direct consequence of the more general Theorem 3.3.2 below. These two theorems are our main result and, contrary to the previous theorem, it seems that they cannot be deduced directly from the approach of [BLMZ12c].

3.1.2 Two criteria with hypoellipticity assumption

In the previous subsection, we have supposed that some of the underlying dynamics contract at sufficiently high rate in a Wasserstein distance. This is of course not a necessary condition for geometric ergodicity in general. Using some arguments of the proof of Theorem 3.1.4 and Theorem 3.1.5, we can deduce a different criterion which uses instead a Lyapunov-type argument to prove that $\mathbf{X}$ converges. We begin by stating an assumption similar to Assumption 3.1.3:

Assumption 3.1.6 (Existence of a Lyapunov function)**.** *There exist $K \geq 0$, a function $V \geq 0$, and for every $i \in F$ there exists $\lambda(i) \in \mathbb{R}$ such that*

$$\forall t \geq 0, \forall x \in E,\ P_t^{(i)} V(x) \leq e^{-\lambda(i)t} V(x) + K. \tag{3.5}$$

Note again that we have not supposed that $\lambda(i) > 0$. One way to prove this kind of bound is to use the classical drift condition on the generator (see (3.7) below). With this assumption we are able to prove

Theorem 3.1.7 (Exponential ergodicity in the constant case)**.** *Suppose that assumptions 3.1.1, 3.1.2 and 3.1.6 hold, that $a(x, i, j)$ does not depend on x and that I has an invariant probability measure ν verifying*

$$\sum_{i \in F} \nu(i)\lambda(i) > 0\ .$$

If there exists $i_0 \in F$ and $t_0 \geq 0$ such that the sublevel sets $\{x \in E \mid V(x) \leq K\}$ are small for $P_t^{(i_0)}$ for every $K > 0$ and $t \geq t_0$, then there exist a probability measure $\boldsymbol{\pi}$ and two constants $C, \lambda > 0$ such that

$$\forall t \geq 0,\ d_{\mathrm{TV}}\left(\delta_{\mathbf{x}} \mathbf{P}_t, \boldsymbol{\pi}\right) \leq C e^{-\lambda t}(1 + V(x)),$$

for every $\mathbf{x} = (x, i) \in \mathbf{E}$.

The definition of a small set is recalled in Definition 3.2.9. We give also the analogous of Theorem 3.1.5:

Theorem 3.1.8 (Exponential ergodicity with an on-off type criterion)**.** *Let us suppose that Assumptions 3.1.1, 3.1.2, 3.1.3 hold. We set*

$$F_0 = \{i \in F \mid \lambda(i) > 0\} \ \text{ and } \ F_1 = \{i \in F \mid \lambda(i) \leq 0\},$$

$$\lambda_0 = \min_{i \in F_0} \lambda(i) > 0 \ \text{ and } \ \lambda_1 = \min_{i \in F_1} \lambda(i) \leq 0,$$

$$a_0 = \max_{i\in F_0} \sup_{x\in E} \sum_{j\in F_1} a(x,i,j) \;\; \textit{and} \;\; a_1 = \min_{i\in F_1} \inf_{x\in E} \sum_{j\in F_0} a(x,i,j).$$

If

$$\lambda_0 a_1 + \lambda_1 a_0 > 0,$$

and there exists $i_0 \in F$ *and* $t_0 \geq 0$ *such that the sublevel sets* $\{x \in E \mid V(x) \leq K\}$ *are small for* $P_t^{(i_0)}$ *for every* $K > 0$ *and* $t \geq t_0$*, then there exist a probability measure* $\boldsymbol{\pi}$ *and two constants* $C, \lambda > 0$ *such that*

$$\forall t \geq 0, \; d_{\mathrm{TV}}\left(\delta_{\mathbf{x}}\mathbf{P}_t, \boldsymbol{\pi}\right) \leq Ce^{-\lambda t}(1 + V(x)),$$

for every $\mathbf{x} = (x,i) \in \mathbf{E}$.

Note that in general it is not necessary to assume that sublevel sets of V are small for any single one of the underlying dynamics. For example, using the results of [BH12, BLMZ12b], Section 3.4.2 gives results analogous to the two previous theorems, in the special case of piecewise deterministic Markov processes where the only small sets for the underlying dynamics consist of single points.

The remainder of the paper is organised as follows. The proofs of our four main theorems are split over two sections: Section 3.2 deals with the proof of Theorem 3.1.4 and Theorem 3.1.7. In Section 3.3, we begin by giving a more general assumption in the non-constant case than our on-off criterion. Then, we introduce a weak form of Harris' Theorem that we will use to prove Theorem 3.1.5. The proof of this theorem is then decomposed in such a way to verify each point of the weak Harris' Theorem. Section 3.4.1 gives sufficient conditions to verify our main assumption in the special case of diffusion processes. The section which follows deals with the special case of switching dynamical system. We conclude with Section 3.5, where we give some very simple examples illustrating the sharpness of our conditions.

3.2 Constant jump rates

In this section, we begin by proving that under Assumptions 3.1.3 or 3.1.6, the process $\mathbf{X}$ cannot wander off to infinity, i.e. its semigroup possesses a Lyapunov function. We then prove Theorems 3.1.4 and 3.1.7 using a similar argument to [BLMZ12c] for the first one and Harris' Theorem for the second one.

3.2.1 Construction of a Lyapunov function

We begin by recalling the definition of a Lyapunov function

Definition 3.2.1 (Lyapunov function)**.** *A Lyapunov function for a Markov semigroup $(P_t)_{t\geq 0}$ over a Polish space (X, d_X) is a function $V : X \mapsto [0, \infty]$ such that V is integrable with respect to $P_t(x, \cdot)$ for every $x \in X$ and $t > 0$ and such that there exist constants $C_V, \gamma, K_V > 0$ verifying*

$$P_t V(x) = \int_X V(y) P_t(x, dy) \leq C_V e^{-\gamma t} V(x) + K_V, \tag{3.6}$$

for every $x \in X$ and $t \geq 0$.

A well know sufficient condition for finding a Lyapunov function is the following drift condition:

$$\mathcal{L}V \leq -\gamma V + C, \tag{3.7}$$

where $\mathcal{L}$ is the generator of the semigroup $(P_t)_{t\geq 0}$. The condition (3.7) implies a bound like (3.5) and is clearly stronger than (3.6). In general, our switching Markov process $\mathbf{X}$ may not verify the drift condition (3.7) but, in Lemmas 3.2.7 and 3.3.8, we give a sharp condition under which it verifies (3.6). In this section, we first prove that a Wasserstein contraction as in Assumption 3.1.3 implies the existence of a Lyapunov-type function as in Assumption 3.1.6. Then, we will prove that Assumption 3.1.6 implies the existence of a Lyapunov function for $\mathbf{X}$.

Lemma 3.2.2 (Wasserstein contraction implies the existence of a Lyapunov-type function)**.** *Let $(P_t)_{t\geq 0}$ be the semigroup of a Markov process, on a Polish space (X, d_X), such that there exists $\lambda \in \mathbb{R}^*$ verifying*

$$\mathcal{W}_{d_X}(\delta_x P_t, \delta_y P_t) \leq e^{-\lambda t} d_X(x, y), \tag{3.8}$$

for every $x, y \in X$ and $t \geq 0$. If there exist $x_0 \in X$ and $t_{x_0} > 0$ such that the function $V_{x_0} : x \mapsto d(x, x_0)$ verifies

$$\sup_{t\in[0,t_{x_0}]} P_t V_{x_0}(x_0) < +\infty, \tag{3.9}$$

then there exist $C_1, C_2 > 0$ such that

$$P_t V_{x_0}(x) \leq e^{-\lambda t}(V_{x_0}(x) + C_1) + C_2, \tag{3.10}$$

for every $x \in X$ and $t \geq 0$.

Proof. Firstly note that Inequality (3.8) is equivalent to

$$\forall t \geq 0,\ \mathcal{W}_{d_X}(\mu P_t, \nu P_t) \leq e^{-\lambda t} \mathcal{W}_{d_X}(\mu, \nu),$$

for every probability measure μ and ν.

For any $t \geq t_{x_0}$ and $n \geq 0$, it follows that

$$\begin{aligned} P_t V_{x_0}(x_0) = \mathcal{W}_{d_X}(\delta_{x_0} P_t, \delta_{x_0}) &\leq \sum_{k=0}^{n-1} \mathcal{W}_{d_X}(\delta_{x_0} P_{(k+1)\frac{t}{n}}, \delta_{x_0} P_{k\frac{t}{n}}) \\ &\leq \sum_{k=0}^{n-1} \mathcal{W}_{d_X}((\delta_{x_0} P_{\frac{t}{n}}) P_{k\frac{t}{n}}, \delta_{x_0} P_{k\frac{t}{n}}) \leq \frac{e^{-\lambda t} - 1}{e^{-\lambda t/n} - 1} P_{t/n} V_{x_0}(x_0). \end{aligned}$$

Taking $n = \lfloor t/t_{x_0} \rfloor + 1$, where $\lfloor t/t_{x_0} \rfloor$ is the integer part of t/t_{x_0}, we conclude that

$$P_t V_{x_0}(x_0) \leq (e^{-\lambda t} + 1) C', \qquad C' = \sup_{u \in \left[t_{x_0}/2, t_{x_0}\right]} \frac{P_u V_{x_0}(x_0)}{|e^{-\lambda u} - 1|},$$

which is finite by (3.9). Finally, for every $x \in X$ and $t \geq 0$, we have

$$\begin{aligned} P_t V_{x_0}(x) = \mathcal{W}_{d_X}(\delta_x P_t, \delta_{x_0}) &\leq \mathcal{W}_{d_X}(\delta_x P_t, \delta_{x_0} P_t) + \mathcal{W}_{d_X}(\delta_{x_0} P_t, \delta_{x_0}) \\ &\leq e^{-\lambda t} V_{x_0}(x) + (e^{-\lambda t} + 1) C' \,, \end{aligned}$$

thus concluding the proof. □

We deduce that Assumption 3.1.3 implies Assumption 3.1.6 with $V = V_{x_0}$ and $\lambda = \rho$.

Remark 3.2.3. *The point of this lemma is to also allow for negative values of λ. When $\lambda > 0$, then it is immediate that P_t admits a unique invariant measure and exhibits geometric ergodicity.*

Remark 3.2.4. *If V_{x_0} is in the domain of the generator $\mathcal{L}$ of $(P_t)_{t \geq 0}$ then we have*

$$\forall t \geq 0, \; P_t V_{x_0}(x_0) \leq \frac{e^{-\lambda t} - 1}{e^{-\lambda t/n} - 1} P_{t/n} V_{x_0}(x_0),$$

for some $n \geq 1$. Now, taking the limit $n \to +\infty$, we deduce the following bound:

$$\mathcal{W}_{d_X}(\delta_{x_0} P_t, \delta_{x_0}) \leq \frac{e^{-\lambda t} - 1}{-\lambda} \mathcal{L} V(x_0).$$

Finally, for every $x \in X$, we have

$$\begin{aligned} P_t V(x) = \mathcal{W}_{d_X}(\delta_x P_t, \delta_{x_0}) &\leq \mathcal{W}_{d_X}(\delta_x P_t, \delta_{x_0} P_t) + \mathcal{W}_{d_X}(\delta_{x_0} P_t, \delta_{x_0}) \\ &\leq e^{-\lambda t} V(x) + \frac{e^{-\lambda t} - 1}{-\lambda} \mathcal{L} V(x_0). \end{aligned}$$

However, V_{x_0} does not belong to the domain of the generator in general, as can be seen already in the example of simple Brownian motion.

Remark 3.2.5 (The special case $\lambda = 0$)**.** *In the previous lemma, we have supposed that $\lambda \neq 0$, and this assumption is necessary for our conclusion to hold. Indeed, if $(B_t)_{t\geq 0}$ is a Brownian motion then*

$$\lim_{t\to+\infty} \mathbb{E}\left[|B_t|\right] = +\infty,$$

and inequality (3.10) does not hold. Indeed note that if $\lambda = 0$ in the previous lemma then we have

$$P_t V_{x_0}(x) \leq V_{x_0}(x) + Kt,$$

for some constant $K > 0$ and every $x \in E$ and $t \geq 0$. In what follows, we shall say that, by the previous lemma, Assumption 3.1.3 implies Assumption 3.1.6 with $\rho(i) = \lambda(i)$, for every $i \in F$. But it is false when $\rho(i) = 0$, however if $\rho(i) = 0$ then we have that

$$\mathcal{W}_d(\mu P_t, \nu P_t) \leq e^{-\widetilde{\rho}(i)t}\mathcal{W}_d(\mu, \nu),$$

for every probability measure μ, ν and $t \geq 0$, for any $\widetilde{\rho}(i) < 0$ as closed to zero as we want, and so we can choose it in order to preserve the positivity of the means in our main theorems.

We now show that if Assumption 3.1.6 holds and the mean of $(\lambda(i))_{i\in F}$ is positive, then $\mathbf{X}$ admits a Lyapunov function. As in [BLMZ12c], this result comes from the following lemma:

Lemma 3.2.6. *Let (K_t) be a continuous time Markov chain on a finite set S, and assume that it is irreducible and positive recurrent with invariant measure ν_K. If $\alpha\colon S \to \mathbb{R}$ is a function verifying*

$$\sum_{n\in S} \nu_K(n)\alpha(n) > 0,$$

then there exist $C, c, \eta > 0$ and $p \in (0, 1]$ such that

$$ce^{-\eta t} \leq \mathbb{E}\left[e^{-\int_0^t p\alpha(K_s)ds}\right] \leq Ce^{-\eta t},$$

for any initial condition K_0 and every $t \geq 0$.

Proof. It is a consequence of Perron-Frobenius Theorem and the study of eigenvalues. See [BGM10, Proposition 4.1] and [BGM10, Proposition 4.2] for further details. □

Now we are able to prove that $\mathbf{P}$ possesses a Lyapunov function in the case where the switching rates do not depend on the location of the process.

Lemma 3.2.7. *Under Assumption 3.1.1, 3.1.2 and 3.1.6, if $a(x,i,j)$ does not depend on x and I has an invariant measure ν satisfying*

$$\sum_{i\in F}\lambda(i)\nu(i) > 0,$$

then there exist $C_V, K_V, \lambda_V > 0$ and $q \in (0,1]$ such that

$$\forall t \geq 0, \forall x \in E,\ \mathbf{P}_t V^q(x,i) \leq C_V e^{-\lambda_V t} V^q(x) + K_V.$$

In the previous lemma, we used a slight abuse of notation. Indeed, if f is a function defined on E, we also denote by f the mapping $(x,i) \mapsto f(x)$ on $\mathbf{E}$.

Proof. First, Jensen's inequality gives this weaker form of (3.5):

$$P_t^{(i)}(V^q)(x) \leq e^{-q\lambda(i)t}V^q(x) + K^q,$$

for every $q \in (0,1]$. Now, for all $t \geq 0$ and $(x,i) \in \mathbf{E}$, a straightforward recurrence gives

$$\begin{aligned}\mathbf{P}_t V^q(x,i) &= \mathbb{E}\left[P_{t-T_{N_t}}^{(I_{T_{N_t}})} \circ P_{T_{N_t}-T_{N_t-1}}^{(I_{T_{N_t-1}})} \circ \cdots \circ P_{T_1-T_0}^{(I_0)}(V^q)(x)\right] \\ &\leq \mathbb{E}\left[e^{-\int_0^t q\lambda(I_s)ds}\right] V^q(x) + K^q \sum_{n\geq 0} \mathbb{E}\left[e^{-q\int_0^{T_n}\lambda(I_s)ds}\right],\end{aligned}$$

where $(T_k)_{k\geq 0}$ is the sequence of jump times of I, with $T_0 = 0$, and N_t the number of jumps before t. By Lemma 3.2.6, there exist $C > 0, \eta > 0$ and $q \in (0,1]$ such that

$$\mathbb{E}\left[e^{-\int_0^t q\lambda(I_s)ds}\right] \leq Ce^{-\eta t}.$$

Furthermore, one can show that T_n is of order n and that

$$K_V = K^q \sum_{n\geq 0} \mathbb{E}\left[e^{-q\int_0^{T_n}\lambda(I_s)ds}\right] \lesssim K^q \sum_{n\geq 0} e^{-\varepsilon n} < +\infty,$$

for some $\varepsilon > 0$. We do not detail this argument now, but we will prove it in the slightly more difficult context of non-constant rate a in Lemma 3.3.8. This concludes the proof. □

Remark 3.2.8 (On the assumption that F is finite)**.** *It is natural to extend our results to the case where F is countably infinite. Obviously, we then have to add the assumption that I is positive recurrent, but this is not enough. Indeed, if for each $i \in F$, $C_1(i)$ and $C_2(i)$ denote the constants C_1, C_2, appearing*

in Lemma 3.2.2 applied on $Z^{(i)}$, then we should furthermore assume that

$$\sup_{i\in F}(C_1(i)+C_2(i))<+\infty\ ,$$

for the argument to go through.

3.2.2 Proof of Theorem 3.1.4

This section is split into two parts. We begin by introducing our coupling construction, and we then proceed to prove Theorem 3.1.4. In both parts, we make the standing assumption that the hypotheses of Theorem 3.1.4 hold. In particular, I is an ergodic Markov chain.

3.2.2.1 Our coupling

Let $\mathbf{x}=(x,i)$ and $\mathbf{y}=(y,j)$ be two points of $\mathbf{E}$, we will build a coupling $(\mathbf{X},\mathbf{Y})$, starting from $(\mathbf{x},\mathbf{y})$, such that each component is an instance of the Markov process generated by $\mathbf{L}$, and $\mathbf{d}(\mathbf{X}_t,\mathbf{Y}_t)$ is small.
Remark that it is well-known that if I and J are two independent processes with transition rate a then there exist $C_c,\theta_c>0$ such that

$$\forall t\geq 0,\ \mathbb{P}\left(T_c>t\right)\leq C_c e^{-\theta_c t}, \tag{3.11}$$

for any initial conditions I_0,J_0, where $T_c=\inf\{t\geq 0\mid I_t=J_t\}$ is their first meeting time. From now on, we fix the starting points of our coupling $\mathbf{x}=(x,i)$, $\mathbf{y}=(y,j)$ and the time $t\geq 0$. The processes $(\mathbf{X}_t)_{t\geq 0}=(X_t,I_t)_{t\geq 0}$ and $(\mathbf{Y}_t)_{t\geq 0}=(Y_t,J_t)_{t\geq 0}$ are then coupled as follow:

- if $i\neq j$ then we consider that $\mathbf{X}$ and $\mathbf{Y}$ evolve independently of each other until the first meeting time T_c.
- for all $s\geq T_c$, we set $I_s=J_s$ and we couple X and Y in such a way that

$$\forall k\geq 0,\ \mathbb{E}\left[d(X_{S_k},Y_{S_k})\mid\mathcal{F}_{S_{k-1}}\right]\leq e^{-\rho(I_{S_{k-1}})(S_k-S_{k-1})}d(X_{S_{k-1}},Y_{S_{k-1}}),$$

  where $(T_k)_{k\geq 0}$ is the sequence of jumps times of I, $S_k=T_k\wedge t$ and $(\mathcal{F}_s)_{s\geq 0}$ is the natural filtration associated to $(\mathbf{X},\mathbf{Y})$.

Note that if $i=j$ then $T_c=0$. Note also that we can build our coupling since the Assumption 3.1.3 is satisfied and that there exists an optimal map in the definition of the Wasserstein distance.

3.2.2.2 Estimation of the distance

Proof of Theorem 3.1.4. If $i = j$, then by Jensen's inequality and iteration, we have similarly to before

$$\mathbb{E}\left[d(X_t, Y_t)^q\right] \leq \mathbb{E}\left[e^{-q\int_0^t \rho(I_s)ds}\right] d(x,y)^q,$$

where $q \in (0,1]$. By Lemma 3.2.6, there exist $C, \eta > 0$ and $q \in (0,1]$ such that

$$\mathbb{E}\left[d(X_t, Y_t)^q\right] \leq Ce^{-\eta t} d(x,y)^q.$$

Now, for general i and j, we have

$$\begin{aligned}\mathbb{E}\left[\mathbf{d}(\mathbf{X}_t, \mathbf{Y}_t)\right] \leq\ & \mathbb{E}\left[\sqrt{\mathbf{1}_{T_c \geq t/2}\left(1 + V^q(X_t) + V^q(Y_t)\right)}\right] \\ & + \mathbb{E}\left[\sqrt{\mathbf{1}_{T_c \leq t/2}\, d(X_t, Y_t)^q (1 + V^q(X_t) + V^q(Y_t))}\right],\end{aligned}$$

where $V(x) = d(x, x_0)$. Now, Cauchy-Schwarz inequality, Equation (3.11), Lemma 3.2.2 and Lemma 3.2.7 give

$$\begin{aligned}\mathbb{E}\left[\sqrt{\mathbf{1}_{T_c \geq t/2}(1 + V^q(X_t) + V^q(Y_t))}\right] &\leq \mathbb{P}\left(T_c \geq t/2\right)^{1/2} \mathbb{E}\left[1 + V^q(X_t) + V^q(Y_t)\right]^{1/2} \\ &\leq C_c e^{-\theta_c t/4}\left(1 + C_V e^{-\lambda_V t}(V^q(x) + V^q(y)) + 2K_V\right)^{1/2}.\end{aligned}$$

In the other hand, one has the bound

$$\begin{aligned}&\mathbb{E}\left[\sqrt{\mathbf{1}_{T_c \leq t/2} d(X_t, Y_t)^q (1 + V^q(X_t) + V^q(Y_t))}\right] \\ &\qquad \leq \mathbb{E}\left[\mathbf{1}_{T_c \leq t/2} d(X_t, Y_t)^q\right]^{1/2} \mathbb{E}\left[1 + V^q(X_t) + V^q(Y_t)\right]^{1/2}. \end{aligned} \tag{3.12}$$

As a consequence of Lemmas 3.2.2 and 3.2.7, we also have the bound

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}_{T_c \leq t/2} d(X_t, Y_t)^q\right]^{1/2} &\leq Ce^{-\eta t/2} \mathbb{E}\left[d(X_{T_c}, Y_{T_c})^q \mathbf{1}_{T_c \leq t/2}\right]^{1/2} \\ &\leq Ce^{-\eta t/2} \mathbb{E}\left[\left(V(X_{T_c})^q + V(Y_{T_c})^q\right)\mathbf{1}_{T_c \leq t/2}\right]^{1/2} \\ &\leq Ce^{-\eta t/2}\left[C_V V^q(x_0) + C_V V^q(y_0) + 2K_V\right]^{1/2}.\end{aligned}$$

Assembling these inequalities and using again Lemma 3.2.7 to bound the second factor in (3.12), we find the existence of $C > 0$ and $\lambda > 0$ verifying

$$\mathbb{E}[\mathbf{d}(\mathbf{X}_t, \mathbf{Y}_t)] \leq Ce^{-\lambda t}(1 + V(x) + V(y)),$$

for every $t \geq 0$ and $x, y \in E$ such that $X_0 = x$ and $Y_0 = y$. From the previous inequality, we deduce that

$$\mathcal{W}_{\mathbf{d}}\left(\boldsymbol{\mu}\mathbf{P}_t, \boldsymbol{\nu}\mathbf{P}_t\right) \leq Ce^{-\lambda t}\left(1 + \sum_{i\in F}\int_E (V(x)\boldsymbol{\nu}(dx, i) + V(x)\boldsymbol{\mu}(dx, i))\right), \tag{3.13}$$

for every probability measure $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$. Now, mimicking the proof of [HM11, Corollary 4.10], we can prove the existence of an invariant measure. More precisely, let us fix a probability measure $\boldsymbol{\mu}$, using the previous inequality, we can prove that $(\boldsymbol{\mu}\mathbf{P}_n)_{n\geq 0}$ is a measure-valued Cauchy sequence (with respect to the $\mathcal{W}_{\mathbf{d}}$ distance). We deduce that it converges to a measure $\boldsymbol{\mu}_\infty$ verifying

$$\boldsymbol{\mu}_\infty \mathbf{P}_1 = \boldsymbol{\mu}_\infty.$$

Finally we can see that $\boldsymbol{\pi} = \int_0^1 \boldsymbol{\mu}_\infty \mathbf{P}_u du$ is invariant. The inequality of Theorem 3.1.4 comes from equation (3.13). Indeed, with the notation of the statement of Theorem 3.1.4, this equation gives

$$\mathcal{W}_{\mathbf{d}}\left(\delta_{\mathbf{y}_0}\mathbf{P}_t, \boldsymbol{\pi}\right) \leq Ce^{-\alpha t}\left(1 + \sum_{i\in F}\int_E d(x, x_0)\boldsymbol{\pi}(dx, i) + \int_E d(x_0, y_0)\right).$$

Integrating with respect to $\boldsymbol{\pi}(dx_0)$ and changing the constants end the proof.

□

3.2.3 Proof of Theorem 3.1.7

We divide again the proof in two parts. First, we recall some tools on Harris' Theorem. Second, we give the proof of Theorem 3.1.7.

3.2.3.1 Classical Harris' Theorem

Here, we recall a version of Harris' Theorem (also called Foster, Lyapunov, Meyn-Tweedie, Doeblin) that is suitable for our needs. This theorem yields exponential convergence to stationarity for a process which does not "escape to infinity" and verifies furthermore a Doeblin-type condition. More precisely, we use the following notion of a small set:

Definition 3.2.9. *A set $A \subset X$ is small for the semigroup $(P_t)_{t\geq 0}$ over a Polish space (X, d_X), if there exists a time $t > 0$ and a constant $\varepsilon > 0$ such that*

$$d_{\mathrm{TV}}(\delta_x P_t, \delta_y P_t) \leq 1 - \varepsilon$$

for every $x, y \in A$.

The classical Harris theorem [HM11, MT93] then states that

Theorem 3.2.10 (Harris)**.** *Let $(P_t)_{t\geq 0}$ be a Markov semigroup over a Polish space (X, d_X) such that there exists a Lyapunov function V with the additional property that the sublevel sets $\{x \in X \mid V(x) \leq C\}$ are small for every $C > 0$. Then $(P_t)_{t\geq 0}$ has a unique invariant measure π and*

$$d_{\mathrm{TV}}(\delta_x P_t, \pi) \leq C e^{-\gamma_* t}(1 + V(x)).$$

for some positive constants C and γ_*.

Note that one does not really need that *all* sublevel sets are small and one can have a slightly stronger conclusion by using a total variation distance weighted by V [HM11, Theorem 1.3].

Proof of Theorem 3.1.7. By Lemma 3.2.7, $\mathbf{P}$ admits V as Lyapunov function so, by Harris' Theorem, it only remains to show that $\{V \leq C\}$ is small for $\mathbf{P}$, for every $C > 0$. Since V is a Lyapunov function, there exists $t_*^{(1)} > 0$ and $K > K_V$ (with K_V as in Lemma 3.2.7) such that

$$\forall t \geq t_*^{(1)}, \quad \mathbb{E}\left[V(X_t)\right] \leq K,$$

uniformly over all $x \in E$ such that $V(x) \leq C$. Therefore, if $\mathbf{X}$ is a process generated by $\mathbf{L}$, it follows from Markov's inequality that

$$\mathbb{P}\left(V(X_t) \leq 2K\right) \geq \frac{1}{2}\,,$$

uniformly over $t \geq t_*^{(1)}$.

Let now $i_0 \in F$ be as in the statement. Since $A = \{V \leq 2K\}$ is small for $P^{(i_0)}$, we obtain some $t_0 > 0$ and $\varepsilon > 0$, such that for all $x, y \in A$ there exists a coupling $(Z_t^{i_0,x}, Z_t^{i_0,y})$ verifying

$$\mathbb{P}\left(Z_t^{i_0,x} = Z_t^{i_0,y}\right) \geq \varepsilon\,, \qquad t \geq t_0\,, \tag{3.14}$$

and $Z_t^{i_0,x}$, $Z_t^{i_0,y}$ have respective law $\delta_x P_t^{(i_0)}$, $\delta_y P_t^{(i_0)}$.

By the irreducibility of the process I, one can find $t_* > t_*^{(1)}$ and $\delta > 0$ such that $\mathbb{P}(I_s = i_0, \forall s \in [t_*, t_* + t_0]) > \delta$, uniformly over the starting distributions. Let now $(\mathbf{X}_t, \mathbf{Y}_t)$ be the following coupling:

- the Markov chains I and J are independent over $t \in [0, t_* + t_0]$;
- the processes X and Y are independent over $t \in [0, t_*]$;
- conditionally on the set

$$B = \{V(X_{t_*}) \leq 2K, V(Y_{t_*}) \leq 2K, I_s = J_s = i_0, \forall s \in [t_*, t_* + t_0]\},$$

  the processes X and Y are coupled in such a way to verify (3.14), over $t \in [t_*, t_* + t_0]$;
- conditionally on B^c, they are coupled independently from each other.

The Markov property gives

$$\mathbb{P}(V(X_{t_*}) \leq 2K, I_s = i_0, \forall s \in [t_*, t_* + t_0]) \geq \frac{\delta}{2},$$

and so $\mathbb{P}(B) \geq \delta^2/4$. Combining this inequality with (3.14), we conclude that $\mathbb{P}(\mathbf{X}_{t_*+t_0} = \mathbf{Y}_{t_*+t_0}) \geq \delta^2\varepsilon/4$, uniformly over all initial conditions $\mathbf{x}$ and $\mathbf{y}$ with $V(x) \leq C$ and $V(y) \leq C$, as required. □

3.3 Non-constant jump rates

In all of this section, we now assume that a depends non-trivially on its first component, so that I by itself is not a Markov process anymore. We want to use again Lemma 3.2.6 to show that $\mathbf{X}$ converges, but this time we cannot use it directly on I. The idea is to consider an auxiliary process which does not depend to X and which will bound $(\rho(I_t))_{t\geq 0}$ or $(\lambda(I_t))_{t\geq 0}$. More precisely, we will assume

Assumption 3.3.1 (Birth-death type criterion in the non constant case)**.** *There exist $\bar{n} \in \mathbb{N}$ and a partition $(F_n)_{0\leq n\leq \bar{n}}$ of F such that*

$$\forall n \leq \bar{n}, \forall i \in F_n,\ \forall j \notin F_{n-1} \cup F_n \cup F_{n+1},\ \forall x \in E,\ a(x, i, j) = 0,$$

where we set $F_{-1} = F_{\bar{n}+1} = \emptyset$. Let $(L_t)_{t\geq 0}$ be the continuous time Markov chain with generator

$$Gf(n) = b(n)\left(f(n+1) - f(n)\right) + d(n)\left(f(n-1) - f(n)\right), \tag{3.15}$$

for every $n \leq \bar{n}$*, where*

$$b(n) = \inf_{x\in E} \inf_{i\in F_n} \sum_{j\in F_{n+1}} a(x,i,j) > 0,$$

and

$$d(n) = \sup_{x\in E} \sup_{i\in F_n} \sum_{j\in F_{n-1}} a(x,i,j) > 0,$$

if $n \notin \{0,\bar{n}\}$*,* $d(0) = 0$ *and* $b(\bar{n}) = 0$*. This process is irreducible, non-explosive and positive recurrent with invariant measure* ν*.*

If this assumption holds then, for every $i \in F$, we denote by n_i the only $n \leq \bar{n}$ verifying $i \in F_n$. Let us recall that ν is defined, for every $n \leq \bar{n}$, by

$$\nu(n) = \nu(0) \prod_{k=1}^{n} \frac{b(k-1)}{d(k)} \text{ and } \nu(0) = (1+\Xi)^{-1},$$

where

$$\Xi = \sum_{n=1}^{\bar{n}} \frac{b(0)\dots b(n-1)}{d(1)\dots d(n)}.$$

Now we can state two slight generalisations of Theorem 3.1.5 and 3.1.8. The first one is

Theorem 3.3.2 (Wasserstein exponential ergodicity)**.** *Let us suppose that Assumptions 3.1.1, 3.1.2, 3.1.3 and 3.3.1 hold. If*

$$\sum_{n=0}^{\bar{n}} \nu(n)\alpha(n) > 0,$$

where $(\alpha(n))_{n\geq 0}$ *is an increasing sequence verifying* $\alpha(n) \leq \inf_{i\in F_n} \rho(i)$*, then there exist a probability measure* $\boldsymbol{\pi}$ *and some constants* $C, \lambda, t_0 > 0$ *and* $q \in (0,1]$ *such that*

$$\forall t \geq t_0,\ \mathcal{W}_{\mathbf{d}}\left(\delta_{\mathbf{y}_0}\mathbf{P}_t, \boldsymbol{\pi}\right) \leq Ce^{-\lambda t}\left(1 + \sum_{i\in F}\int_E d(y_0,x)^q \boldsymbol{\pi}(dx,i)\right),$$

for every $\mathbf{y}_0 = (y_0, j_0) \in \mathbf{E}$*, where the distance* $\mathbf{d}$*, on* $\mathbf{E}$*, is defined in* (3.4)*.*

If Assumption 3.3.1 holds with $\bar{n} = 0$ then all contraction parameters are positive and we recover [BLMZ12c, Theorem 1.15]. If it holds with $\bar{n} = 1$, then we have the on-off criterion which was given in introduction. We can also state the analogous result in the setting of Theorem 3.1.8:

Theorem 3.3.3 (Exponential ergodicity)**.** *Let us suppose that Assumptions 3.1.1, 3.1.2, 3.1.3 and 3.3.1 hold and there exists* $i_0 \in F$ *and* $t_0 \geq 0$ *such that the sublevel sets of* V *are small for* $P_t^{(i_0)}$*, for*

every $t \geq t_0$. *If*

$$\sum_{n=0}^{\bar{n}} \nu(n)\alpha(n) > 0,$$

where $(\alpha(n))_{n\geq 0}$ *is an increasing sequence verifying* $\alpha(n) \leq \inf_{i\in F_n} \lambda(i)$, *then there exist a probability measure* $\boldsymbol{\pi}$ *and two constants* $C, \lambda > 0$ *such that*

$$\forall t \geq 0,\ d_{\mathrm{TV}}\left(\delta_{\mathbf{x}}\mathbf{P}_t, \boldsymbol{\pi}\right) \leq Ce^{-\lambda t}(1 + V(x))$$

for every $\mathbf{x} = (x, i) \in \mathbf{E}$.

We do not give the proofs of Theorem 3.1.8 and Theorem 3.3.3, as their proofs are very similar to the proof of Theorem 3.1.7, combined with the argument of Lemma 3.3.8 below. To prove Theorem 3.3.2 however, we cannot use classical Harris' Theorem. Its proof follows the same idea as the proof of Theorem 3.1.4, but there is no direct equivalent to the meeting time. Instead, we use a weak version of Harris' Theorem which yields geometric ergodicity under the existence of a Lyapunov function and a modified "small set" condition. This theorem was previously applied to the stochastic Navier-Stokes equation [HM08], stochastic delay differential equations [HMS11], and linear response theory [HM10]. It is an extension of the classic Harris' Theorem which allows to deal with some degenerate examples like the one given in (3.1).

3.3.1 Weak form of Harris' Theorem

As already mentioned earlier, there are situations in which we cannot expect convergence in total variation. The problem here is that bounded sets may not be small sets. We will therefore replace the notion of small set by the following notion of "closedness" between transition probabilities introduced in [HMS11], which takes into account the topology of the underlying space X.

Definition 3.3.4 (d-small set)**.** *Let* P *be a Markov operator over a Polish space* X *endowed with a distance* $d_X \colon X \times X \mapsto [0, 1]$. *A set* $A \subset X$ *is said to be* d_X*-small if there exists a constant* ε *such that*

$$\mathcal{W}_{d_X}(\delta_x P, \delta_y P) \leq 1 - \varepsilon,$$

for every $x, y \in A$.

This notion is a generalisation of the notion of small set, since small sets are d-small for the trivial distance. This definition can also be extended to situations when d is not a distance [HMS11]. As remarked in that paper, having a Lyapunov function V with d-small sublevel sets cannot be sufficient

to imply the ergodicity of a Markov semigroup. To obtain some convergence result, we further impose that d is contracting for our semigroup:

Definition 3.3.5 (d-contracting operator)**.** *Let P be a Markov operator over a Polish space X endowed with a distance $d_X : X \times X \mapsto [0,1]$. The distance d_X is said to be contracting for P if there exists $\alpha < 1$ such that the bound*

$$\mathcal{W}_{d_X}(\delta_x P, \delta_y P) \leq \alpha d_X(x,y)$$

holds for every $x, y \in X$ verifying $d(x,y) < 1$.

Note that this condition alone is not sufficient to guarantee the convergence of transition probabilities toward a unique invariant measure since we only impose a contraction when $d(x,y) < 1$. In typical situations, "most" pairs (x,y) may satisfy $d(x,y) = 1$, as would be the case for the total variation distance. However, when combined with the existence of a Lyapunov function V that has d-small sublevel sets, it gives geometrical ergodicity [HMS11, Theorem 4.7]:

Theorem 3.3.6 (Weak form of Harris' Theorem)**.** *Let $(P_t)_{t\geq 0}$ be a Markov semigroup over a Polish space X admitting a continuous Lyapunov function V. Assume furthermore that there exist $t^* > t_* > 0$ and a distance $d_X : X \times X \mapsto [0,1]$ which is contracting for P_t and such that the sublevel set $\{x \in X \mid V(x) \leq 4K_V\}$ is d_X-small for P_t, for every $t \in [t_*, t^*]$. Here K_V is as in definition 3.2.1. Then, $(P_t)_{t\geq 0}$ has an invariant probability measure π. Furthermore, defining*

$$\delta_X(x,y) = \sqrt{d_X(x,y)(1 + V(x) + V(y))},$$

there exist $r > 0$ and $t_0 > 0$ such that

$$\forall t \geq t_0,\ \mathcal{W}_{\delta_X}(\mu P_t, \nu P_t) \leq e^{-rt}\mathcal{W}_{\delta_X}(\mu, \nu),$$

for all of probability measures μ, ν on X.

Remark 3.3.7 (On the contracting distances)**.** *The main difficulty when applying the previous theorem is to find a contracting distance. The construction of this distance represents the main part of our paper. In [HM10], there is a general way to build a contracting distance of a Markov operator P over a Banach space $(\mathbb{B}, \|\cdot\|)$, based on a gradient estimate for P and the existence of a super-Lyapunov function. This technique was efficient in [HM10, HM08].*

3.3.2 Construction of a Lyapunov function

As in the constant case, we first show that if each underlying Markov process verifies a weaker form of the drift condition (3.7) then $\mathbf{X}$ possesses a Lyapunov function:

Lemma 3.3.8 (Construction of a Lyapunov function)**.** *Let us suppose that Assumptions 3.1.1, 3.1.2, 3.1.6 and 3.3.1 hold, if*

$$\sum_{n\geq 0} \nu(n)\alpha(n) > 0,$$

where $(\alpha(n))_{n\geq 0}$ *is an increasing sequence verifying* $\alpha(n) \leq \inf_{i\in F_n} \lambda(i)$*, then there exist* $C_V, K_V, \lambda_V > 0$ *and* $q \in (0,1)$ *such that*

$$\forall t \geq 0, \forall (x,i) \in \mathbf{E},\ \mathbf{P}_t V^q(x,i) \leq C_V e^{-\lambda_V t} V^q(x) + K_V.$$

Proof. Recall again that Jensen's inequality gives this weaker form of (3.5):

$$P_t^{(i)}(V^q)(x) \leq e^{-q\alpha(i)t} V^q(x) + K^q,$$

for every $x \in E$ and $q \in (0,1]$. Now, we will describe a construction of $\mathbf{X}$ which will permit to have a better control of the jump mechanism. Let $r \geq 2\bar{a}$ and $(N_t)_{t\geq 0}$ be a Poisson process of intensity r; namely

$$N_t = \sum_{n\geq 0} \mathbf{1}_{\{\tau_n \leq t\}},$$

where $\tau_n = \sum_{k=1}^n E_k$ and $(E_k)_{k\geq 0}$ is a family of i.i.d. exponentially distributed random variables with mean $1/r$. We set $\tau_0 = 0$. At this stage, we do not fix the value of r, but we allow ourselves the freedom to tune it at the end of the proof. We will couple $\mathbf{X} = (X, I)$ with a process L that has generator (3.15). Let us fix $n \in \mathbb{N}$, on $[\tau_n, \tau_{n+1}]$, the process $(\mathbf{X}, L)$ is built as follow:

- conditionally on $\mathbf{X}_{\tau_n}$, $(L_s)_{s\geq 0}$, $(\tau_k)_{k\geq 0}$, the process $(X_s)_{s\in[\tau_n,\tau_{n+1}]}$ moves as $(Z_{t-\tau_n}^{(I_{\tau_n})})_{t\in[\tau_n,\tau_{n+1}]}$ starting from X_{τ_n}; more precisely,

$$\mathbb{E}\left[f(X_t)\mathbf{1}_{t\in[\tau_n,\tau_{n+1}]} \mid \mathcal{G}_n\right] = P_{t-\tau_n}^{(I_{\tau_n})} f(X_{\tau_n}),$$

  where f is a continuous and bounded function and $\mathcal{G}_n = \sigma\{\mathbf{X}_{\tau_n}, (L_s)_{s\geq 0}, (\tau_k)_{k\geq 0}\}$; recall that for every $i \in F$, $(Z_t^{(i)})_{t\geq 0}$ is a Markov process generated by $\mathcal{L}^{(i)}$;
- on $[\tau_n, \tau_{n+1})$, the discrete processes I and L remain constant;
- at time τ_{n+1}, we consider a Bernoulli random variable B with parameter $1/2$ independent of the previous variables and we have two situations:

- if $n_{I_{\tau_n}} \neq L_{\tau_n}$ then
  - if $B = 0$ then I does not jump but L can jump,
  - if $B = 1$ then L does not jump but I can jump;
- if $n_{I_{\tau_n}} = L_{\tau_n}$ then
  - if $B = 0$ then neither I nor L can jump,
  - if $B = 1$ then we have the following possibilities:
    - $L_{\tau_{n+1}} = L_{\tau_n} + 1$ and $I_{\tau_{n+1}} \in F_{n_{I_{\tau_n}}+1}$,
    - $L_{\tau_{n+1}} = L_{\tau_n}$ and $I_{\tau_{n+1}} \in F_{n_{I_{\tau_n}}} \cup F_{n_{I_{\tau_n}}+1}$,
    - $L_{\tau_{n+1}} = L_{\tau_n} - 1$ and $I_{\tau_{n+1}} \in F_{n_{I_{\tau_n}}} \cup F_{n_{I_{\tau_n}}-1}$.

Here, the respective probabilities of those jumps that are admissible are chosen in such a way that $\mathbf{X}$ and L takes separately are indeed Markov processes with respective generators $\mathbf{L}$ and G. Note that we used again the notation $F_{-1} = F_{\bar{n}+1} = \emptyset$.

In words, if $L \neq n_I$, L and $\mathbf{X}$ move independently from each other until the time where n_I and L agree. After that time, it is guaranteed that one always has $n_I \geq L$. We have not detailed precisely where I jumps exactly to be concise. But, if we ignore N, the couple $(\mathbf{X}, L)$ is just the Markov process generated by

$$\begin{aligned}\mathcal{G}f(x,i,l) = \mathcal{L}^{(i)}f(x,i,l) \\ + r\left(\frac{1}{2}\mathcal{A}f(x,i,l) + \frac{1}{2}\mathcal{B}f(x,i,l) - f(x,i,l)\right)\end{aligned}$$

if $l \neq n_i$, where

$$\mathcal{A}_r f(x,i,l) = \frac{2}{r}\sum_{j\in F} a(x,i,j)f(x,j,l) + \left(1 - \frac{2}{r}\sum_{j\in F} a(x,i,j)\right) f(x,i,l),$$

and

$$\mathcal{B}_r f(x,i,l) = \frac{2}{r}b(l)f(x,i,l+1) + \frac{2}{r}d(l)f(x,i,l-1) + \left(1 - \frac{2}{r}b(l) - \frac{2}{r}d(l)\right) f(x,i,l).$$

Otherwise,

$$\begin{aligned}\mathcal{G}f(x,i,n_i) &= \mathcal{L}^{(i)}f(x,i,n_i)\\ &\quad + r\left(\frac{1}{2}f(x,i,n_i)+\frac{1}{2}\mathcal{C}_r f(x,i,n_i)-f(x,i,n_i)\right),\end{aligned}$$

where

$$\begin{aligned}\mathcal{C}_r f(x,i,n_i) &= \frac{2}{r}\sum_{j\in F_{n_i-1}} a(x,i,j)f(x,j,n_i-1)\\ &\quad + \frac{2}{r}\Big(d(n_i)-\sum_{j\in F_{n_i-1}} a(x,i,j)\Big)f(x,i,n_i-1)\\ &\quad + \frac{2}{r}\sum_{j\in F_{n_i}} a(x,i,j)\left(f(x,j,n_i)-f(x,i,n_i)\right)\\ &\quad + \frac{2}{r}\frac{b(n_i)}{\sum_{k\in F_{n_i+1}} a(x,i,k)}\sum_{j\in F_{n_i+1}} a(x,i,j)f(x,j,n_i+1)\\ &\quad + \frac{2}{r}\frac{\sum_{k\in F_{n_i+1}} a(x,i,k)-b(n_i)}{\sum_{k\in F_{n_i+1}} a(x,i,k)}\sum_{j\in F_{n_i+1}} a(x,i,j)f(x,j,n_i)\\ &\quad + \frac{2}{r}\left(1-d(n_i)-\sum_{j\in F_{n_i}\cup F_{n_i+1}} a(x,i,j)\right)f(x,i,n_i).\end{aligned}$$

This construction ensures that

$$\forall t\geq 0, \alpha(I_t)\geq\alpha(L_t).$$

Now, for all $t\geq 0$ and $x\in E$, we have

$$\begin{aligned}\mathbf{P}_t V^q(x) &= \mathbb{E}\left[P_{t-\tau_{N_t}}^{(I_{\tau_{N_t}})}V^q(X_{\tau_{N_t}})\right]\leq\mathbb{E}\left[e^{-\alpha(I_{\tau_{N_t}})(t-\tau_{N_t})}V^q(X_{\tau_{N_t}})+K^q\right]\\ &\leq \mathbb{E}\left[e^{-\alpha(L_{\tau_{N_t}})(t-\tau_{N_t})}V^q(X_{\tau_{N_t}})\right]+K^q\\ &= \mathbb{E}\left[e^{-\alpha(L_{\tau_{N_t}})(t-\tau_{N_t})}P_{\tau_{N_t}-\tau_{N_t-1}}^{(I_{\tau_{N_t}-1})}V^q(X_{\tau_{N_t-1}})\right]+K^q\\ &\leq\ldots\leq\mathbb{E}\left[e^{-\int_0^t q\alpha(L_s)ds}\right]V(x)+K^q\sum_{n\geq 0}\mathbb{E}\left[e^{-q\int_0^{\tau_n}\alpha(L_s)ds}\right].\end{aligned}\tag{3.16}$$

Now, using Lemma 3.2.6, there exist $C,\eta>0$ and $q\in(0,1]$ such that

$$\mathbb{E}\left[e^{-\int_0^t q\alpha(L_s)ds}\right]\leq Ce^{-\eta t},\tag{3.17}$$

Hence, it only remains to prove that

$$\sum_{n\geq 0} \mathbb{E}\left[e^{-q\int_0^{\tau_n}\alpha(L_s)ds}\right] < +\infty.$$

We cannot deduce this directly from equation (3.17) but, heuristically, if r and n are large enough then the law of large number gives that $\tau_n \approx n/r$ and thus

$$\mathbb{E}\left[e^{-\int_0^{\tau_n} q\alpha(L_s)ds}\right] \approx \mathbb{E}\left[e^{-\int_0^{n/r} q\alpha(L_s)ds}\right] \leq Ce^{-\eta n/r},$$

and the previous sum is finite. Now, we estimate the left hand side of the previous equation following that τ_n is lower than n/r, close to n/r or higher than n/r. Note that we have not

$$e^{-\int_0^{\tau_n} q\alpha(L_s)ds} \leq 1,$$

because α can be negative. Let $\epsilon > 0$ and denote by ϱ the worst case of decay:

$$\varrho = -\min\{\, q\alpha(k) \mid k \in F\}. \tag{3.18}$$

If τ_n is close to n/r then we have from (3.17) the bound

$$\begin{aligned}
&\mathbb{E}\left[e^{-\int_0^{\tau_n} q\alpha(L_s)ds}\mathbf{1}_{\{\tau_n\in[nr^{-1}(1-\epsilon),nr^{-1}(1+\epsilon)]\}}\right] \\
&\qquad\leq e^{2\rho\varepsilon n/r}\mathbb{E}\left[e^{-\int_0^{nr^{-1}(1-\epsilon)} q\alpha(L_s)ds}\mathbf{1}_{\{\tau_n\in[nr^{-1}(1-\epsilon),nr^{-1}(1+\epsilon)]\}}\right] \\
&\qquad\leq e^{2\rho\varepsilon n/r}\mathbb{E}\left[e^{-\int_0^{nr^{-1}(1-\epsilon)} q\alpha(L_s)ds}\right] \leq Ce^{-nr^{-1}(\eta-\epsilon(2\varrho+\eta))}.
\end{aligned}$$

Thereafter, we therefore fix $\epsilon < \eta(2\varrho+\eta)^{-1}$. Now, if τ_n is lower than n/r then, using Markov's inequality, we have

$$\begin{aligned}
\mathbb{E}\left[e^{-\int_0^{\tau_n} q\alpha(L_s)ds}\mathbf{1}_{\{\tau_n<nr^{-1}(1-\epsilon)\}}\right] &\leq e^{\varrho nr^{-1}(1-\epsilon)}\mathbb{P}\left(\tau_n < nr^{-1}(1-\epsilon)\right) \\
&\leq e^{\varrho nr^{-1}(1-\epsilon)}e^{\theta nr^{-1}(1-\epsilon)}\mathbb{E}\left[e^{-\theta\tau_n}\right] \\
&\leq \exp\left(-n\left(\ln\left(1+\frac{\theta}{r}\right)-(1-\epsilon)r^{-1}(\varrho+\theta)\right)\right),
\end{aligned}$$

for every $\theta \geq 0$. And finally, if τ_n is higher than n/r then, using the Cauchy-Schwarz and Markov

inequalities, we have

$$\begin{aligned}\mathbb{E}\left[e^{-\int_0^{\tau_n} q\alpha(L_s)ds}\mathbf{1}_{\{\tau_n>nr^{-1}(1+\epsilon)\}}\right] &\leq \mathbb{E}\left[e^{2\varrho\tau_n}\right]^{1/2}\mathbb{P}\left(\tau_n > nr^{-1}(1+\epsilon)\right)^{1/2}\\ &\leq \mathbb{E}\left[e^{2\varrho\tau_n}\right]^{1/2}\left(e^{-\theta' nr^{-1}(1+\epsilon)}\mathbb{E}\left[e^{\theta'\tau_n}\right]\right)^{1/2}\\ &\leq \exp\left(-\frac{n}{2}\left(\ln\left(1-\frac{2\varrho}{r}\right)+\ln\left(1-\frac{\theta'}{r}\right)+\frac{\theta'(1+\epsilon)}{r}\right)\right),\end{aligned}$$

where $\theta' \geq 0$. Note that in the previous inequality, we have supposed that $r > 2\varrho$. Let $\gamma \in (0,1)$, we set $\theta = \theta' = \gamma r$. We can find a large r and a small γ verifying

$$\ln(1+\gamma) - (1-\epsilon)\gamma - (1-\epsilon)r^{-1}\varrho > 0,$$

and

$$\ln\left(1-\frac{2\varrho}{r}\right) + \ln(1-\gamma) + \gamma(1+\epsilon) > 0.$$

and thus there exist $C' > 0$ and $\varepsilon > 0$ such that

$$\sum_{n\geq 1}\mathbb{E}\left[e^{-\int_0^{\tau_n} q\alpha(I_s)ds}\right] \leq \sum_{n\geq 1} C' e^{-\varepsilon n} < +\infty\,,$$

thus concluding the proof by combining this with (3.16) and (3.17). □

Remark 3.3.9. *If all Markov processes contract, then the proof simplifies considerably. Indeed, if for all $i \in F$, one has $\alpha(i) \geq \zeta > 0$, then one has*

$$\sum_{n\geq 1}\mathbb{E}\left[e^{-\int_0^{\tau_n} q\alpha(L_s)ds}\right] \leq \sum_{n\geq 1}\mathbb{E}\left[e^{-\zeta\tau_n}\right] \leq \sum_{n\geq 1}\left(\frac{r}{r+\zeta}\right)^n = \frac{r}{\zeta} < \infty\,.$$

3.3.3 The contracting distance

This section is divided in three parts. We introduce the distance $\widetilde{d}$ that we will use in Theorem 3.3.6, we build our coupling in such a way that $\widetilde{d}$ will be contracting for it, and we finally prove that it is indeed contracting.

3.3.3.1 Definition of $\widetilde{d}$

Here, we build a distance $\widetilde{d}\colon (E\times F)\times(E\times F)\to[0,1]$ such that there exist $t_*>0$ and $\alpha\in(0,1)$ verifying

$$\widetilde{d}(\mathbf{x},\mathbf{y})<1 \quad\Rightarrow\quad \forall t\geq t_*,\quad \mathcal{W}_{\widetilde{d}}(\delta_{\mathbf{x}}P_t,\delta_{\mathbf{y}}P_t)\leq\alpha\widetilde{d}(\mathbf{x},\mathbf{y}). \tag{3.19}$$

where $\mathbf{x}=(x,i)$ and $\mathbf{y}=(y,j)$ belong to $E\times F$. Since we can say nothing when $i\neq j$, we will take $\widetilde{d}(\mathbf{x},\mathbf{y})$ constant equal to 1 in this case. When $i=j$ we want to use Assumption 3.1.3 to prove a decay. But it is more useful to "decrease the contraction" of the underlying Markov semigroup. More precisely, by Jensen inequality, Assumption 3.1.3 gives

$$\mathcal{W}_{d^q}(\mu P_t^{(i)},\nu P_t^{(i)})\leq e^{-q\rho(i)t}\mathcal{W}_{d^q}(\mu,\nu),$$

for all $t\geq 0$, $q\in(0,1]$ and every probability measures μ,ν. Finally, we define $\widetilde{d}$ by

$$\widetilde{d}(\mathbf{x},\mathbf{y})=\mathbf{1}_{i\neq j}+\mathbf{1}_{i=j}\left(\delta^{-1}d^q(x,y)\wedge 1\right),$$

where $\delta>0$ will be determined later. Now, if a coupling $(\mathbf{X}_t,\mathbf{Y}_t)_{t\geq0}=((X_t,I_t),(Y_t,J_t))_{t\geq0}$ starting from $(\mathbf{x},\mathbf{y})$, verifies $\widetilde{d}(\mathbf{x},\mathbf{y})<1$, then $I_0=J_0=i=j$. So, we will try to build our coupling in such a way that I and J remain equal for as long as possible. More precisely, if we set

$$T=\inf\{s\geq0\mid I_s\neq J_s\}, \tag{3.20}$$

then we will prove that there exists $K>0$ and a choice of coupling such that

$$\mathbb{P}(T<\infty)\leq Kd(x,y).$$

3.3.3.2 Construction of our coupling

Here, we fix $\mathbf{x}=(x,i)$, $\mathbf{y}=(y,j)$ in $\mathbf{E}$ and we let $t>0$. Let $r\geq0$ and $(N_t)_{t\geq0}$ be a Poisson process of intensity r with $N_t=\sum_{n\geq0}\mathbf{1}_{\{\tau_n\leq t\}}$ and $\tau_n=\sum_{k=1}^n E_k$ for a family $(E_k)_{k\geq0}$ of i.i.d. exponential variables as before and $\tau_0=0$. We assume that $r\geq2\bar{a}$, i.e. that is r is bigger than the jump rates of I or J. As in the proof of Lemma 3.3.8 and Theorem 3.1.4, we give the construction of our coupling $(\mathbf{X},\mathbf{Y})$ at the jump times of N. Let $n\in\{0,..,N_t\}$, we consider the following dynamics:

– If $I_{\tau_n}\neq J_{\tau_n}$ then X_s and Y_s evolve independently for every $s\in[\tau_n,\tau_{n+1}\wedge t)$.

– If $I_{\tau_n} = J_{\tau_n}$ then by Assumption 3.1.3, we can couple X and Y in such a way that

$$\mathbb{E}\left[d(X_{\tau_{n+1}\wedge t}, Y_{\tau_{n+1}\wedge t}) \mid \mathcal{G}_{\tau_n}\right] \leq e^{-\rho(I_{\tau_n})(\tau_{n+1}\wedge t-\tau_n)} d(X_{\tau_n}, Y_{\tau_n}),$$

where $\mathcal{G}_n = \sigma\{(\mathbf{X}_{\tau_n}, \mathbf{Y}_{\tau_n}), (\tau_k)_{k\geq 0}\}$.

At the jump times of N the situation is different since I or J may jump. We will optimise the chance that I and J jump simultaneously. For each $n \in \mathbb{N}^*$, we cut $[0,1]$ in four parts $I_0^n, I_1^n, I_2^n, I_3^n$ in such a way that

$$\begin{aligned}
\lambda(I_0^n) &= \frac{1}{r}\sum_{j\in F}\left(a(X_{\tau_n-}, I_{\tau_n}, j) - a(Y_{\tau_n-}, I_{\tau_n}, j)\right)_+, \\
\lambda(I_1^n) &= \frac{1}{r}\sum_{j\in F}\left(a(Y_{\tau_n-}, I_{\tau_n}, j) - a(X_{\tau_n-}, I_{\tau_n}, j)\right)_+, \\
\lambda(I_2^n) &= \frac{1}{r}\sum_{j\in F} a(X_{\tau_n-}, I_{\tau_n}, j) \wedge a(Y_{\tau_n-}, I_{\tau_n}, j), \\
\lambda(I_3^n) &= 1 - \frac{1}{r}\sum_{j\in F} a(X_{\tau_n-}, I_{\tau_n}, j) \vee \sum_{j\in F} a(Y_{\tau_n-}, I_{\tau_n}, j),
\end{aligned}$$

where λ is the Lebesgue measure and $(x)_+ = \max(x, 0)$. Let $(U_n)_{n\geq 0}$ be a sequence of i.i.d. random variables uniformly distributed on $[0,1]$, we couple I and J at the jump times as follows:

– For $U_n \in I_0^n$, I jumps, but J does not jump.
– For $U_n \in I_1^n$, J jumps, but I does not jump.
– For $U_n \in I_2^n$, I and J both jump simultaneously to the same location.
– For $U_n \in I_3^n$, I and J both stay in place.

The second components, X and Y, do not jump. Finally, we also couple $\mathbf{X}$ and $\mathbf{Y}$ with a continuous Markov chain L which only depend to U and N and which verifies

$$\forall t \geq 0,\ \rho(I_t) \geq \alpha(L_t).$$

This Markov chain L is constructed as in the proof of Lemma 3.3.8.

Remark 3.3.10. *This coupling is not quite Markovian since, between times τ_n and τ_{n+1}, it already uses information about the pair (X_t, Y_t) at time τ_{n+1}. However, in many situations to which our results apply there exists a Markovian coupling with generator $\mathbb{L}^{(i)}$ which minimises the Wasserstein distance for each of the underlying processes. In this case, we can make our coupling Markovian with*

generator

$$\begin{aligned}\mathbb{L}f(\mathbf{x},\mathbf{y},n) = \mathbb{L}^{(i)}f(\mathbf{x},\mathbf{y},n) &+ \sum_{k\in F}\left(a(x,i,k)-a(y,j,k)\right)_+ f((x,k),\mathbf{y},n+1)\\ &+ \sum_{k\in F}\left(a(y,j,k)-a(x,i,k)\right)_+ f(\mathbf{x},(y,k),n+1)\\ &+ \sum_{k\in F} a(x,i,k)\wedge a(y,j,k) f((x,k),(y,k),n+1)\\ &+ \Big(r-\sum_{k\in F} a(x,i,k)\vee a(y,j,k)\Big) f(\mathbf{x},\mathbf{y},n+1) - rf(\mathbf{x},\mathbf{y},n).\end{aligned}$$

3.3.3.3 The distance $\widetilde{d}$ is contracting for P

In this subsection, we show that the distance $\widetilde{d}$ defined above is indeed contracting for the coupling constructed in the previous subsection. This is formulated in the following result.

Lemma 3.3.11. *Let $(\mathbf{X}_t,\mathbf{Y}_t)_{t\geq 0}$ be the coupling of the previous section. Under the assumptions of Theorem 3.3.2, we can choose r and δ in such a way that*

$$\forall t\geq t_*,\ \mathbb{E}\left[\widetilde{d}(\mathbf{X}_t,\mathbf{Y}_t)\right]\leq \gamma\widetilde{d}(\mathbf{x},\mathbf{y}),$$

for some $\gamma\in(0,1)$ and $t_*>0$, and all $\mathbf{x},\mathbf{y}\in E\times F$ verifying $\widetilde{d}(\mathbf{x},\mathbf{y})<1$.

Proof. Recall that since $\widetilde{d}(\mathbf{x},\mathbf{y})<1$ one has $I_0=J_0$ and that T, defined in (3.20), denotes the first time of separation of I and J. Using Lemma 3.2.6, there exist $q\in(0,1]$ and $C,\eta>0$ such that

$$\begin{aligned}\mathbb{E}\left[\widetilde{d}(\mathbf{X}_t,\mathbf{Y}_t)\right] &\leq \mathbb{E}\left[\mathbf{1}_{\{T=\infty\}}\frac{1}{\delta}d^q(X_t,Y_t)+\mathbf{1}_{\{T<+\infty\}}\right]\\ &\leq \frac{1}{\delta}\mathbb{E}\left[e^{-\int_0^t q\alpha(L_s)ds}\right]\mathbb{E}[d^q(x,y)]+\mathbb{P}\left(T<+\infty\right).\\ &\leq Ce^{-\eta t}\widetilde{d}(\mathbf{x},\mathbf{y})+\mathbb{P}\left(T<+\infty\right).\end{aligned}$$

Here, we have used the fact that

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}_{\{T=\infty\}}d^q(X_t,Y_t)\right] &\leq \mathbb{E}\left[\mathbf{1}_{\{T\geq\tau_{N_t}\}}e^{-q\alpha(L_{\tau_{N_t}})(t-\tau_{N_t})}d^q(X_{\tau_{N_t}},Y_{\tau_{N_t}})\right]\\ &\leq \mathbb{E}\left[\mathbf{1}_{\{T\geq\tau_{N_t}\}}e^{-q\alpha(L_{\tau_{N_t}})(t-\tau_{N_t})}\mathbb{E}\left[d^q(X_{\tau_{N_t}},Y_{\tau_{N_t}})\mid\mathcal{G}_n\right]\right]\\ &\leq \mathbb{E}\left[\mathbf{1}_{\{T\geq\tau_{N_t-1}\}}e^{-\int_{\tau_{N_t-1}}^t q\alpha(L_s)ds}d^q(X_{\tau_{N_t-1}},Y_{\tau_{N_t-1}})\right]\\ &\leq \mathbb{E}\left[e^{-\int_0^t q\alpha(L_s)ds}\right]\mathbb{E}[d^q(x,y)]\ .\end{aligned}$$

It remains to obtain a bound on $\mathbb{P}(T < +\infty)$. Since I and J can only jump when N jumps, T can be finite only if it is one of the jump times of N. So, we set

$$A_n = \{T = \tau_n\} = \{T \geq \tau_n \text{ and } I_{\tau_n} \neq J_{\tau_n}\}.$$

By Assumption 3.1.1, we have

$$\begin{aligned}
\mathbb{P}(A_n) &= \mathbb{P}(\{U_n \in I_0^n \cup I_1^n \cup I_3^n\} \cap \{T \geq \tau_n\}) \\
&\leq \mathbb{E}\left[\frac{2\mathbf{1}_{\{T \geq \tau_n\}} \sum_{j \in F} |a(X_{\tau_n -}, I_{\tau_n -}, j) - a(Y_{\tau_n -}, I_{\tau_n -}, j)|}{r}\right] \\
&\leq \mathbb{E}\left[\left(\frac{2\mathbf{1}_{\{T \geq \tau_n\}} \sum_{j \in F} |a(X_{\tau_n -}, I_{\tau_n -}, j) - a(Y_{\tau_n -}, I_{\tau_n -}, j)|}{r}\right)^q\right] \\
&\leq \frac{2^q \kappa^q}{r^q} \mathbb{E}\left[d(X_{\tau_n -}, Y_{\tau_n -})^q\right] \leq \frac{2^q \kappa^q}{r^q} \mathbb{E}\left[e^{-q \int_0^{\tau_n} \alpha(L_s) ds}\right] d(x,y)^q.
\end{aligned}$$

Hence

$$\mathbb{P}(T < \infty) = \sum_{n \geq 1} \mathbb{P}(A_n) \leq \frac{2^q \kappa^q}{r^q} d(x,y)^q \sum_{n \geq 1} \mathbb{E}\left[e^{-q \int_0^{\tau_n} \alpha(L_s) ds}\right].$$

Now, similarly to the proof of Lemma 3.3.8, if r is large enough then there exist $C' > 0$ and $\varepsilon > 0$ verifying

$$\sum_{n \geq 1} \mathbb{E}\left[e^{-q \int_0^{\tau_n} \alpha(L_s) ds}\right] \leq \sum_{n \geq 1} C' e^{-\varepsilon n} =: \tilde{C} < +\infty.$$

Combining these bounds, we obtain the estimate

$$\mathbb{E}\left[\tilde{d}(\mathbf{X}_t, \mathbf{Y}_t)\right] \leq \left(Ce^{-\eta t} + \frac{(2\kappa)^q \tilde{C}}{r^q} \delta\right) \tilde{d}(\mathbf{x}, \mathbf{y}).$$

First making δ sufficiently small and then taking t large enough, we thus obtain the announced result. □

3.3.4 Bounded sets are $\tilde{d}$-small

Here, we prove that if a set is bounded then it is $\tilde{d}$-small.

Lemma 3.3.12. *Under the assumptions of Theorem 3.3.2, if $S \subset E \times F$ is of bounded diameter in the sense that*

$$R = \sup\{d(x,y) \mid \mathbf{x}, \mathbf{y} \in S\} < +\infty,$$

then there exist $t_*, t^* > 0$ such that S is $\tilde{d}$-small for P_t, for all $t \in [t_*, t^*]$.

Proof. Let $\mathbf{x} = (x, i)$ and $\mathbf{y} = (y, j)$ be two different points of S. By the assumptions of Theorem 3.3.2, there exists $i_0 \in F$ such that $\rho(i_0) > 0$. Let $(\mathbf{X}_t)_{t\geq 0}$ and $(\mathbf{Y}_t)_{t\geq 0}$ be two independent processes generated by (3.2) and starting respectively from $\mathbf{x}$ and $\mathbf{y}$. Let us denote

$$\tau_{\text{in}} = \inf\{t \geq 0 \mid I_t = J_t = i_0\} \quad \text{and} \quad \tau_{\text{out}} = \inf\{t \geq \tau_{\text{in}} \mid I_t \neq i_0 \text{ or } J_t \neq i_0\}.$$

For every $b, c > 0$ such that $b > c$, we define

$$p_{c,b}(\mathbf{x}, \mathbf{y}) = \mathbb{P}(\tau_{\text{in}} < c, \tau_{\text{out}} > b) \ .$$

By Assumptions 3.1.1 and 3.1.2, we have $p_{c,b}(\mathbf{x}, \mathbf{y}) > 0$. Using the fact that a is bounded, a coupling argument shows that $p_{c,b}$ is lower bounded by a positive quantity which only depends on i and j. We then obtain the bound

$$\begin{aligned}
\mathbb{E}\left[\widetilde{d}(\mathbf{X}_t, \mathbf{Y}_t)\right] &\leq \mathbb{E}\left[\mathbf{1}_{\{\tau_{\text{in}}<c,\tau_{\text{out}}>b\}}\widetilde{d}(\mathbf{X}_t, \mathbf{Y}_t)\right] + 1 - p_{c,b}(\mathbf{x}, \mathbf{y}) \\
&\leq 1 - p_{c,b}(\mathbf{x}, \mathbf{y})\left(1 - \delta^{-1}e^{\varrho c}e^{-\rho(i_0)t}d(x, y)\right) \\
&\leq 1 - p_{c,b}(\mathbf{x}, \mathbf{y})\left(1 - \delta^{-1}e^{\varrho c}e^{-\rho(i_0)t}R\right),
\end{aligned}$$

where ϱ was defined in (3.18). There exist $c > 0$ and $t_* > c$ such that $1 - \delta^{-1}e^{\varrho c}e^{-\rho(i_0)t_*}R > 0$. Since F is finite, we can furthermore bound $p_{c,b}$ from below by the minimum over all $i, j \in F$, and the result follows for any $b > t_*$ and $t^* \in (t_*, b)$. □

Remark 3.3.13. *One can see from this proof that it is not necessary that the jump rates are lower bounded, as in Assumption 3.1.2. Indeed, we need that, for each $i, j \in F$, the jump times of I are stochastically smaller than a variable which does not depend of the dynamics of X.*

3.3.5 Proofs of Theorem 3.1.5 and Theorem 3.3.2

Lemma 3.2.2 and Lemma 3.3.8 give the existence of a Lyapunov function $V = V_{x_0}$, for some $x_0 \in E$, Lemma 3.3.11 shows that $\widetilde{d}$ is contracting for $\mathbf{P}$, and Lemma 3.3.12 proves that sublevel sets of V are $\widetilde{d}$-small. So we can use Theorem 3.3.6 to deduce that there exist a probability measure $\boldsymbol{\pi}$ and some constants $C, \lambda, t_0 > 0$ such that

$$\forall t \geq t_0, \ \mathcal{W}_{\widetilde{\mathbf{d}}}(\boldsymbol{\mu}\mathbf{P}_t, \boldsymbol{\pi}) \leq Ce^{-\lambda t}\mathcal{W}_{\widetilde{\mathbf{d}}}(\boldsymbol{\mu}, \boldsymbol{\pi}),$$

for every probability measure μ on $\mathbf{E}$. In this expression, $\widetilde{\mathbf{d}}$ is defined by

$$\widetilde{\mathbf{d}}(\mathbf{x},\mathbf{y}) = \sqrt{(\mathbf{1}_{i\neq j} + \mathbf{1}_{i=j}(1 \wedge d^q(x,y)))(1 + d^q(x,x_0) + d^q(y,x_0))},$$

where $\mathbf{x} = (x,i)$, $\mathbf{y} = (y,j)$ belong to $\mathbf{E}$, x_0 is as in Assumption 3.1.3 and $q \in (0,1]$. Noting that $\mathbf{d} \leq \widetilde{\mathbf{d}}$ we find

$$\forall t \geq t_0,\ \mathcal{W}_{\mathbf{d}}\left(\delta_{\mathbf{y}_0}\mathbf{P}_t, \boldsymbol{\pi}\right) \leq Ce^{-\lambda t}\left(1 + \sum_{i\in F}\int_E d(y_0,x)^q \boldsymbol{\pi}(dx,i)\right),$$

and as

$$\forall t \leq t_0,\ \mathcal{W}_{\mathbf{d}}\left(\delta_{\mathbf{y}_0}\mathbf{P}_t, \boldsymbol{\pi}\right) \leq 1,$$

This ends the proof.

3.4 Two special cases

Here, we give some sufficient conditions allowing to verify our main assumptions in situations where the underlying processes are deterministic or diffusive. Note that we can find sufficient conditions in [Clo12] for stochastically monotone processes, in [**?**] for birth-death processes and in [Ebe11] for diffusion processes.

3.4.1 The case of diffusion processes

Let us recall that a diffusion process on $\mathbb{R}^d$, $d \in \mathbb{N}^*$, is a process generated by

$$\forall x \in \mathbb{R}^d,\ \mathcal{L}f(x) = \sum_{i=1}^d b_i(x)\partial_i f(x) + \sum_{i,j=1}^d (\sigma(x)\sigma(x)^t)_{i,j}\partial_{i,j} f(x), \tag{3.21}$$

where f is a smooth enough function and b, σ are regular enough, say

$$\forall x,y \in \mathbb{R}^d,\ \|\sigma(x) - \sigma(y)\| + \|b(x) - b(y)\| \leq K\|x-y\|. \tag{3.22}$$

for some $K > 0$. In the previous expression, $\|\cdot\|$ denotes both the Euclidean norm and the subordinate norm.

Lemma 3.4.1. *Let $(P_t)_{t\geq 0}$ be the Markov semigroup generated by (3.21). If (3.22) holds and*

$$\forall x,y \in \mathbb{R}^d,\ \langle b(x) - b(y), x - y\rangle \leq -\alpha\|x-y\|^2,$$

for some $\alpha \in \mathbb{R}$, then

$$\forall t \geq 0,\ \mathcal{W}_{\|\cdot\|}(\mu P_t, \nu P_t) \leq e^{-\alpha t}\mathcal{W}_{\|\cdot\|}(\mu, \nu),$$

for any probability measures μ and ν.

Proof. Let $p > 1$. Considering the same Brownian motion for two different solutions $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ of the SDE starting with different initial measures, we can show that if

$$\frac{p-1}{2}(\sigma(x)-\sigma(y))(\sigma(x)-\sigma(y))^t + \langle b(x)-b(y), x-y\rangle \leq -\alpha_p\|x-y\|^2 \tag{3.23}$$

then

$$\forall t \geq 0,\ \mathbb{E}\left[\|X_t - Y_t\|^p\right] \leq e^{-p\alpha_p t}\mathbb{E}\left[\|X_0 - Y_0\|^p\right]. \tag{3.24}$$

Now, if $u \in \mathbb{R}^d$ verifies $\|u\| = 1$ then by (3.22) we have

$$\begin{aligned} u(\sigma(x)-\sigma(y))(\sigma(x)-\sigma(y))^t u &\leq \|u(\sigma(x)-\sigma(y))\|\|(\sigma(x)-\sigma(y))^t u\| \\ &\leq \|\sigma(x)-\sigma(y)\|^2 \\ &\leq K^2\|x-y\|^2. \end{aligned}$$

Hence, under our assumptions, the inequality (3.23) is satisfied with

$$\alpha_p = \alpha - \frac{p-1}{2}K^2.$$

Taking the limit $p \to 1$ in (3.24) and the infimum over the starting distribution ends the proof. □

Assumptions of Theorem 3.1.7 or Theorem 3.1.8 are satisfied if one of the underlying diffusions verifies Hörmander's hypoellipticity assumption. See for instance [Hai11] for an introduction on this subject.

Remark 3.4.2 (Exponential convergence for an infinite dimensional process)**.** *The previous result gives also the convergence for switching Fokker-Planck processes. Indeed, we can consider that each underlying Markov process $(Z_t^{(i)})_{t\geq 0}$ is deterministic, belongs to the space of smooth density functions, and verifies*

$$\partial_t Z_t^{(i)}(x) = \sum_{k=1}^{d} -\partial_k(b_k Z_t^{(i)})(x) + \sum_{k,l=1}^{d} \partial_{k,l}(\sigma_{k,l} Z_t^{(i)})(x)$$

for all $x \in \mathbb{R}^d$, and $t \geq 0$. The previous lemma gives a contraction as in Assumption 3.1.3, for each underlying process, where d is the Wasserstein metric.

3.4.2 Case of piecewise deterministic Markov processes

Let us assume that each one of the underlying Markov processes is actually deterministic. More precisely, we consider that E is an open of $\mathbb{R}^d$, $d \in \mathbb{N}^*$ and $\mathcal{L}^{(i)} f = G^{(i)} \cdot \nabla f$, for every $i \in F$, where $(G^{(i)})_{i \in F}$ is a family of vector fields such that the ordinary differential equations $x' = G^{(i)}(x)$ have a unique and global solution for any initial condition, for every $i \in F$. Lemma 3.4.1 gives the assumption in order to apply Theorem 3.1.4 and Theorem 3.1.5. In general, we can not apply Theorem 3.1.7 or Theorem 3.3.3 but [BH12, BLMZ12b] give a sufficient condition ensuring that $\mathbf{X}$ generates densities:

Assumption 3.4.3 (Hörmander-type bracket conditions)**.** *Let* $\mathcal{G}_0 = \{G^{(i)} - G^{(j)},\ i \neq j\}$ *and for all* $k \geq 0$,

$$\mathcal{G}_{k+1} = \{[G^{(i)}, G] \mid i \in F,\ G \in \mathcal{G}_k\},$$

where $[,]$ *designs the Lie bracket. We have* $\mathcal{G}_k(x) = \{G(x) \mid G \in \mathcal{G}_k\} = \mathbb{R}^d$, *for every* $x \in E$.

In this case our main theorem gives

Theorem 3.4.4. *Let us suppose that Assumptions 3.1.1, 3.1.2 and 3.4.3 hold. If one of the two following assumptions is satisfied:*

– $a(x, i, j)$ *does not depend to* x *and* I *is ergodic with an invariant measure* ν *satisfying*

$$\sum_{i \in F} \nu(i)\lambda(i) > 0;$$

– *Assumption 3.3.1 holds and*

$$\sum_{i \in F} \nu(i)\alpha(i) > 0,$$

for some increasing sequence α *satisfying* $\alpha(n) \leq \min_{i \in F_n} \lambda(i)$, *for all* $n \leq \bar{n}$.

then there exist a probability measure $\boldsymbol{\pi}$ *and three constants* $C, \lambda, t_0 > 0$ *such that*

$$\forall t \geq t_0,\ d_{\mathrm{TV}}\left(\delta_{\mathbf{x}} \mathbf{P}_t, \boldsymbol{\pi}\right) \leq C e^{-\lambda t}(1 + V(x)),$$

for every $\mathbf{x} = (x, i) \in \mathbf{E}$.

Proof. Using [BLMZ12b, Theorem 6.6], we see that compact sets are small for $\mathbf{X}$. Using Lemma 3.2.7 in the first case and Lemma 3.3.8 in the second case, we see that we can apply Theorem 3.2.10. □

3.5 Examples

Here, we give three simple examples to illustrate our results.

3.5.1 The most elementary example

Let us consider the example where X belongs to $\mathbb{R}$ and verifies

$$\forall t \geq 0,\ \partial_t X_t = I_t X_t,$$

where $(I_t)_{t\geq 0}$ is the continuous time Markov chain, on $\{-1, 1\}$, which jumps from 1 to -1 with rate $a_1 > 0$ and from -1 to 1 with rate $a_{-1} > 0$. If $a_1 > a_{-1}$ then Theorems 3.1.4 and 3.1.5 give the exponential ergodicity of $\mathbf{X}$ in the Wasserstein distance. Here, the invariant law is

$$\delta_0 \otimes \frac{1}{a_{-1} + a_1}(a_1\delta_{-1} + a_{-1}\delta_1),$$

and there is clearly no convergence in total variation. Thus, classical Harris' Theorem does not work here. Furthermore, the classical law of large number gives

$$\lim_{t\to+\infty} X_t = \begin{cases} 0 \text{ a.s. , if } a_1 > a_{-1}, \\ +\infty \text{ a.s. , if } a_1 < a_{-1}. \end{cases}$$

In particular, there is no convergence when $a_1 < a_{-1}$.

Remark 3.5.1. *In our main theorems, we use a Wasserstein distance associated to a distance comparable to d^q rather than d. We choose this distance because, in general, moments of $\mathbf{X}$ can explode even though $\mathbf{X}$ converges in law. For instance, in the above example, one has $\lim_{t\to\infty} \mathbb{E}X_t = \infty$ as soon as $a_1 < 1$. See also [BGM10] for comments on the optimal choice of the parameter q.*

3.5.2 Wasserstein contraction of some switching dynamical systems

Let us consider a slight generalisation of the previous example; that is X belongs to $\mathbb{R}$ and verifies

$$\forall t \geq 0,\ \partial_t X_t = -a(I_t)X_t, \tag{3.25}$$

where $(I_t)_{t\geq 0}$ is a recurrent continuous time Markov chain on a finite state space F and a a function

from F to $\mathbb{R}$. Theorem 3.1.4 gives the exponential-Wasserstein ergodicity under the condition that

$$\sum_{i\in F} a(i)\nu(i) > 0, \tag{3.26}$$

where ν is a invariant measure of I. This simple example satisfies a bound like in Assumption 3.1.3. Indeed we have

Lemma 3.5.2. *If* (3.25) *and* (3.26) *are satisfied then there is a distance* $\boldsymbol{\delta}$ *on* $\mathbf{E}$ *such that the Wasserstein curvature of the semigroup of* $\mathbf{X}$ *is positive, i.e. there exists* $\lambda > 0$ *such that*

$$\forall t \geq 0,\ \mathcal{W}_{\boldsymbol{\delta}}\left(\delta_{\mathbf{x}}\mathbf{P}_t, \delta_{\mathbf{y}}\mathbf{P}_t\right) \leq e^{-\lambda t}\boldsymbol{\delta}(\mathbf{x},\mathbf{y}),$$

for all $\mathbf{x},\mathbf{y} \in \mathbf{E}$.

Proof. Firstly, let us give a complement on the conclusion of Lemma 3.2.6. The Markov chain I satisfies its assumptions and using the results of [BGM10], there exist a function ψ on F, $\rho > 0$ and $p \in (0,1)$ verifying

$$\forall t \geq 0,\ \mathbb{E}\left[\psi(I_t)e^{-\int_0^t pa(I_s)ds}\right] = e^{-\rho t}\mathbb{E}\left[\psi(I_0)\right].$$

Now let $\boldsymbol{\delta}$ be the distance, on $\mathbf{E}$, defined by

$$\forall \mathbf{x},\mathbf{y} \in \mathbf{E},\ \boldsymbol{\delta}(\mathbf{x},\mathbf{y}) = \mathbf{1}_{\{i=j\}}\psi(i)|x-y|^p + \mathbf{1}_{\{i\neq j\}}\frac{\overline{\psi}}{\underline{\psi}}(\psi(i)|x|^p + \psi(j)|y|^p + 1),$$

where

$$\overline{\psi} = \max_{k\in F}\psi(k) \ \text{ and } \ \underline{\psi} = \min_{k\in F}\psi(k).$$

Now, using the fact that

$$\forall t \geq 0, X_t = X_0 e^{-\int_0^t a(I_s)ds},$$

the proof is straightforward. □

3.5.3 Surprising blow-up under exponential ergodicity assumptions

Here we give some comments on [BLMZ12a, Example 1.4], which also illustrate the sharpness of our criteria. Let us consider $E = \mathbb{R}^2$, $F = \{0,1\}$, $\mathcal{L}^{(i)}f = A_i \cdot \nabla f$ where

$$A_0 = \begin{pmatrix} -1 & 3 \\ -1/3 & -1 \end{pmatrix} \ \text{ and } \ A_1 = \begin{pmatrix} -1 & -1/3 \\ 3 & -1 \end{pmatrix},$$

$a(x,0,0) = a(x,1,1) = 0$, and $a(x,1,0) = a(x,0,1) = a > 0$, for all $x \in \mathbb{R}^2$. In short, $\mathbf{X}$ is generated, for all $x \in \mathbb{R}^2$ and $i \in \{0,1\}$, by

$$\mathbf{L}f(x,i) = A_i \cdot \nabla f(x,i) + a\left(f(x,1-i) - f(x,i)\right). \tag{3.27}$$

Since a does not depend on its first component, I is a Markov process and it converges exponentially to

$$\nu = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1.$$

For each $i \in \{0,1\}$, we have $\partial_t Z_t^{(i)} = A_i Z_t^{(i)}$ and thus we easily prove that

$$\left\|Z_t^{(i)}\right\|_i \leq e^{-t}\left\|Z_0^{(i)}\right\|_i \quad \text{and} \quad \left\|Z_t^{(i)}\right\|_{1-i} \leq 3e^{-t}\left\|Z_0^{(i)}\right\|_{1-i}, \tag{3.28}$$

for every $t \geq 0$, where the norms $\|\cdot\|_0$ and $\|\cdot\|_1$ are defined by

$$\forall u = (u_1,u_2) \in \mathbb{R}^2, \|u\|_0 = \sqrt{(u_1/3)^2 + u_2^2} \quad \text{and} \quad \|u\|_1 = \sqrt{u_1^2 + (u_2/3)^2}.$$

Thus each flow $i \in \{0,1\}$ contracts, with the norm $\|\cdot\|_i$, and converges geometrically, with the norm $\|\cdot\|_{1-i}$, to the same limit. Nevertheless, if a is large enough then [BLMZ12a, Example 1.4] shows that

$$\lim_{t\to+\infty} \|X_t\| = +\infty.$$

In particular, the conclusion of Theorem 3.1.4 is not satisfied. This illustrates the fact that assuming that each underlying dynamics converges geometrically is not sufficient in general to guarantee the convergence of X. Moreover, this shows that it is essential in Theorem 3.1.4 to measure the constants $\rho(i)$ with respect to the *same* distance for every i. Note that the Wasserstein curvature of $Z^{(i)}$, with respect to $\|\cdot\|_{1-i}$, is negative and given by $-37/3$.

3.5.4 Non-convergence when I is recurrent but not positive recurrent

A last example is the following: the process X verifies

$$\forall t \geq 0,\ dX_t = -(X_t - a_{I_t})dt,$$

where $(a_n)_{n\geq 0}$ is a bounded real sequence and I is an irreducible and recurrent continuous time Markov chain which is not positive recurrent. It is easy to see that the sequence of laws of $(X_t)_{t\geq 0}$ is

tight and we can hope that there exists a probability measure π verifying

$$\lim_{t\to+\infty} \mathbb{E}\left[f(X_t)\right] = \int f d\pi,$$

for every continuous and bounded function f and any starting distribution. But in general, this is false. To illustrate it, let us consider the case when I is the classical continuous-time random walk on $\mathbb{N}$ reflected at 0. Namely, I is generated by

$$Jf(i) = \frac{1}{2}f(i+1) + \frac{1}{2}f(i-1) - f(i).$$

if $i \neq 0$ and

$$Jf(0) = f(1) - f(0).$$

The sequence a on the other hand is defined recursively by:

$$a_{n+1} = \begin{cases} a_n \text{ if } n \notin \{2^k \mid k \in \mathbb{N}\}, \\ -a_n \text{ if } n \in \{2^k \mid k \in \mathbb{N}\}. \end{cases}$$

In this case, the central limit theorem gives that $I_t \approx \sqrt{t}$ and so, for very large times, I and a do not switch on the same time scale. As a matter of fact, the process a_{I_t} stays constant during longer and longer stretches of time. It is then possible to find two sequences of *deterministic* times $(t_n)_{n\geq 0}$ and $(s_n)_{n\geq 0}$, both converging to infinity, and such that

$$\lim_{n\to+\infty} \mathbb{E}\left[f(X_{t_n})\right] = f(0) \text{ and } \lim_{n\to+\infty} \mathbb{E}\left[f(X_{s_n})\right] = f(1).$$

Thus this process exhibits ageing and is not exponentially stable, even though there exists $C > 0$, such that for any two starting points $\mathbf{x} = (x, i)$ and $\mathbf{y} = (y, j)$, we have

$$\forall t \geq 0,\ \mathcal{W}_{\mathbf{d}_0}(\delta_{\mathbf{x}}\mathbf{P}_t, \delta_{\mathbf{y}}\mathbf{P}_t) \leq \frac{C}{\sqrt{t}}|i-j|,$$

where $\mathbf{d}_0(\mathbf{x}, \mathbf{y}) = \mathbf{1}_{i=j}\|x-y\| \wedge 1 + \mathbf{1}_{i\neq j}$.

Deuxième partie

Comportement asymptotique de populations structurées

Chapitre 4

De l'équation de Schrödinger aux formules *many-to-one*

4.1 Introduction

Avant de présenter nos résultats sur des processus avec branchement, il nous a paru intéressant de mettre en avant une technique récurrente dans nos articles : la $h-$transformée de Doob. Elle a été introduite pour la première fois par celui-ci en 1957 dans [Doo57]. Cette transformation, de type Girsanov, permet d'interpréter un semigroupe de type Feynman-Kac en un semigroupe de Markov, sous condition de connaître une fonction harmonique h. Elle intervient dans beaucoup de domaines d'applications des probabilités. Nous nous en servons dans trois différents contextes dans les chapitres 2, 3 et la partie II. Le but de ce chapitre introductif est de mettre en relief où elle intervient dans nos différents résultats et d'introduire les formules *many-to-one*. Ces formules, liées aux processus avec branchement, nous disent que le comportement moyen d'une population de particules ne se résume qu'à l'étude d'une seule particule au comportement biaisé. Cette formule est une généralisation de la formule de Wald, dans le cas où les variables ont une dépendance de type branchement. Rappelons que ce résultat élémentaire nous dit que si $(X_n)_{n\geq 0}$ est une suite de variables aléatoires i.i.d. et N un nombre aléatoire sur $\mathbb{N}^*$ indépendant de cette suite alors

$$\frac{1}{\mathbb{E}[N]}\mathbb{E}\left[\sum_{n=1}^{N} f(X_n)\right] = \mathbb{E}\left[f(X_1)\right],$$

pour toute fonction positive f. L'utilisation des $h-$transformée pour l'étude des processus de branchement date approximativement de la fin des années 80, mais ces formules sont actuellement à la mode dans l'étude des processus branchants [BDMT11, EHK10] et des superprocessus [EW06, Eng09]. Dans le chapitre 5, nous commençons par généraliser ces formules dans le contexte qui nous intéresse puis nous démontrons des théorèmes de convergence à l'aide de celles-ci. Dans le chapitre 6, nous étudions un cas particulier du modèle introduit dans le chapitre 5. Dans la section qui suit, nous introduisons plus précisément cette transformation et explicitons exactement là ou elle intervient dans les différents chapitres. Ensuite, dans la section 4.3, nous introduisons deux exemples simples de modèle de croissance-fragmentation. Ces modèles nous permettront d'introduire simplement les résultats des chapitres 5 et 6.

4.2 Formule de Feynman-Kac et $h-$transformée

Soit $(P_t)_{t\geq 0}$ un semigroupe de Markov, sur un espace polonais E, généré par un opérateur $\mathcal{L}$; c'est-à-dire que

$$\forall t \geq 0,\ \partial_t P_t = \mathcal{L}P_t = P_t\mathcal{L}.$$

S'il existe une fonction harmonique h, c'est-à-dire que

$$\mathcal{L}h = 0,$$

qui est strictement positive, alors il est facile de voir que $(P_t h)_{t\geq 0}$ est une application constante et la famille d'opérateurs $(P_t^h)_{t\geq 0}$, définie par

$$\forall t \geq 0,\ P_t^h f = \frac{P_t(fh)}{P_t h},$$

pour toute fonction positive f, est un semigroupe de Markov. La fonction h est appelée état fondamental, ou *ground state*, et le nouveau semigroupe est appelée $h-$transformée de P. Remarquons que la propriété de Markov nous donne que si $(X_t)_{t\geq 0}$ est un processus de Markov, généré par $\mathcal{L}$, alors $(h(X_t))_{t\geq 0}$ est une martingale. Illustrons maintenant un peu plus précisément le lien entre cette transformation et l'équation de Schrödinger. Soit V une application régulière et $(S_t)_{t\geq 0}$ la famille d'opérateurs, définie pour toute fonction positive f et $t \geq 0$ par

$$S_t f(x) = \mathbb{E}\left[f(X_t)e^{\int_0^t V(X_s)ds} \mid X_0 = x\right].$$

Si V est non nul alors $(S_t)_{t\geq 0}$ n'est pas un semigroupe de Markov. On dit parfois que c'est un semigroupe de Feynman-Kac. Lorsque $V \leq 0$, on parle aussi de semigroupe sous-markovien. En utilisant la formule d'Itô-Dynkin, on montre facilement que, pour toute fonction f suffisamment régulière, on a

$$\forall t \geq 0,\ \partial_t S_t f = (\mathcal{L}+V)S_t f = S_t(\mathcal{L}+V)f.$$

C'est-à-dire que S est solution d'une équation de type Schrödinger. Supposons qu'il existe une fonction $\psi > 0$ suffisamment régulière et un nombre réel λ tel que

$$(\mathcal{L}+V)\psi = \lambda\psi. \tag{4.1}$$

Sous cette condition, la fonction $h : (x,t) \mapsto e^{-\lambda t}\psi(x)$ est harmonique pour l'opérateur $\mathcal{L}+V+\partial_t$, défini sur $E \times \mathbb{R}_+$, et il existe un semigroupe de Markov $(Q_t)_{t\geq 0}$ tel que

$$S_t f = e^{\lambda t}\psi \times Q_t\left(\frac{f}{\psi}\right). \tag{4.2}$$

Ce semigroupe est généré par $\mathcal{A}_\psi$, défini pour toute fonction régulière f par

$$\mathcal{A}_\psi f = \frac{\mathcal{L}(f\psi) - f\mathcal{L}\psi}{\psi}.$$

Dans la suite de cette section, nous décrivons comment nous nous servons de cette transformation dans cette thèse.

4.2.1 Processus de diffusion de Kolmogorov-Langevin

Soit $(X_t)_{t\geq 0}$ le processus, sur $\mathbb{R}$, solution de

$$\forall t \geq 0,\ dX_t = \sqrt{2}dB_t - q'(X_s)ds,$$

où $(B_t)_{t\geq 0}$ est un mouvement brownien standard et q une application régulière vérifiant

$$\int_{\mathbb{R}} e^{-q(x)}dx < +\infty.$$

Dans le chapitre 2, on a montré que, si $(P_t)_{t\geq 0}$ désigne le semigroupe de X, alors on a

$$\nabla P_t f(x) = \mathbb{E}\left[f'(X_t)e^{-\int_0^t q''(X_s)ds} \mid X_0 = x\right],$$

pour toute fonction positive f et $x \in \mathbb{R}$. Si on suppose que

$$\lim_{t\to+\infty} q''(x) = +\infty,$$

alors il est connu qu'il existe une fonction suffisamment régulière et strictement positive ψ et $\lambda > 0$ tel que

$$\psi'' - q'\psi' - q''\psi = -\lambda\psi. \tag{4.3}$$

Voir [BS83, HS96, RS78, Pin95] par exemple. Utilisant une $h-$transformée, on peut donc montrer que

$$\nabla P_t f = e^{-\lambda t}\psi Q_t\left(\frac{f}{\psi}\right),$$

et que donc, $\delta_x P_t$ converge vers son équilibre, en distance de Wasserstein, exponentiellement vite à taux λ. Maintenant, un calcul rapide montre que si ψ vérifie (4.3) alors $x \mapsto e^{q(x)}\psi(-x)$ est un vecteur propre du générateur de X. On en déduit donc que la convergence en Wasserstein se déroule à la même vitesse que celle donnée par le trou spectral. Les détails sont donnés dans le chapitre 2

4.2.2 Processus de Markov modulé

Les preuves du chapitre 3 reposent essentiellement sur le lemme suivant

Lemme 4.2.1 (Exponentielle des fonctionnelles additives des processus de Markov ergodiques)**.** *Supposons que I est une chaîne de Markov récurrente, sur un espace fini F, avec comme mesure invariante ν, et α une fonction sur F vérifiant*

$$\sum_{i\in F} \alpha(i)\nu(i) > 0,$$

alors il existe $q \in (0,1]$, $C, \lambda > 0$ tel que

$$\forall t \geq 0,\ \mathbb{E}\left[e^{-\int_0^t q\alpha(I_s)ds}\right] \leq Ce^{-\lambda t}$$

La démonstration de ce lemme repose sur une $h-$transformée. En effet, l'espace d'état de I étant fini, et I étant irréductible, le théorème de Perron-Frobenius nous donne, pour tout $q > 0$, l'existence d'un couple d'éléments propres (λ_q, ψ_q), tel que ψ_q est strictement positive, associé au problème

$$(Q - q\alpha)\psi_q = -\lambda_q\psi_q,$$

où Q est le générateur de I. En particulier, on a $\lambda_0 = 0$ et $\psi_0 = 1$. En intégrant, on trouve

$$-q \sum_{i \in F} \alpha(i)\psi_q(i)\nu(i) = -\lambda_q \sum_{i \in F} \psi_q(i)\nu(i),$$

puis en dérivant par rapport à q,

$$\sum_{i \in F} (\alpha(i)\psi_q(i) + q\alpha(i)\partial_q\psi_q(i))\nu(i) = \sum_{i \in F} (\partial_q\lambda_q\psi_q(i) + \lambda_q\partial_q\psi_q(i))\nu(i).$$

Maintenant, on prend $q = 0$ pour trouver que

$$\partial_q\lambda_q = \sum_{i \in F} \alpha(i)\nu(i) > 0.$$

En particulier, $q \mapsto \lambda_q$ est croissante prés de l'origine et si q est proche de 0 alors $\lambda_q > 0$. Finalement, pour tout $t \geq 0$, on a

$$\frac{\min \psi_q}{\max \psi_q} e^{-\lambda_q t} \leq \mathbb{E}\left[e^{-\int_0^t q\alpha(I_s)ds}\right] \leq \frac{\max \psi_q}{\min \psi_q} e^{-\lambda_q t}.$$

4.2.3 Formule *many-to-one*

Beaucoup d'auteurs ont utilisé des $h-$transformée dans l'étude des processus de branchement. Beaucoup de références sont données dans l'introduction de l'article [BDMT11] par exemple. Dans mes recherches, j'ai étudié des processus en temps continu indexés par un arbre discret comme dans [Ber06, EHK10]. Dans ce contexte, cette transformation est appelée fragment marqué, décomposition en épine ou formule *many-to-one*. Montrons plus précisément comment cette transformation intervient dans notre modèle. Supposons que

- les particules se déplacent suivant un générateur de Markov G, sur un espace d'état E ;
- les particules se divisent à taux r, en un nombre aléatoire distribué suivant une loi $(p_k)_{k \in \mathbb{N}}$;
- conditionnellement à avoir k enfants, les nouvelles particules sont données par $(F_j^{(k)}(x, \Theta))_{1 \leq j \leq k}$, où x est la position de la particule mère, Θ est une variable aléatoire de loi uniforme sur $[0, 1]$, et $(F_j^{(k)})_{j,k}$ une famille de fonctions mesurables.

Dans ce cas, on peut montrer que la mesure empirique $(\mathbf{Z}_t)_{t \geq 0}$, représentant la population, vérifie

$$\partial_t \mathbf{Z}_t(f) = \partial_t \int_E f(x)\mathbf{Z}_t(dx) = \int_E \mathcal{G}f(x)\mathbf{Z}_t(dx) = \mathbf{Z}_t(\mathcal{G}f),$$

pour toute fonction suffisamment régulière f, où $\mathcal{G}$ est un opérateur décrit par

$$\mathcal{G}f(x) = Gf(x) + r(x)\sum_{k\geq 0} p_k \sum_{j=1}^{k} \int_0^1 f(F_j^{(k)}(x,\theta))d\theta - r(x)f(x),$$

pour toute fonction dans le domaine de G. Par exemple, dans le cas particulier d'un mouvement brownien, se divisant à taux r, en 2 particules, se situant à la même place que leur mère, cet opérateur se réduit à $\Delta + r$. De manière général, il est facile de voir que cet opérateur s'écrit sous la forme d'un opérateur de type Feynman-Kac. Pour pouvoir utiliser une $h-$transformée, il faut donc avoir l'existence d'un vecteur propre positif. Pour l'exemple des processus de fragmentation [Ber06], on voit que les monômes sont des fonctions propres. Pour le cas des diffusions branchantes, on utilise les résultats classiques sur l'équation de Schrödinger [HS96, Pin95, RS78]. Dans le cas général, $\mathcal{G}$ est un opérateur intégro-différentiel et il est difficile de trouver des résultats sur l'existence d'éléments propres. Notons tout de même que lorsque r est constant, on a

$$\mathcal{G}1 = r\sum_{k\geq 0} p_k(k-1),$$

et donc $x \mapsto 1$ est un vecteur propre. On retrouve en particulier que

$$\mathbb{E}\left[N_t\right] = e^{tr\sum_{k\geq 0} p_k(k-1)},$$

où N_t est le nombre de particules au temps $t \geq 0$. Ce résultat est bien connu car N est un processus de branchement [AN04]. Dans notre cadre, nous avons utilisé des résultats récents de [Per07, DG09, BCnG12] pour l'existence d'éléments propres.

4.3 Deux exemples simples de population structurée en taille

Dans cette section, nous décrivons deux exemples simples pour mettre en avant les techniques utilisées dans cette partie. Nous allons supposer que la croissance entre les divisions est déterministe : la taille X de chaque particule est solution de

$$X_t = g(X_t)dt,$$

où g sera la fonction donnée par $g : x \mapsto \mu x$, $\mu \in \mathbb{R}$, ou $g : x \mapsto 1$ est constante. On supposera que le taux de division est constant et que la division est dyadique. Nous supposerons aussi que la taille des

enfants vaut la moitié de la position du parent. Dans ce cas, $\mathcal{G}$ est donné par

$$\mathcal{G}f(x) = g(x)f'(x) + r\left(2f\left(\frac{x}{2}\right) - f(x)\right).$$

Avant de rentrer dans les détails, remarquons que le nombre d'individus N_t, au temps $t \geq 0$, ne dépend pas de la dynamique induite par g. Ce processus est appelée processus de Yule. En utilisant les propriétés élémentaires des lois exponentielles, on peut montrer que N_t suit une loi géométrique de paramètre e^{-rt} et que $N_t e^{-rt}$ converge, presque sûrement et dans L^2, lorsque $t \to +\infty$, vers une variable aléatoire W de loi exponentielle de paramètre 1.

4.3.1 Notations

Ici on rappelle rapidement les notations standards que nous allons utiliser. Pour tenir compte de la généalogie, chaque individu sera étiqueté par un élément

$$u \in \mathcal{T} = \bigcup_{n\geq 0} \{0,1\}^n,$$

avec $\{0,1\}^0 = \{\emptyset\}$. Le label $\emptyset$ est attribué au premier individu, et pour un individu $u = (u_1, ..., u_n) \in \mathcal{T}$, ses enfants sont labellisés par $u1 = (u_1, ..., u_n, 1)$ et $u2 = (u_1, ..., u_n, 2)$. La position de la cellule u est donnée par X^u, sa date de naissance par $b(u)$ et sa date de mort par $d(u)$. La variable aléatoire $d(u) - b(u)$ est distribuée selon une variable aléatoire de loi exponentielle de paramètre r. On note V_t l'ensemble des cellules en vie au temps t :

$$V_t = \{u \in \mathcal{T} \mid b(u) \leq t < d(u)\}.$$

4.3.2 Croissance linéaire : $g = 1$

Commençons par le cas ou g est constante égale à 1. Cet exemple simple va nous permettre de comprendre les résultats du chapitre 5. En utilisant le fait que $\mathcal{G}$ est un opérateur de type Schrödinger et que les fonctions constantes sont vecteurs propres, on montre facilement le lemme suivant

Lemme 4.3.1 (formule *many-to-one*)**.** *Pour toute fonction positive f, on a*

$$\forall t \geq 0, \frac{1}{\mathbb{E}[N_t]}\mathbb{E}\left[\sum_{u\in V_t} f(X_t^u)\right] = \mathbb{E}[f(Y_t)],$$

où Y est un processus de Markov généré par

$$Af(x) = f'(x) + 2r\left(f\left(\frac{x}{2}\right) - f(x)\right).$$

On voit apparaître un phénomène de biais contre-intuitif. Le taux de saut de Y est deux fois plus grand que le taux de division. Ce biais par la taille peut s'expliquer de la manière suivante : le processus Y représente une cellule choisie uniformément au hasard, mais les individus les plus présents dans la population sont ceux issus d'une famille nombreuse, c'est-à-dire pour lesquelles les parents se sont divisés assez vite, donc Y a plus de chance d'être issu de parents qui se sont beaucoup divisés. Dans [BDMT11], la preuve de ce lemme est différente de la notre. Donnons la rapidement.

Preuve du lemme 4.3.1 via la méthode de [BDMT11] . Pour $u = (u_1, .., u_n) \in \mathcal{T}$, on note $|u| = n$. On a

$$\mathbb{E}\left[\sum_{u\in V_t} f(X_t^u)\right] = \sum_{u\in\mathcal{T}}\sum_{q\geq 0} \mathbf{1}_{|u|=q}\mathbb{E}\left[\mathbf{1}_{\{b(u)\leq t<d(u)\}} f(X_t^u)\right].$$

Maintenant, si $|u| = n$ alors X_t^u est une fonction déterministe des temps de division ; c'est-à-dire que

$$X_t^u = \Phi(t, \tau_1, ..., \tau_n),$$

où $\tau_1 = b(u_1) - b(\emptyset), ..., \tau_n = b(u_n) - b(u_{n-1})$, pour une certaine fonction Φ. On peut donc calculer la somme de façon explicite ; plus précisément,

$$\begin{aligned}
\mathbb{E}\left[\sum_{u\in V_t} f(X_t^u)\right] &= \sum_{u\in\mathcal{T}}\sum_{q\geq 0} \mathbf{1}_{|u|=q}\mathbb{E}\left[\mathbf{1}_{\{b(u)\leq t<d(u)\}} f(\Phi(t, \tau_1, ..., \tau_q))\right] \\
&= \sum_{u\in\mathcal{T}}\sum_{q\geq 0} \mathbf{1}_{|u|=q}\int_{\mathbb{R}_+^{q+1}} r^{q+1} e^{-r\sum_1^{q+1} t_i} \\
&\qquad \mathbf{1}_{\{\sum_1^q t_i\leq t<\sum_1^{q+1} t_i\}} f(\Phi(t, t_1, ..., t_q))dt_1...dt_{q+1} \\
&= \sum_{q\geq 0}\sum_{u\in\mathcal{T}} \mathbf{1}_{|u|=q}\int_{\mathbb{R}_+^{q}} r^{q} e^{-rt} \mathbf{1}_{\{\sum_1^q t_i\leq t\}} f(\Phi(t, t_1, ..., t_q))dt_1...dt_q \\
&= \sum_{q\geq 0}\int_{\mathbb{R}_+^{q}} (2r)^{q} e^{-rt} \mathbf{1}_{\{\sum_1^q t_i\leq t\}} f(\Phi(t, t_1, ..., t_q))dt_1...dt_q \\
&= e^{rt}\sum_{q\geq 0}\int_{\mathbb{R}_+^{q}} (2r)^{q} e^{-2rt} \mathbf{1}_{\{\sum_1^q t_i\leq t\}} f(\Phi(t, t_1, ..., t_q))dt_1...dt_q \\
&= e^{rt}\mathbb{E}\left[f(Y_t)\right].
\end{aligned}$$

On conclut la preuve en utilisant le fait que $\mathbb{E}[N_t] = e^{rt}$. □

Lorsque r n'est pas constant, cette démonstration ne marche pas au contraire de la démonstration utilisant une $h-$transformée. Par exemple si $r(x) = x$ alors $\psi : x \mapsto x + 1$ est vecteur propre et on peut démontrer la formule *many-to-one* à poids suivante

$$\frac{1}{\mathbb{E}\left[\sum_{u\in V_t}(1+X_t^u)\right]}\mathbb{E}\left[\sum_{u\in V_t} f(X_t^u)(1+X_t^u)\right] = \mathbb{E}\left[f(Y_t)\right],$$

où Y est généré par

$$Af(x) = f'(x) + \frac{x(x+2)}{(x+1)}\left(f\left(\frac{x}{2}\right) - f(x)\right).$$

On remarque que le biais est différent. Lorsque r n'est pas constant le biais n'est pas seulement présent dans le taux de saut. Le mouvement entre les sauts peut aussi être biaisé. Lorsque r varie, on va montrer, dans le prochain chapitre, le théorème suivant :

Théorème 4.3.2 (Loi des grands nombres)**.** *Si r est continu, $0 < r < \bar{r}$ et $r = \bar{r}$ hors d'un compact alors il existe une mesure π tel que*

$$\lim_{t\to+\infty}\frac{1}{N_t}\sum_{u\in V_t} g(X_t^u) = \int g d\pi,$$

en probabilité, pour toute fonction continue et bornée g.

Notons que lorsque r est constant, la mesure π est explicite. La démonstration est basée sur le lemme précédent et une sorte de propagation du chaos. C'est-à-dire que si le nombre de particules est grand alors elles sont presque indépendantes.

Remarque 4.3.3 (Unicité des vecteurs propres)**.** *Notons que sous les hypothèses du théorème précédent, l'opérateur $\mathcal{G}$ admet un unique vecteur propre positif, à multiplication par une constante prés.*

4.3.3 Croissance exponentielle : $g(x) = \mu x$

Lorsque $g(x) = \mu x$, le générateur $\mathcal{G}$ admet une infinité de vecteurs propres positifs. En effet, comme les processus de fragmentation, cet opérateur admet les monômes $x \mapsto x^\alpha$, $\alpha \geq 0$, comme vecteurs propres. Cependant les valeurs propres ne suivent pas le même comportement que dans le cas des processus de fragmentation. En utilisant cette famille de vecteurs propres, on a une famille de martingales. Puis en utilisant des inégalités du type

$$\max_{u\in V_t}(X_t^u)^\alpha \leq \sum_{u\in V_t}(X_t^u)^\alpha$$

ou

$$\max_{u\in V_t}(X_t^u)^\alpha \geq \frac{\sum_{u\in V_t}(X_t^u)^\beta}{\sum_{u\in V_t},(X_t^u)^{\beta-\alpha}}$$

pour tout $\beta \geq \alpha \geq 0$, on trouve le comportement de la taille de la plus grande cellule dans le chapitre 6.

Chapter 5

Limit theorems for some branching measure-valued processes

5.1 Introduction

In this work, we study the evolution of a Markov process indexed by a tree in continuous time. The tree can represent a population of cells, polymers or particles. On this population, we consider the evolution of an individual characteristic. This characteristic can represent the size, the age or the rate of a nutriment. During the life of an individual, its characteristic evolves according to an underlying Markov process. At non-homogeneous time, the individuals die and divide. When one divides, the characteristics of the offspring depend on those and their number. This model was studied in [ABBS11, BDMT11, BT11, Ber06, EHK10, HW96]. Here, we study the asymptotic behaviour of the empirical measure which describes the population. Following [BDMT11], we begin to prove a many-to-one formula (or spinal decomposition, tagged fragment ...) and then deduce its long time behaviour. This formula looks like the Wald formula and reduces the problem to the study of a "typical" individual. Closely related, we can find a limit theorem in discrete time in [DM10], in continuous time with a continuous population in [EW06] and for a space-structured population model in [EHK10]. Our approach is closer to [BDMT11] and extends their law of large number to a variable rate of division. This extension is essential in application [BT11]. In our model, the population is discrete. It is the microscopic version of some deterministic equations studied in [LP09, Per07, PR05]. Following [FM04, Tra06], we scale our empirical measure and prove that these P.D.E. are macroscopic versions of our model. Before expressing our main results, we begin by giving some notations. If we start with

one individual then we will use the Ulam-Harris-Neveu notation [BDMT11]:

- the first individual is labelled by $\emptyset$;
- when the individual u divides, then his K descendants are labelled by $u1, ..., uK$;
- we denote by $\mathcal{T}$ the random set of individuals which are dead, alive or will be alive;
- it is a subset of $\mathcal{U} = \cup_{m\geq 0} (\mathbb{N}^*)^m$, where $\mathbb{N} = \{0, 1, ...\}$ and $(\mathbb{N}^*)^0 = \{\emptyset\}$;
- we denote by $\mathcal{V}_t$ the set of individuals which are alive at time t;
- for each $u \in \mathcal{T}$, $\alpha(u)$ and $\beta(u)$ denote respectively the birth and the death date of the individual u;
- we denote by N_t the number of individuals alive at time t;
- for each $u \in \mathcal{T}$ and $t \in [\alpha(u), \beta(u))$, the characteristic of the individual u is denoted by X_t^u.

The dynamics of our model is then as follows.

- The characteristic of the first individual, $(X_t^\emptyset)_{t\in[0,\beta(\emptyset))}$ is distributed according to an underlying càdlàg strong Markov process $(X_t)_{t\geq 0}$. For sake of simplicity, we will assume that $X = (X_t)_{t\geq 0}$ is a Feller process, takes values in an open subset E of $\mathbb{R}^d$ and is generated by

$$Gf(x) = b(x) \cdot \nabla f(x) + \sigma \Delta f(x), \tag{5.1}$$

  for every $f \in C_c^2(E)$, where $d \in \mathbb{N}^*$, $b : \mathbb{R}^d \to \mathbb{R}^d$ is a C^∞ function and $\sigma \in \mathbb{R}_+$. Here, $C_c^2(E)$ is the set of C^2 functions with compact support.
- The death time $\beta(\emptyset)$ of the first individual verifies

$$\mathbb{P}\left(\beta(\emptyset) > t \mid X_s^\emptyset, s \leq t\right) = \int_0^t r(X_s^\emptyset) ds,$$

  where r is a non negative, measurable and locally bounded function. Notice that $\alpha(\emptyset) = 0$.
- At time $\beta(\emptyset)$, the first individual splits into a random number of children given by an independent random variable K of law $(p_k)_{k\in\mathbb{N}^*}$. We have $\alpha(0) = ... = \alpha(K-1) = \beta(\emptyset)$.
- We assume that the mean offspring number, which is defined by $m : x \mapsto \sum_{k\geq 0} k p_k(x)$, is locally bounded on E.
- The characteristics of the new individuals are given by $(F_j^{(K)}(X_{\beta(\emptyset)-}^\emptyset, \Theta))_{1\leq j\leq K}$, where Θ is a uniform variable on $[0, 1]$. The sequence $(F_j^{(k)})_{j\leq k, k\in\mathbb{N}^*}$ is supposed to be a family of measurable functions.
- Finally, the children evolve independently from each other like the first individual.

The last point is the branching property. To obtain a limit theorem, we follow the approach of [BDMT11]. In this paper, the cell's death rate r and the law of the number of descendants $(p_k)_{k\geq 1}$ are

constant. A many-to-one formula is proved: for every continuous and bounded function f, we have

$$\frac{1}{\mathbb{E}[N_t]}\mathbb{E}\left[\sum_{u\in\mathcal{V}_t} f(X_t^u)\right] = \mathbb{E}[f(Y_t)], \tag{5.2}$$

where Y is generated, for any $f \in C_c^2(E)$ and $x \in E$, by

$$A_0 f(x) = Gf(x) + rm\sum_{k\geq 1}\frac{kp_k}{m}\int_0^1\left(\frac{1}{k}\sum_{j=1}^{k} f(F_j^{(k)}(x,\theta)) - f(x)\right)d\theta. \tag{5.3}$$

This process evolves as X, until it jumps, at an exponential time with mean $1/rm$. We observe that r is not the jump rate of the auxiliary process. There is a biased phenomenon. It is described in [BDMT11, HW96] and their references. We can interpret it by the fact that the faster the cells divide, the more descendants they have. That is why a uniformly chosen individual has an accelerated rate of division. A possible generalisation of (5.2) is a Feynman-Kac formula as in [HW96]: for every continuous and bounded function f, we have

$$\mathbb{E}\left[\sum_{u\in\mathcal{V}_t} f(X_t^u)\right] = \mathbb{E}\left[f(Y_t)e^{\int_0^t r(Y_s)(m(Y_s)-1)ds}\right],$$

where Y is an auxiliary process starting from x_0 and generated by (5.3). Using a Poisson point process, in [BT11], we get also another representation of the empirical measure to prove the extinction of a parasite population. However, it is difficult to exploit these formulas. Inspired by [EW06, LP09, Per07, PR05], we follow an alternative approach. In (5.2), Y can be understood as a uniformly chosen individual. The problem is: if r is not constant then a uniformly chosen individual does not follow Markovian dynamics. Our solution is to choose this individual with an appropriate weight. This weight is the eigenvector V of the operator $\mathcal{G}$ be defined, for every $f \in C_c^2(E)$ and $x \in E$, by

$$\mathcal{G}f(x) = Gf(x) + r(x)\left[\left(\sum_{k\geq 0}\sum_{j=1}^{k}\int_0^1 f(F_j^{(k)}(x,\theta))d\theta p_k(x)\right) - f(x)\right].$$

It is not the generator of a Markov process on E. It is described in the next section. Under some assumptions, we are able to prove that the following weighted many-to-one formula holds for every continuous and bounded function f:

$$\frac{1}{\mathbb{E}[\sum_{u\in\mathcal{V}_t} V(X_t^u)]}\mathbb{E}\left[\sum_{u\in\mathcal{V}_t} f(X_t^u)V(X_t^u)\right] = \mathbb{E}[f(Y_t)], \tag{5.4}$$

where Y is an auxiliary Markov process, starting from x_0. It is generated by $A = M + J$, where M describes the motion between the jumps and is defined by

$$Mf(x) = \frac{G(f \times V)(x) - f(x)GV(x)}{V(x)} = Gf(x) + 2\sigma \frac{\nabla V(x).\nabla f(x)}{V(x)},$$

and J describes the jump dynamics and is given by

$$Jf(x) = \Lambda(x) \left[\frac{\sum_{k\in\mathbb{N}} \sum_{j=1}^{k} \int_0^1 V\left(F_j^{(k)}(x,\theta)\right) f\left(F_j^{(k)}(x,\theta)\right) d\theta p_k(x)}{\sum_{k\in\mathbb{N}} \sum_{j=1}^{k} \int_0^1 V\left(F_j^{(k)}(x,\theta)\right) d\theta p_k(x)} - f(x) \right],$$

where

$$\Lambda(x) = \left[\sum_{k\in\mathbb{N}} \sum_{j=1}^{k} \int_0^1 V\left(F_j^{(k)}(x,\theta)\right) \, d\theta \, p_k(x) \right] \times \frac{r(x)}{V(x)},$$

for every $f \in C_c^2(E)$ and $x \in E$. These formulas seem to be complicated but they are very simple when applied. We also observe a biased phenomenon. But contrary to the previous formulas, the bias is present in the motion and the branching mechanism. This bias has been already observed in another context [EW06]. Also note that we do not assume that λ_0 is the first eigenvalue. So, it is possible to have different many-to-one formulas as can be seen in Remark 5.3.8. We can find some criteria for existence of eigenelements in [BCnG12, DG09, Mic06, Pin95] and theirs references. If Y is ergodic with invariant measure π then Formula (5.4) gives

$$\lim_{t\to+\infty} \frac{1}{\mathbb{E}\left[\sum_{u\in\mathcal{V}_t} V(X_t^u)\right]} \mathbb{E}\left[\sum_{u\in\mathcal{V}_t} f(X_t^u) V(X_t^u) \right] = \int f \, d\pi,$$

for all continuous and bounded function f. We improve this result:

Theorem 5.1.1 (Long time behaviour of the empirical measure)**.** *If the following assumptions holds,*

- $X_0^{\emptyset}$ *is deterministic;*
- *the system is non explosive; namely* $N_t < +\infty$ *a.s. for all* $t \geq 0$*;*
- *there exists* (V, λ_0) *such that* $\mathcal{G}V = \lambda_0 V$*,* $V > 0$ *and* V *is "smooth" (see Assumption 5.3.1);*
- Y *is a Feller process and is ergodic with invariant measure* π*;*

then for any continuous function g *such that:*

- *there exists* $C > 0$*, such that for all* $x \in E$*,* $|g(x)| \leq CV(x)$*;*

– *there exists* $\alpha < \lambda_0$*, such that* $\mathbb{E}[V^2(Y_t)] \leq Ce^{\alpha t}$ *and*

$$\mathbb{E}\left[\frac{r(Y_t)}{V(Y_t)}\int_0^1 \sum_{\substack{a,b\in\mathbb{N}^* \\ a\neq b}} \sum_{k\geq \max(a,b)} p_k(Y_s)V\left(F_a^{(k)}(Y_s,\theta)\right)V\left(F_b^{(k)}(Y_s,\theta)\right)d\theta\right] \leq Ce^{\alpha t};$$

then we have

$$\lim_{t\to+\infty} e^{-\lambda_0 t}\sum_{u\in\mathcal{V}_t} g(X_t^u) = W\int \frac{g}{V}\,d\pi,$$

where $W = \lim_{t\to+\infty} e^{-\lambda_0 t}V(x_0)^{-1}\sum_{u\in\mathcal{V}_t} V(X_t^u)$ *and the convergences hold in probability. If furthermore* V *is lower bounded by a positive constant then*

$$\lim_{t\to+\infty} \frac{\mathbf{1}_{W\neq 0}}{N_t}\sum_{u\in\mathcal{V}_t} g(X_t^u) = \mathbf{1}_{W\neq 0}\int \frac{g}{V}\,d\pi \Big/ \int \frac{1}{V}\,d\pi \text{ in probability .}$$

If r and $(p_k)_{k\in\mathbb{N}}$ are constant then $V \equiv 1$ is an eigenvector, and so this theorem generalises [BDMT11, Theorem 1.1]. On the other hand, our model is microscopic and is a scaled version of some deterministic models. More precisely, let $(\mathbf{Z}_t)_{t\geq 0}$ be the empirical measure. It is defined, for all $t \geq 0$, by

$$\mathbf{Z}_t = \sum_{u\in\mathcal{V}_t} \delta_{X_t^u}.$$

Now, let $\mathbf{Z}^{(n)}$ be distributed as $\mathbf{Z}$ and let us consider the following scaling $\mathbf{X}^{(n)} = \frac{1}{n}\mathbf{Z}^{(n)}$. We have:

Theorem 5.1.2 (Law of large number for the large population)**.** *If the following assumptions hold*

– $T > 0$*;*
– r *is upper bounded;*
– *there exist* $\bar{k} \geq 0$ *such that* $p_k \equiv 0$ *for all* $k \geq \bar{k}$*;*
– *either* $E \subset \mathbb{R}$ *and* $F_j^{(k)}(x,\theta) \leq x$ *for all* $j \leq k$ *and* $\theta \in [0,1]$ *or* E *is compact;*
– *the equation* (5.5) *below admits almost one solution.*
– *The starting distribution* $\mathbf{X}_0^{(n)}$ *converges in distribution to* $\mathbf{X}_0 \in \mathcal{M}(E)$*, embedded with the weak topology;*
– *we have*

$$\sup_{n\geq 0}\mathbb{E}\left[\mathbf{X}^{(n)}(E)\right] < +\infty.$$

then $\mathbf{X}^{(n)}$ *converges in distribution in* $\mathbb{D}([0,T],\mathcal{M}(E))$ *to* $\mathbf{X}$ *which verifies*

$$\int_E f(x)\,\mathbf{X}_t(dx) = \int_E f(x)\,\mathbf{X}_0(dx) + \int_0^t\int_E \mathcal{G}f(x)\,\mathbf{X}_s(dx)ds. \tag{5.5}$$

Here, $\mathbb{D}([0,T],\mathcal{M}(E))$ is the space of càd-làg functions embedded with the skohorod topology [Bil99, JM86]. We observe that if $\mathbf{X}_0$ is deterministic then $\mathbf{X}_t$ is deterministic for any time $t \geq 0$. The equation (5.5) can be written as

$$\partial_t n(t,x) + \nabla\left(b(x)n(t,x)\right) + r(x)n(t,x) = \sigma\partial_{xx}n(t,x) + \sum_{k\geq 1}\sum_{j=1}^{k} K_j^k\left(r \times p_k \times n(t,\cdot)\right).$$

where $\mathbf{X}_t = n(t,x)dx$ and K_j^k is the adjoint operator of $f \mapsto \int_0^1 f(F_j^{(k)}(x,\theta)d\theta$. This equation was studied in [LP09, Per07, PR05] and Theorem 5.1.1 is relatively close to their limit theorems. We will see in the next section that it is also the Kolmogorov equation associated to $\mathbf{Z}$. So, we observe that $\mathbf{X}$ is equal to the mean measure of $\mathbf{Z}$; that is $f \mapsto \mathbb{E}[\int_E f(x)\ \mathbf{Z}_t(dx)]$. This average phenomenon comes from the branching property. After a branching event, each cell evolves independently from each other, there is not interaction or mutation. Another reason is the linearity of the operator $\mathcal{G}$. From the many-to-one formula, we also deduce that, in large population, the empirical measure behaves as the auxiliary process. The proof is based on the Aldous-Rebolledo criterion [JM86, RC86] and is inspired by [FM04, MT12, Tra06].

In the end of the paper, these two theorems are applied to some structured population models. Our main example is a size-structured population. In this example, the size of cells grows linearly and if a cell dies then it divides to two descendants. Thus, there is motion between the branching events and discontinuity during division. This model is a branching version of the well known TCP windows size process [CMP10, GRZ04, LvL08, OKM96]. For this example, we are able to give some explicit formulas of the invariant distribution, the moments or the rate of convergence. We also prove that, in large population, the empirical measure behaves according to the deterministic equation (5.5) plus a Gaussian noise.

Outline. In the next section, we introduce some properties of the empirical measure. In Section 5.3, we focus our interest on the long time behaviour. We prove some many-to-one formulas and deduce a general limit theorem which implies Theorem 5.1.1. Section 5.4 is devoted to the study of large populations. In this one, we prove Theorem 5.1.2. Note that Section 5.3 and Section 5.4 are independent. In Section 5.5, we give our main example, which describes the cell mitosis. Moreover, we give two theorems for the long time behaviour of our empirical measure in addition to some explicit formulas. We also give a central limit theorem for asymmetric cell division for the macroscopic limit. In section 5.6, we finish by two classical examples which are branching diffusions and self-similar fragmentation.

5.2 Preliminaries

In this section, we describe a little more the empirical measure $(\mathbf{Z}_t)_{t\geq 0}$. We recall that

$$\forall t \geq 0,\ \mathbf{Z}_t = \sum_{u\in\mathcal{V}_t} \delta_{X_t^u}.$$

It belongs to the space $\mathbb{D}(\mathbb{R}_+, \mathcal{M}(E))$ of càd-làg functions with values in $\mathcal{M}(E)$, which is the set of finite measures on E. Let us add the following notations:

$$\mathbf{Z}_t(f) = \int_E f(x)\mathbf{Z}_t(dx) = \sum_{u\in\mathcal{V}_t} f(X_t^u),$$

for every continuous and bounded function f and

$$\mathbf{Z}_t(1+x^p) = \int_E 1 + x^p\ \mathbf{Z}_t(dx) = \sum_{u\in\mathcal{V}_t} 1 + (X_t^u)^p.$$

We can describe the dynamics of the population with a stochastic differential equation. Let $C_c^{2,1}(E,\mathbb{R}_+)$ be the set of functions $f:(x,t)\mapsto f(x,t)=f_t(x)$ that are C^1 in time, with bounded derivative, such that $f_t \in C_c^2(E)$. For any function f belonging to $C_c^{1,2}(E\times\mathbb{R}_+)$, we have

$$\begin{aligned}\mathbf{Z}_t(f_t) =& \mathbf{Z}_0(f_0) + \int_0^t \int_E \mathcal{G}f_s(x) + \partial_t f_s(x)\mathbf{Z}_s(dx)ds \\ &+ \int_0^t \sum_{u\in V_s} \sqrt{2\sigma}\partial_x f_s(X_s^u)dB_s^u \\ &+ \int_0^t \int_{\mathcal{U}\times\mathbb{R}_+\times\mathbb{N}^*\times[0,1]} [\mathbf{1}_{\{u\in V_{s-}, l\leq r(X_{s-}^u)\}} \\ &\quad (\sum_{j=1}^k f_s(F_j^{(k)}(X_{s-}^u,\theta)) - f_s(X_{s-}^u))]\ \rho(ds,du,dl,dk,d\theta),\end{aligned}$$

where $(B^u)_{u\in\mathcal{U}}$ is a family of independent standard Brownian motions and $\rho(ds,du,dl,dk,d\theta)$ is Poisson point process on $\mathbb{R}_+\times\mathcal{U}\times\mathbb{R}_+\times\mathbb{N}^*\times[0,1]$ of intensity

$$\bar{\rho}(ds,du,dl,dk,d\theta) = ds\ n(du)\ dl\ dp_k\ d\theta.$$

It is also independent from the Brownian motions. We have denoted by $n(du)$ the counting measure on $\mathcal{U}$ and $ds,\ dl,\ d\theta$ are Lebesgue measures. A necessary and sufficient condition for the existence of our process is the non-explosion of $\mathbf{Z}$:

Assumption 5.2.1 (Non explosion)**.** *For all $t\geq 0$, $N_t < +\infty$ a.s..*

For instance, we have

Lemma 5.2.2 (Sufficient condition to non explosion)**.** *If $r \leq \bar{r}$ and $p_k = 0$ for all $k \geq \bar{k}$, where $\bar{r}, \bar{k} > 0$, then Assumption 5.2.1 holds. Moreover, for any $T > 0$, we have*

$$\forall t \leq T,\ \mathbb{E}[N_t] \leq \mathbb{E}[N_0]\, e^{(\bar{k}-1)\bar{r}T}.$$

Proof. In this case, we can bound N_t by a branching process independent of the underlying dynamics. □

Lemma 5.2.3 (Semi-martingale Decomposition)**.** *If Assumption 5.2.1 holds, then for all bounded $f = (f_t)_{t\geq 0} \in C_c^{2,1}(E, \mathbb{R}_+)$ and $t \geq 0$, we have*

$$\mathbf{Z}_t(f_t) = \mathbf{Z}_0(f_0) + \mathbf{M}_t(f) + \mathbf{V}_t(f)$$

where

$$\mathbf{V}_t(f) = \int_0^t \mathbf{Z}_s(\mathcal{G} f_s + \partial_s f_s) ds,$$

and the bracket of $\mathbf{M}_t(f)$ is given by

$$\begin{aligned}\langle \mathbf{M}(f)\rangle_t = &\int_0^t G(f_s^2)(x) - 2 f_s(x) G f_s(x) \mathbf{Z}_s(dx) \\ &+ \int_E r(x) \int_0^1 \sum_{k\in\mathbb{N}^*} \left(\sum_{j=1}^k f_s(F_j^{(k)}(x,\theta)) - f_s(x) \right)^2 p_k(x) d\theta \mathbf{Z}_s(dx) ds\end{aligned}$$

Proof. It is an application of Dynkin and Itô formulas, see for instance [JS03, Lemma 3.68 p .487] and [IW89, Theorem 5.1, p.67]. □

We define the mean measure $(\mathbf{z}_t)_{t\geq 0}$, for any continuous and bounded function f on E, by

$$\forall t \geq 0,\ \mathbf{z}_t(f) = \mathbb{E}(\mathbf{Z}_t(f)) = \mathbb{E}\left[\sum_{u\in\mathcal{V}_t} f(X_t^u) \right].$$

Corollary 5.2.4 (Evolution equation for the mean measure)**.** *Under Assumption 5.2.1, if $f \in C_c^2(E)$ and $t \geq 0$ then we have:*

$$\mathbf{z}_t(f) = \mathbf{z}_0(f) + \int_0^t \mathbf{z}_s(Gf) + \int_E r(x) \sum_{k\geq 1} \sum_{j=1}^k \int_0^1 f\left(F_j^{(k)}(x,\theta)\right) d\theta\, p_k(x) - f(x)\, \mathbf{z}_s(dx) ds.$$

The previous equation can be written as

$$\partial_t n(t,x) + \nabla\left(b(x)n(t,x)\right) + r(x)n(t,x) = \sigma\partial_{xx}n(t,x) + \sum_{k\geq 1}\sum_{j=1}^{k} K_j^k\left(r \times p_k \times n(t,\cdot)\right).$$

where $\mathbf{z}_t = n(t,x)dx$ and K_j^k is the adjoint of $f \mapsto \int_0^1 f(F_j^{(k)}(x,\theta)d\theta$.

5.3 Long time's behaviour

Let us recall that

$$\mathcal{G}f(x) = Gf(x) + r(x)\left[\sum_{k\geq 0}\sum_{j=1}^{k}\int_0^1 f(F_j^{(k)}(x,\theta))\,d\theta\; p_k(x) - f(x)\right],$$

for every $f \in C_c^2(E)$ and $x \in E$. In the following, we will prove some formulas which characterise the mean behaviour of our model. Then we will use them to prove our limit theorems.

5.3.1 Eigenelements and auxiliary process

As said in introduction, the existence of eigenelements is fundamental in our approach. Henceforth, we assume the following.

Assumption 5.3.1 (Existence of eigenelements)**.** *Assumption 5.2.1 holds, and there exist $\lambda_0 > 0$ and a measurable and positive function V such that there exists a sequence $(V_n)_{n\geq 0}$ of functions belonging to $C_c^2(E)$ verifying, for all $x \in E$,*

$$\lim_{n\to\infty} V_n(x) = V(x)$$

$$\lim_{n\to\infty} \mathcal{G}V_n(x) = \lambda_0 V(x)$$

and the mappings $x \mapsto \sup_{n\geq 0} V_n(x)$ and $x \mapsto \sup_{n\geq 0} \mathcal{G}V_n(x)$ are integrable with respect to $\mathbf{z}_t$, for every $t \geq 0$.

The integrability condition is essentially a consequence that V is an eigenfunction. This can be proved using a suitable sequence of stopping times. Under this assumption, we introduce the martingale $(\mathbf{Z}_t(V)e^{-\lambda_0 t})_{t\geq 0}$ which plays an important role in the proof of theorem 5.1.1.

Lemma 5.3.2 (Martingale properties)**.** *If Assumption 5.3.1 holds and*

$$\mathbf{z}_0(V) < +\infty,$$

then the process $(\mathbf{Z}_t(V)e^{-\lambda_0 t})_{t\geq 0}$ *is a martingale. Moreover, it converges almost surely to a random variable* W.

Proof. First, by corollary 5.2.4 and dominated convergence Theorem, we have

$$\begin{aligned}\mathbf{z}_t(V) &= \mathbf{z}_0(V) + \int_0^t \mathbf{z}_s(\mathcal{G}V)ds \\ &= \mathbf{z}_0(V) + \lambda_0 \int_0^t \mathbf{z}_s(V)ds.\end{aligned}$$

Hence for all $t \geq 0$, we have $\mathbf{z}_t(V) = \mathbf{z}_0(V)e^{\lambda_0 t}$. Then if $\mathcal{F}_t = \sigma\{Z_s \mid s \leq t\}$ then the Markov properties, applied on $\mathbf{Z}$, gives

$$\mathbb{E}[\mathbf{Z}_{t+s}(V)|\mathcal{F}_s] = \mathbb{E}[\widetilde{\mathbf{Z}}_t(V)|\widetilde{\mathbf{Z}}_0 = \mathbf{Z}_s],$$

where $\widetilde{\mathbf{Z}}$ is distributed as $\mathbf{Z}$. Then $\mathbb{E}[\mathbf{Z}_{t+s}(V)|\mathcal{F}_s] = \mathbf{Z}_s(V)e^{\lambda_0 t}$ and thus

$$\mathbb{E}[\mathbf{Z}_{t+s}(V)e^{-\lambda_0(t+s)} \mid \mathcal{F}_s] = \mathbf{Z}_s(V)e^{\lambda_0 s}.$$

Since $(\mathbf{Z}_t(V)e^{-\lambda_0 t})_{t\geq 0}$ is a positive martingale, it converges almost surely. □

To have our many-to-one formula, we add the following natural assumption:

Assumption 5.3.3 (Auxiliary process)**.** *The generator* A*, be defined in the introduction, generates a Feller process.*

Lemma 5.3.4 (Weighted many-to-one formula)**.** *Under Assumptions 5.3.1 and 5.3.3, if* $\mathbf{Z}_0 = \delta_{x_0}$*, where* $x_0 \in E$*, then we have*

$$\frac{1}{\mathbb{E}\left[\sum_{u\in\mathcal{V}_t} V(X_t^u)\right]}\mathbb{E}\left[\sum_{u\in\mathcal{V}_t} V(X_t^u)f(X_t^u,t)\right] = \mathbb{E}[f(Y_t,t) \mid Y_0 = x_0], \tag{5.6}$$

for any non negative function f *on* $E\times\mathbb{R}_+$ *and* $t \geq 0$*, where* Y *is a Markov process generated by* A *starting from* x_0.

Proof. If $\gamma_t : f \mapsto \mathbf{z}_t(f\times V)e^{-\lambda_0 t}V(x_0)^{-1}$ then, for all $t \geq 0$ and $f \in C^{2,1}(E\times R_+)$, we have

$$\partial_t\gamma_t(f) = \mathbf{z}_t(\mathcal{G}(Vf) + V\partial_t f - f\mathcal{G}V)e^{-\lambda_0 t}V(x_0)^{-1} = \gamma_t(Af + \partial_t f).$$

Now, by Dynkin formula, the right hand side of (5.6) verifies the same equation. Uniqueness comes from classical arguments; see [Kol11, Theorem 4.1.2] and [Kol11, Theorem 3.6.6]. □

Remark 5.3.5 (The first characteristic can be random)**.** *If* $\mathbf{Z}_0 = \delta_{X_0^\emptyset}$*, where* $X_0^\emptyset$ *is random and distributed according to a probability measure* μ*, then* (5.6) *holds, where* Y *starts form* Y_0 *which is distributed according to* μ*.*

Remark 5.3.6 (Schrödinger operator and $h-$transform)**.** *The operator* $\mathcal{G}$ *is not a Markov generator. As we have, for all* $f \in C_c^2(E)$*,*

$$\mathcal{G}f = Bf - r(m-1)f,$$

where B *is a Markov generator and* $\mathcal{G}$ *a so-called a Schrödinger operator. Its study is connected to the Feynman-Kac formula. The key point of our weighted many-to-one formula is an* $h-$*transform (Girsanov type transformation) of the Feynman-Kac semigroup as in [Pin95]. This transformation is usual in the superprocesses study [EW06].*

Remark 5.3.7 (Galton-Watson tree and Malthus parameter)**.** *If* r *and* p *are constant, then* $V \equiv 1$ *is an eigenvector with respect to the eigenvalue* $\lambda_0 = r(m-1)$*, where* $m = \sum_{k\geq 0} kp_k$ *denotes the mean offspring number. So,* $\mathbf{Z}_t(V) = N_t$ *and the population grows exponentially. This result is already know for* N_t*. It is a continuous branching process [AN04, BDMT11]. Furthermore, since Thomas Malthus (1766-1834) has introduced the following simple model to describe the population evolution:*

$$\partial_t N_t = \textit{ birth} - \textit{death} = bN_t - dN_t = \lambda_0 N_t \implies N_t = e^{\lambda_0 t},$$

in biology and genetic population study, λ_0 *is sometimes called the Malthus parameter.*

Remark 5.3.8 (Many eigenelements are possible!)**.** *In the previous lemmas,* λ_0 *was not required to be the first eigenvalue. So, it is possible to have different eigenelements and auxiliary processes. Consider the example of [BT11], where some eigenelements are explicit; that is:*

$$Gf(x) = axf'(x) + bxf''(x),$$

for every $f \in C_c^2(E)$ *and* $x \in E = \mathbb{R}_+$*, where* a, b *are two non-negative numbers. We also consider that* $p_2 = 2$ *and for all* $j \in \{1, 2\}$*,*

$$\mathbb{E}[f(F_j^{(2)}(x, \Theta))] = \mathbb{E}\left[f(Hx)\right],$$

where H *is a symmetric random variable on* $[0, 1]$ *i.e.* $H \overset{d}{=} 1 - H$*. This example models cell division with parasite infection. In this case,*

$$\mathcal{G}f(x) = axf'(x) + bxf''(x) + r(x)\left(2\mathbb{E}[f(Hx)] - f(x)\right),$$

for every continuous and bounded function f, where a is an eigenvalue of $\mathcal{G}$ and $V(x) = x$ is its eigenvector. So, we should have

$$\mathbb{E}\left[\sum_{u\in\mathcal{V}_t} X_t^u f(X_t^u)\right] = \mathbb{E}[f(Y_t)]e^{at}x_0,$$

for every continuous and bounded function f, where Y is a Markov process generated by G_Y be defined by

$$G_Y f(x) = (ax + 2b)\, f'(x) + bxf''(x) + r(x)\,(2\mathbb{E}[Hf(Hx)] - f(x))\,,$$

for every $f \in C_c^2(E)$ and $x \in E$. We can see a bias in the drift terms and jumps mechanism which is not observed in [BDMT11, BT11]. When r is affine, we obtain a second formula. Indeed, if $r(x) = cx + d$, with $c \geq 0$ and $d > a$ (or $d > 0$ and $c = 0$) then $V_1(x) = x(c/(d-a)) + 1$ is an eigenvector with respect to the eigenvalue $\lambda_1 = d$ ($\Rightarrow \lambda_1 > \lambda_0 = a$). Thus, we should also write

$$\mathbb{E}\left[\sum_{u\in\mathcal{V}_t} f(X_t^u)\right] e^{-dt} = \mathbb{E}\left[\frac{f(U_t)}{\tau U_t + 1}\right](\tau x_0 + 1),$$

for every continuous and bounded function f, where $\tau = \frac{c}{d-a}$ and U is generated by G_U be defined, for every $f \in C_c^2(E)$ and $x \in E$, by

$$G_U f(x) = \left(ax + \frac{2bx\tau}{\tau x + 1}\right) f'(x) + bxf''(x) + \frac{r(x)(\tau x + 2)}{\tau x + 1}\left(\frac{2\mathbb{E}[(\tau Hx + 1)f(Hx)]}{\tau x + 2} - f(x)\right).$$

5.3.2 Many-to-one formulas

In order to compute our limit theorem, we need to control the second moment. As in [BDMT11], we begin by describing the population over the whole tree. Then we give a many-to-one formula for forks. Let $\mathcal{T}$ be the random set representing cells that have lived at a certain moment. It is defined by

$$\mathcal{T} = \{u \in \mathcal{U} \mid \exists t > 0, u \in \mathcal{V}_t\}.$$

Lemmas 5.3.9 and 5.3.10, that follow, are respectively the generalisation of [BDMT11, proposition 3.5] and [BDMT11, proposition 3.9].

Lemma 5.3.9 (Many-to-one formula over the whole tree)**.** *Under Assumption 5.3.1, if* $\mathbf{Z}_0 = \delta_{x_0}$,

where $x_0 \in E$, then for any non-negative measurable function $f : E \times \mathbb{R}_+ \to \mathbb{R}$, we have

$$\mathbb{E}\left[\sum_{u\in\mathcal{T}} f\left(X^u_{\beta(u)-}, \beta(u)\right)\right] = \int_0^{+\infty} \mathbb{E}\left[f(Y_s, s)\, \frac{r(Y_s)}{V(Y_s)}\right] V(x_0)e^{\lambda_0 s} ds$$

Proof. First we have, for all $u \in \mathcal{U}$,

$$\mathbb{E}\left[\mathbf{1}_{\{u\in\mathcal{T}\}} f\left(X^u_{\beta(u)-}, \beta(u)\right)\right] = \mathbb{E}\left[\mathbf{1}_{\{u\in\mathcal{T}\}} \int_{\alpha(u)}^{\beta(u)} f(X^u_s, s) r(X^u_s) ds\right]$$

because

$$\begin{aligned}
&\mathbb{E}\left[\mathbf{1}_{\{u\in\mathcal{T}\}} \int_{\alpha(u)}^{\beta(u)} f(X^u_s, s) r(X^u_s) ds\right] \\
=&\mathbb{E}\left[\mathbf{1}_{\{u\in\mathcal{T}\}} \int_0^{+\infty} \int_{\alpha(u)}^{\tau} f(X^u_s, s) r(X^u_s) ds\; r(X^u_\tau) e^{-\int_{\alpha(u)}^{\tau} r(X^u_t) dt}\, d\tau\right] \\
=&\mathbb{E}\left[\mathbf{1}_{\{u\in\mathcal{T}\}} \int_{\alpha(u)}^{+\infty} \int_{s}^{+\infty} r(X^u_\tau) e^{-\int_{\alpha(u)}^{\tau} r(X^u_t) dt}\, d\tau\; f(X^u_s, s) r(X^u_s) ds\right] \\
=&\mathbb{E}\left[\mathbf{1}_{\{u\in\mathcal{T}\}} \int_{\alpha(u)}^{+\infty} e^{-\int_{\alpha(u)}^{s} r(X^u_t) dt} f(X^u_s, s) r(X^u_s) ds\right] \\
=&\mathbb{E}\left[\mathbf{1}_{\{u\in\mathcal{T}\}} f\left(X^u_{\beta(u)-}, \beta(u)\right)\right].
\end{aligned}$$

Thus,

$$\mathbb{E}\left[\mathbf{1}_{\{u\in\mathcal{T}\}} f\left(X^u_{\beta(u)-}, \beta(u)\right)\right] = \mathbb{E}\left[\int_0^{+\infty} \mathbf{1}_{\{u\in V_s\}} f(X^u_s, s) r(X^u_s) ds\right],$$

and finally,

$$\begin{aligned}
\mathbb{E}\left[\sum_{u\in\mathcal{T}} f\left(X^u_{\beta(u)-}, \beta(u)\right)\right] &= \int_0^{+\infty} \mathbb{E}\left[\sum_{u\in V_s} f(X^u_s, s) r(X^u_s)\right] ds \\
&= \int_0^{+\infty} \mathbb{E}\left[f(Y_s, s) \frac{r(Y_s)}{V(Y_s)}\right] V(x_0) e^{\lambda_0 s} ds.
\end{aligned}$$

□

If we set $g(x,s) = f(x,s)/V(x)$ then we have:

$$\mathbb{E}\left[\sum_{u\in\mathcal{T}} g\left(X^u_{\beta(u)-},\beta(u)\right)\ V\left(X^u_{\beta(u)-}\right)\right] = \int_0^{+\infty} \mathbb{E}\left[g(Y_s,s)r(Y_s)\right] \times \mathbb{E}[\mathbf{Z}_s(V)]\ ds.$$

This equality means that adding the contributions over all the individuals corresponds to integrating the contribution of the auxiliary process over the average number of living individuals at time s. Let $(P_t)_{t\geq 0}$ be the semigroup of the auxiliary process; it is defined for any continuous and bounded f by

$$P_t f(x) = \mathbb{E}[f(Y_t) \mid Y_0 = x]$$

Lemma 5.3.10 (Many-to-one formula for forks)**.** *Under Assumption 5.3.1, if* $\mathbf{Z}_0 = \delta_{x_0}$*, where* $x_0 \in E$*, then for all non-negative and measurable function* f, g *on* E*, we have*

$$\begin{aligned}&\mathbb{E}\left[\sum_{u,v\in\mathcal{V}_t, u\neq v} f(X^u_t)V(X^u_t)g(X^v_t)V(X^v_t)\right]\\ =&\mathbb{E}[\mathbf{Z}_t(V)]^2 \int_0^t \frac{1}{\mathbb{E}[\mathbf{Z}_s(V)]}\mathbb{E}\left[J_2(VP_{t-s}f, VP_{t-s}g)(Y_s)\ \frac{r(Y_s)}{V(Y_s)}\right] ds\end{aligned}$$

where J_2 *is defined by*

$$J_2(f,g)(x) = \int_0^1 \sum_{a\neq b} \sum_{k\geq \max(a,b)} p_k(x) f\left(F^{(k)}_a(x,\theta)\right) g\left(F^{(k)}_b(x,\theta)\right)\ d\theta.$$

The operator J_2 describes the starting positions of two siblings picked at random.

Proof. Let $u, v \in \mathcal{V}_t$ be such that $u \neq v$, then there exists $(w,\tilde{u},\tilde{v}) \in \mathcal{U}^3$ and $a, b \in \mathbb{N}^*$, $a \neq b$ such that $u = wa\tilde{u}$ and $v = wb\tilde{v}$. The cell w is sometimes called the most recent common ancestor (MRCA). We have

$$\begin{aligned}&\mathbb{E}\left[\sum_{u,v\in\mathcal{V}_t, u\neq v} f(X^u_t)V(X^u_t)g(X^v_t)V(X^v_t)\right]\\ =&\sum_{w\in\mathcal{U}}\sum_{a\neq b}\sum_{\tilde{u},\tilde{v}\in\mathcal{U}} \mathbb{E}\left[\mathbf{1}_{\{u\in\mathcal{V}_t\}} f(X^u_t)V(X^u_t)\mathbf{1}_{\{v\in\mathcal{V}_t\}} g(X^v_t)V(X^v_t)\right],\end{aligned}$$

where $u = wa\tilde{u}$ and $v = wb\tilde{v}$. Let $\mathcal{F}_t = \sigma\{\mathbf{Z}_s \mid s \leq t\}$. By the conditional independence between descendants, we have

$$\mathbb{E}\left[\sum_{u,v\in\mathcal{V}_t,u\neq v} f(X_t^u)V(X_t^u)g(X_t^v)V(X_t^v)\right]$$
$$=\sum_{w\in\mathcal{U}}\sum_{a\neq b}\mathbb{E}\left[\mathbb{E}\left[\sum_{\tilde{u}\in\mathcal{U}}\mathbf{1}_{\{u\in\mathcal{V}_t\}}f(X_t^u)V(X_t^u)|\mathcal{F}_{\beta(w)}\right]\mathbb{E}\left[\sum_{\tilde{v}\in\mathcal{U}}\mathbf{1}_{\{v\in\mathcal{V}_t\}}g(X_t^v)V(X_t^v)|\mathcal{F}_{\beta(w)}\right]\right].$$

Therefore, as $\beta(w)$ is a stopping time, then using the strong Markov property and (5.6), we have

$$\begin{aligned}
&\mathbb{E}\left[\sum_{u,v\in\mathcal{V}_t,u\neq v} f(X_t^u)V(X_t^u)g(X_t^v)V(X_t^v)\right]\\
=&\sum_{w\in\mathcal{U}}\sum_{a\neq b}\mathbb{E}[\mathbf{1}_{\{wa,wb\in\mathcal{T},t\geq\beta(w)\}}P_{t-\beta(w)}f(X^{wa}_{\beta(w)})V(X^{wa}_{\beta(w)})\\
&P_{t-\beta(w)}g(X^{wb}_{\beta(w)})V(X^{wb}_{\beta(w)})e^{2\lambda_0(t-\beta(w))}]\\
=&\mathbb{E}\left[\sum_{w\in\mathcal{T}}\mathbf{1}_{\{t\geq\beta(w)\}}J_2(VP_{t-\beta(w)}f,VP_{t-\beta(w)}g)(X^w_{\beta(w)-})\,e^{2\lambda_0(t-\beta(w))}\right]\\
=&e^{2\lambda_0 t}V(x_0)\int_0^t\mathbb{E}\left[J_2(VP_{t-s}f,VP_{t-s}g)(Y_s)\,\frac{r(Y_s)}{V(Y_s)}\right]e^{-\lambda_0 s}ds.
\end{aligned}$$

□

5.3.3 Proof of Theorem 5.1.1

In this section, we give the main limit theorem which implies Theorem 5.1.1.

Theorem 5.3.11 (General Condition for the convergence of the empirical measure)**.** *Under Assumption 5.3.1, if f is a measurable function defined on E and μ a probability measure such that there exists a probability measure π, two constants $\alpha<\lambda_0$ and $C>0$, and a measurable function h such that*

1. $\pi(|f|)<+\infty$ *and* $\forall x\in E, \lim_{t\to+\infty}P_tf(x)=\pi(f)$,

2. $\mu(V)<+\infty$ *and* $\mu P_t(f^2\times V)\leq Ce^{\alpha t}$,

3. $P_t|f|\leq h$ *and* $\mu P_s\left(J_2(Vh,Vh)\,\frac{r}{V}\right)\leq Ce^{\alpha t}$,

and $\mathbf{Z}_0 = \delta_{X_0^\emptyset}$*, where* $X_0^\emptyset \sim \mu$*, then we have*

$$\lim_{t\to+\infty} \frac{1}{\mathbb{E}[\mathbf{Z}_t(V)]} \sum_{u\in\mathcal{V}_t} f(X_t^u)V(X_t^u) = W \times \pi(f),$$

where the convergence holds in probability. If furthermore $(\mathbf{Z}_t(V)e^{-\lambda_0 t})_{t\geq 0}$ *is bounded in* L^2 *then the convergence holds in* L^2*.*

Note that the constants and π may depend on f and μ! Also note that λ_0 is not assumed to be the largest eigenvalue.

Proof. As in [BDMT11, Theorem 4.2], we first prove the convergence for f such that $\pi(f) = 0$. We have $\mathbb{E}[\mathbf{Z}_t(V)] = \mu(V)e^{\lambda_0 t}$ and so

$$\mathbb{E}\left[\left(\frac{1}{\mathbb{E}[\mathbf{Z}_t(V)]}\sum_{u\in\mathcal{V}_t} f(X_t^u)V(X_t^u)\right)^2\right] = \mathbb{E}\left[\mathbf{Z}_t(f\times V)^2 e^{-2\lambda_0 t}\mu(V)^{-2}\right] = A_t + B_t,$$

where

$$A_t = e^{-2\lambda_0 t}\mu(V)^{-2}\mathbb{E}\left[\sum_{u\in\mathcal{V}_t} f^2(X_t^u)V^2(X_t^u)\right] = e^{-\lambda_0 t}\mu(V)^{-1}\mathbb{E}\left[f^2(Y_t)V(Y_t)\right],$$

and

$$\begin{aligned} B_t &= e^{-2\lambda_0 t}\mu(V)^{-2}\mathbb{E}\left[\sum_{u,v\in\mathcal{V}_t,\ u\neq v} f(X_t^u)V(X_t^u)f(X_t^v)V(X_t^v)\right] \\ &= \mu(V)^{-1}\int_0^t \mathbb{E}\left[J_2(VP_{t-s}f, VP_{t-s}f)(Y_s)\,\frac{r(Y_s)}{V(Y_s)}\right] e^{-\lambda_0 s}ds. \end{aligned}$$

From (2), we get $\lim_{t\to+\infty} A_t = 0$. Since $\pi(f) = 0$, from (1), we get $\lim_{t\to} P_t f = 0$. Then, by (3) and Lebesgue's theorem, we obtain that, for all $s \geq 0$ and $x \in E$,

$$\lim_{t\to+\infty} J_2(VP_{t-s}f, VP_{t-s}f)(x) = 0.$$

And again by (3) and Lebesgue's theorem, we obtain that $\lim_{t\to+\infty} B_t = 0$. Now, if $\pi(f) \neq 0$ then we have

$$\mathbf{Z}_t(fV)e^{-\lambda_0 t}\mu(V)^{-1} - W\pi(f) = \mathbf{Z}_t\left((f-\pi(f))V\right)e^{-\lambda_0 t}\mu(V)^{-1} + \pi(f)\left(\mathbf{Z}_t(V)e^{-\lambda_0 t}\mu(V)^{-1} - W\right).$$

Then, as a consequence of the first part of the proof, the first term of the sum, in the right hand side, converges to 0 in L^2. Moreover, the second term converges to 0 in probability thanks to lemma 5.3.2. □

Proof of Theorem 5.1.1. If $f = g/V$ then it is a continuous and bounded function. If $h \equiv 1$ then all assumptions of the previous theorem hold and we get the first convergence. Now if V is lower bounded, we can use the same argument with $g = 1$ and $f = 1/V$ which is also a continuous and bounded function. □

5.4 Macroscopic approximation

To prove Theorem 5.1.2, we need to use different topologies on $\mathcal{M}(E)$. Let $(\mathcal{M}(E), d_v)$ (resp. $(\mathcal{M}(E), d_w)$) be the set of finite measure when it is embedded with the vague (resp. weak) topology. These topologies are defined as follow.

$$\lim_{n\to+\infty} d_v(X_n, X_\infty) = 0 \iff \forall f \in C_0,\ \lim_{n\to+\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X_\infty)],$$

$$\lim_{n\to+\infty} d_w(X_n, X_\infty) = 0 \iff \forall f \in C_b,\ \lim_{n\to+\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X_\infty)],$$

where $(X_n)_{\geq 1}$ is a sequence of $\mathcal{M}(E)$ and $X_\infty \in \mathcal{M}(E)$. Here, C_0 is the set of continuous functions which vanish at infinity, and C_b is the set of continuous and bounded functions. Let $\mathbb{D}([0,T], E)$ and $C([0,T], E)$ be respectively the set of càd-làg functions embedded with the Skohorod topology and continuous functions embedded with the uniform topology [Bil99].

5.4.1 Proof of Theorem 5.1.2

Let $(\mathbf{Z}^{(n)})_{n\geq 1}$ be a sequence of random measure-valued distributed as $\mathbf{Z}$. In this section, we consider the following scaling: $\mathbf{X}^{(n)} = \frac{1}{n}\mathbf{Z}^{(n)}$, and we describe the behaviour of this scaled process when n goes to infinity.

To understand the behaviour of our model in a large population, we can consider that it starts from a deterministic probability measure $\mathbf{X}_0$, and approach it by the interesting sequence defined by

$$\mathbf{X}_0^{(n)} = \frac{1}{n}\sum_{k=0}^{n} \delta_{Y_k},$$

where $(Y_k)_{k\geq 1}$ is a sequence of i.i.d. random variable distributed according to $\mathbf{X}_0$. In other words, we

set

$$\mathbf{Z}_0^{(n)} = \sum_{k=0}^{n} \delta_{Y_k}.$$

The sequence $\mathbf{X}^{(n)}$ converges. Indeed, by the branching property, we have $\mathbf{Z}^{(n)} \overset{d}{=} \sum_{k=0}^{n} \mathbf{Z}^{Y_k}$, where $\mathbf{Z}_t^{Y_k}$ are i.i.d., distributed as $\mathbf{Z}$ and starting from $\mathbf{Z}_0^{Y_k} = \delta_{Y_k}$. Henceforth, if f is a continuous and bounded function then the classical law of large number gives

$$\forall t \geq 0,\ \lim_{n\to\infty} \mathbf{X}_t^{(n)}(f) = \mathbb{E}\left[\mathbf{Z}_t^{Y_1}(f)\right] a.s.$$

So by corollary 5.2.4, it implies that $\mathbf{X}^{(n)}$ (pointwise) converges to the solution $(\mu_t)_{t\geq 0}$ of the following integro-differential equation:

$$\begin{aligned}\mu_t(f) = \mu_0(f) + \int_0^t \mu_s(Gf) \\ + \int_E r(x) \sum_{k\geq 0} p_k(x) \int_0^1 \sum_{j=1}^{k} f(F_j^{(k)}(x,\theta))d\theta - f(x)\mu_s(dx)ds.\end{aligned} \tag{5.7}$$

Theorem 5.1.2 gives a stronger convergence.

Lemma 5.4.1 (Semi-martingale decomposition)**.** *If Assumption 5.2.1, then for all $f \in C_c^2(E)$ and $t \geq 0$,*

$$\mathbf{X}_t^{(n)}(f) = \mathbf{X}_0^{(n)}(f) + \mathbf{M}_t^{(n)}(f) + \mathbf{V}_t^{(n)}(f),$$

where

$$\mathbf{V}_t^{(n)}(f) = \int_0^t \int_E Gf(x) + r(x) \int_0^1 \sum_{k\in\mathbb{N}} \sum_{j=1}^{k} f(F_j^{(k)}(x,\theta)) - f(x) p_k d\theta \mathbf{X}_s^{(n)}(dx)ds,$$

and $\mathbf{M}_t^{(n)}(f)$ is a square-integrable and càdlàg martingale. Its bracket is defined by

$$\begin{aligned}\langle \mathbf{M}^{(n)}(f)\rangle_t = \frac{1}{n}\int_0^t 2\mathbf{X}_s^{(n)}(Gf^2) - 2\mathbf{X}_s^{(n)}(f \times Gf) \\ + \int_E r(x) \int_0^1 \sum_{k\in\mathbb{N}^*} \left(\sum_{j=1}^{k} f(F_j^{(k)}(x,\theta)) - f(x)\right)^2 p_k(x)d\theta\ \mathbf{X}_s^{(n)}(dx)ds.\end{aligned}$$

Proof. It is a direct consequence of Lemma 5.2.3. Indeed, if $\mathbb{L}^{(n)}$ is the generator of $\mathbf{X}^{(n)}$ then it verifies

$$\mathbb{L}^{(n)}F_f(\mu) = \partial_t \mathbb{E}[F_f(\mathbf{X}^{(n)})|\mathbf{X}_0^{(n)} = \mu]\big|_{t=0} = \partial_t \mathbb{E}[F_{f/n}(\mathbf{Z}^{(n)})|\mathbf{Z}_0^{(n)} = n\mu]\big|_{t=0} = \mathbb{L}F_{f/n}(n\mu),$$

where $F_f(\mu) = F(\mu(f))$, F, f are two test functions and $\mathbb{L}$ is the generator of $\mathbf{Z}$. □

Remark 5.4.2 (Non explosion)**.** *Let us recall that, by Lemma 5.2.2, if the assumptions of Theorem 5.1.2 hold then Assumption 5.2.1 holds; that is there is no explosion.*

Let us denote by $\mathcal{L}(U)$ the law of U, for any random variable U.

Lemma 5.4.3. *Under the assumptions of Theorem 5.1.2 the sequence $(\mathcal{L}(\mathbf{X}^{(n)}))_{n\geq 1}$ is uniformly tight in the space of probability measures on $\mathbb{D}([0,T],(\mathcal{M}(E),d_v))$.*

Proof. We follow the approach of [FM04]. According to [RC86], it is enough to show that, for any continuous bounded function f, the sequence of laws of $\mathbf{X}^{(n)}(f)$ is tight in $\mathbb{D}([0,T],\mathbb{R})$. To prove it, we will use the Aldous-Rebolledo criterion. Let C_c^∞ be the sef of functions of class C^∞ with finite support, we set $S = C_c^\infty \cup \{\mathbf{1}\}$, where $\mathbf{1}$ is the mapping $x \mapsto 1$. We have to prove that, for any function $f \in S$, we have

1. $\forall t \geq 0$, $\left(\mathbf{X}_t^{(n)}(f)\right)_{n\geq 0}$ is tight;
2. for all $n \in \mathbb{N}$, and $\varepsilon, \eta > 0$, there exists $\delta > 0$ such that for each stopping time S_n bounded by T,

$$\limsup_{n\to+\infty} \sup_{0\leq u\leq\delta} \mathbb{P}(|\mathbf{V}_{S_n+u}^{(n)}(f) - \mathbf{V}_{S_n}^{(n)}(f)| \geq \eta) \leq \varepsilon.$$

$$\limsup_{n\to+\infty} \sup_{0\leq u\leq\delta} \mathbb{P}(|\langle \mathbf{M}^{(n)}(f)\rangle_{S_n+u} - \langle \mathbf{M}^{(n)}(f)\rangle_{S_n}| \geq \eta) \leq \varepsilon.$$

The first point is the tightness of the family of time-marginals $(\mathbf{X}_t^{(n)}(f))_{n\geq 1}$ and the second point, called the Aldous condition, gives a "stochastic continuity". It looks like the Arzelà-Ascoli Theorem. Using Lemma 5.2.2, there exists $C > 0$ such that

$$\begin{aligned}\mathbb{P}(|\mathbf{X}_t^{(n)}(f)| > k) &\leq \frac{\|f\|_\infty\, \mathbb{E}[\mathbf{X}_t^{(n)}(\mathbf{1})]}{k} \\ &\leq \frac{\|f\|_\infty\, C\, \mathbb{E}[\mathbf{X}_0^{(n)}(\mathbf{1})]}{k},\end{aligned}$$

which tends to 0 as k tends to infinity. This proves the first point. Let $\delta > 0$, we get for all stopping

times $S_n \leq T_n \leq (S_n + \delta) \leq T$, that there exist $C', C_f > 0$ such that

$$\begin{aligned}\mathbb{E}[|\mathbf{V}_{T_n}^{(n)}(f) - \mathbf{V}_{S_n}^{(n)}(f)|] = \mathbb{E}\Bigg[\Bigg|\int_{S_n}^{T_n} \mathbf{X}_s^{(n)}(Gf) \\ + \int_E r(x)\int_0^1 \sum_{k\in\mathbb{N}}\sum_{j=1}^{k} f\left(F_j^{(k)}(x,\theta)\right) - f(x)p_k(x)d\theta\, \mathbf{X}_s^{(n)}(dx)ds\Bigg|\Bigg] \\ \leq C'\left[\|Gf\|_\infty + \|f\|_\infty\right] \times \mathbb{E}\left[|T_n - S_n|\right] \\ \leq C_f\delta.\end{aligned}$$

In the other hand, there exists $C'_f > 0$ such that

$$\begin{aligned}&\mathbb{E}[|\langle \mathbf{M}^{(n)}(f)\rangle_{T_n} - \langle \mathbf{M}^{(n)}(f)\rangle_{S_n}|] \\ =&\frac{1}{n}\mathbb{E}\Bigg[\Bigg|\int_{S_n}^{T_n} 2\mathbf{X}_s^{(n)}(Gf^2) - 2\mathbf{X}_s^{(n)}(fGf) \\ &+ \int_E r(x)\int_0^1 \sum_{k\in\mathbb{N}}\sum_{j=1}^{k}\left(f\left(F_j^{(k)}(x,\theta)\right) - f(x)\right)^2 p_k d\theta\, \mathbf{X}_s^{(n)}(dx)ds\Bigg|\Bigg] \\ \leq&\frac{C'_f\delta}{n.}\end{aligned}$$

Then, for a sufficiently small δ, the second point is verified and we conclude that $(\mathbf{X}^{(n)})_{n\geq 1}$ is uniformly tight in $\mathbb{D}([0,T],(\mathcal{M}(E),d_v))$. □

Proof of Theorem 5.1.2. Let us denote by $\mathbf{X}$ a limit process of $(\mathbf{X}^{(n)})_{n\geq 1}$; namely there exists an increasing sequence $(u_n)_{n\geq 1}$, on $\mathbb{N}^*$, such that $(\mathbf{X}^{(u_n)})_{n\geq 1}$ converges to $\mathbf{X}$. It is almost surely continuous in $(\mathcal{M}(E), v)$ since

$$\sup_{t\geq 0}\sup_{\|f\|_\infty\leq 1}|\mathbf{X}_{t-}^{(n)}(f) - \mathbf{X}_t^{(n)}(f)| \leq \frac{\bar{k}}{n}. \tag{5.8}$$

In the case where E is compact, the vague and weak topologies coincide. By Doob's inequality, there exists $C > 0$ such that

$$\sup_f \mathbb{E}\left[\sup_{t\leq T}\left|\mathbf{M}_t^{(n)}(f)\right|\right] \leq 2\sup_f \mathbb{E}\left[\langle \mathbf{M}^{(n)}(f)\rangle_T\right] \leq \frac{C}{n}$$

where the supremum is taken over all the function $f \in C_c^2(E)$ such that $\|f\|_\infty \leq 1$. Hence,

$$\lim_{n\to+\infty}\sup_f \mathbb{E}\left[\sup_{t\leq T}\left|\mathbf{M}_t^{(n)}(f)\right|\right] = 0. \tag{5.9}$$

But as

$$\mathbf{M}_t^{(n)}(f) = \mathbf{X}_t^{(n)}(f) - \mathbf{X}_0^{(n)}(f) \\ - \int_0^t \int_E Gf(x) + r(x) \int_0^1 \sum_{k\in\mathbb{N}} \sum_{j=1}^{k} f\left(F_j^{(k)}(x,\theta)\right) - f(x)\, p_k(x) d\theta\, \mathbf{X}_s^{(n)}(dx)ds,$$

we have

$$0 = \mathbf{X}_t(f) - \mathbf{X}_0(f) - \int_0^t \mathbf{X}_s(Gf) \\ + \int_E r(x) \left(\sum_{j=1}^{k} f(F_j^{(K)}(x,\theta))\, p_k(x)\, d\theta - f(x) \right) \mathbf{X}_s(dx)ds.$$

Since this equation has a unique solution, it ends the proof when E is compact. This approach fails in the non-compact case. Nevertheless, we can use the Méléard-Roelly criterion [MR93]. We have to prove that $\mathbf{X}$ is in $C([0,T],(\mathcal{M}(E),w))$ and $\mathbf{X}^{(n)}(\mathbf{1})$ converges to $\mathbf{X}(\mathbf{1})$. By (5.8), $\mathbf{X}$ is continuous. To prove that $\mathbf{X}^{(n)}(\mathbf{1})$ converges to $\mathbf{X}(\mathbf{1})$, we use the following lemmas.

Lemma 5.4.4 (Approximation of indicator functions)**.** *For each $k \in \mathbb{N}$, there exists $\psi_k \in C^2(E)$ such that:*

$$\forall x \in E,\ \mathbf{1}_{[k;+\infty[}(x) \leq \psi_k(x) \leq \mathbf{1}_{[k-1;+\infty[}(x) \text{ and } \exists C,\ G\psi_k \leq C\psi_{k-1}.$$

Proof. See [JMW12, lemma 4.2] or [MT12, lemma 3.3]. □

Lemma 5.4.5 (Commutation of limits)**.** *Under the assumptions of Theorem 5.1.2,*

$$\lim_{k\to+\infty} \limsup_{n\to+\infty} \mathbb{E}\left[\sup_{t\leq T} \mathbf{X}_t^{(n)}(\psi_k)\right] = 0,$$

where $(\psi_k)_{k\geq 0}$ are defined as in the previous lemma.

The proof is postponed after. Hence, a same computation to [MT12] gives us the convergence in $\mathbb{D}([0,T],(\mathcal{M}(E),w))$. Thus, each subsequence converges to the equation (5.7). The end of the proof follow with the same argument of the compact case.

We can give another argument, which does not use the Méléard-Roelly criterion [MR93]. As

$$\sup_{t\geq 0} \sup_{\|f\|_\infty \leq 1} |\mathbf{X}_{t-}^{(n)}(f) - \mathbf{X}_t^{(n)}(f)| \leq \frac{\bar{k}}{n},$$

$\mathbf{X}$ is continuous from $[0,T]$ to $(\mathcal{M}(E),d_w)$. Let $\mathbf{G}$ be a Lipschitz function on $C([0,T],(\mathcal{M}(E),d_w))$,

we get,

$$\begin{aligned}|\mathbb{E}[\mathbf{G}(\mathbf{X}^{(u_n)})] - \mathbf{G}(\mathbf{X})| &\leq \mathbb{E}\left[\sup_{t\in[0,T]} d_w\left(\mathbf{X}_t^{(u_n)}, \mathbf{X}_t\right)\right] \\ &\leq \mathbb{E}\left[\sup_{t\in[0,T]} d_w\left(\mathbf{X}_t^{(u_n)}, \mathbf{X}_t^{(u_n)}(\cdot\times(1-\psi_k))\right)\right] \\ &+ \mathbb{E}\left[\sup_{t\in[0,T]} d_w\left(\mathbf{X}_t^{(u_n)}(\cdot\times(1-\psi_k)), \mathbf{X}_t(\cdot\times(1-\psi_k))\right)\right] \\ &+ \sup_{t\in[0,T]} d_w\left(\mathbf{X}_t(\cdot\times(1-\psi_k)), \mathbf{X}_t\right).\end{aligned}$$

According to Lemma 5.4.5, we obtain that

$$\lim_{k\to+\infty} \limsup_{n\to+\infty} \mathbb{E}\left[\sup_{t\in[0,T]} d_w\left(\mathbf{X}_t^{(u_n)}, \mathbf{X}_t^{(u_n)}(\cdot\times(1-\psi_k))\right)\right] = 0$$

and

$$\lim_{k\to+\infty} \sup_{t\in[0,T]} d_w(\mathbf{X}_t(\cdot\times(1-\psi_k)), \mathbf{X}_t) = 0.$$

Then, we have

$$\begin{aligned}&d_w\left(\mathbf{X}_t^{(u_n)}(\cdot\times(1-\psi_k)), \mathbf{X}_t(\cdot\times(1-\psi_k))\right) \\ &= d_v\left(\mathbf{X}_t^{(u_n)}(\cdot\times(1-\psi_k)), \mathbf{X}_t(\cdot\times(1-\psi_k))\right).\end{aligned}$$

Thus,

$$\lim_{k\to+\infty} \limsup_{n\to+\infty} \mathbb{E}\left[\sup_{t\in[0,T]} d_w\left(\mathbf{X}_t^{(u_n)}(\cdot\times(1-\psi_k)), \mathbf{X}_t(\cdot\times(1-\psi_k))\right)\right] = 0,$$

by continuity of $\nu \mapsto \nu(1-\psi_k)$ in $\mathbb{D}(\mathcal{M}(E), d_v)$. And finally,

$$\lim_{n\to+\infty} \mathbf{G}\left(\mathbf{X}^{(u_n)}\right) = \mathbf{G}(\mathbf{X}),$$

which completes the proof. □

proof of Lemma 5.4.5. If $\mu_t^{n,k} = \mathbb{E}(\mathbf{X}_t^{(n)}(\psi_k))$ then we have

$$\begin{aligned}\mu_t^{n,k} &= \mathbb{E}[\mathbf{X}_0^{(n)}(\psi_k)] + \int_0^t \mathbb{E}\left[\int_E G\psi_k(x)\right. \\ &\left. +r(x)\left(\sum_{k\geq 1}\sum_{j=1}^{k} p_k(x)\int_0^1 \psi_k(F_j^{(k)}(x,\theta)) - \psi_k(x)\right)\mathbf{X}_s^{(n)}(dx)\right] ds \\ &\leq \mu_0^{n,k} + C\int_0^t \mu_s^{n,k-1} + \mu_s^{n,k} ds.\end{aligned}$$

Now, by Gronwall's Lemma, iteration and monotonicity, we deduce that

$$\begin{aligned}\mu_t^{n,k} &\leq C_1(\mu_0^{n,k} + \int_0^t \mu_s^{n,k-1} ds) \\ &\leq C_1\mu_0^{n,k} + C_1^2 T\mu_0^{n,k-1} + \int_0^t\int_0^s \mu_u^{n,k-2} du ds \\ &\leq \sum_{l=0}^{k-1} \mu_0^{n,k-l} C_1 \frac{(C_1T)^l}{l!} + C_2 \times \frac{(C_1T)^k}{k!} \\ &\leq \mu_0^{n,\lfloor k/2\rfloor} C_1 e^{C_1T} + C_3 \sum_{l>\lfloor k/2\rfloor} \frac{(C_1T)^l}{l!} + C_2 \times \frac{(C_1T)^k}{k!},\end{aligned}$$

where C_1, C_2 and C_3 are three constants. Thus,

$$\lim_{k\to+\infty} \limsup_{n\to+\infty} \mu_t^{n,k} = 0.$$

And finally the following expression completes the proof,

$$\mathbb{E}\left[\sup_{t\leq T} |\mathbf{X}_t^n(\psi_k)|\right] \leq \mu_0^{n,k} + C\int_0^t \mu_s^{n,k-1} + \mu_s^{n,k} ds + \mathbb{E}\left[\sup_{t\leq T} |\mathbf{M}_t^{(n)}(\psi_k)|\right].$$

□

5.5 Main example : a size-structured population model

Let us introduce our main example. It is a size-structured population model which represents the cell mitosis. It is described as follows: the underlying process X is deterministic and linear and when

a cell dies, it divides in two parts. Formally and with our notations, we have

$$E = (0, +\infty),\ p_2 = 1,\ \text{and}\ Gf = f', \tag{5.10}$$

for every $f \in C_c^2(E)$, and

$$\forall x \in E,\ \forall \theta \in [0,1],\ F_1^{(2)}(x,\theta) = F^{-1}(\theta)x\ \text{ and }\ F_2^{(2)}(x,\theta) = (1 - F^{-1}(\theta))x, \tag{5.11}$$

where F is the cumulative distribution function of the random variable in $[0,1]$. It verifies $F(x) = 1 - F(1-x)$. In this case, one cell lineage is generated by:

$$\forall x \geq 0,\ Lf = f'(x) + r(x)\left[\mathbb{E}[f(Hx)] - f(x)\right],$$

for every $f \in C_c^2(E)$, where H is distributed according to F. This process is sometimes called the TCP (Transmission Control Protocol) process in computer science [CMP10, GRZ04, LvL08, OKM96]. Firstly, we prove the non explosion even if r is not bounded.

Lemma 5.5.1 (Non explosion)**.** *Let $p \geq 1$. If for all $x \in \mathbb{R}_+^*$, $r(x) \leq C_0(1+x^p)$, and $\mathbf{z}_0(1+x^p) < +\infty$, then our process is non explosive. Moreover*

$$\mathbb{E}\left[\sup_{s\in[0,T]} \mathbf{Z}_s(1+x^p)\right] \leq \mathbf{z}_0(1+x^p)e^{C_pT},$$

where C_p is constant and $T > 0$.

Proof. Recall that for every $f \in C_c^2(E)$, we have

$$\begin{aligned}\mathbf{Z}_t(f) = \mathbf{Z}_0(f) &+ \int_0^t \int_E f'(x)\, \mathbf{Z}_s(dx)\, ds \\ &+ \int_0^t \int_{\mathcal{U}\times\mathbb{R}_+\times[0,1]} \mathbf{1}_{\{u\in V_{s-}, l\leq r(X_{s-}^u)\}} f(\theta X_{s-}^u) + f((1-\theta)X_{s-}^u) - f(X_{s-}^u))\, \rho(ds,du,dl,d\theta)\end{aligned}$$

Using the same argument to [FM04, Theorem 3.1], we introduce $\tau_n = \inf\{\, t \geq 0 \mid \mathbf{Z}_t(1+x^p) > n \,\}$;

and we have

$$\begin{aligned}\sup_{u\in[0,t\wedge\tau_n]}\mathbf{Z}_u(1+x^p) \leq & \mathbf{Z}_0(1+x^p)+\int_0^{t\wedge\tau_n}\mathbf{Z}_s(px^{p-1})ds\\ &+\int_0^{t\wedge\tau_n}\int_{\mathcal{U}\times\mathbb{R}_+\times[0,1]}\mathbf{1}_{u\in V_{s-},l\leq r(X^u_{s-})}\\ &\quad(1+(\theta^p+(1-\theta)^p-1)(X^u_{s-})^p)\,\rho(ds,du,dl,d\theta)\\ \leq & \mathbf{Z}_0(1+x^p)+\int_0^{t\wedge\tau_n}p\times\sup_{u\in[0,s\wedge\tau_n]}\mathbf{Z}_u(1+x^p)ds\\ &+\int_0^t\int_{\mathcal{U}\times\mathbb{R}_+\times[0,1]}\mathbf{1}_{\{u\in V_{s-},l\leq r(X^u_{s-})\}}\,\rho(ds,du,dl,d\theta),\end{aligned}$$

because $(\theta^p+(1-\theta)^p-1)\leq 0$. Thus there exist $C>0$ such that

$$\mathbb{E}\left[\sup_{u\in[0,t\wedge\tau_n]}\mathbf{Z}_u(1+x^p)\right]\leq \mathbf{z}_0(1+x^p)+\int_0^t C\,\mathbb{E}\left[\sup_{u\in[0,s\wedge\tau_n]}\mathbf{Z}_u(1+x^p)\right]ds.$$

Finally, the Gronwall Lemma implies the existence of C_p such that

$$\mathbb{E}\left[\sup_{s\in[0,t\wedge\tau_n]}\mathbf{Z}_s(1+x^p)\right]\leq \mathbf{z}_0(1+x^p)e^{C_pt}.$$

We deduce that τ_n tends almost surely to infinity and that there is non explosion. □

5.5.1 Equal mitosis : long time behaviour

In this subsection, we establish the long time behaviour of $\mathbf{Z}$. We assume that

$$\forall x\geq 0,\forall\theta\in[0,1],\ F_1^{(2)}(x,\theta)=F_2^{(2)}(x,\theta)=\frac{x}{2}.$$

That is, the cells divide in two equal parts. In short, we have

$$\forall x\geq 0,\ \mathcal{G}f(x)=f'(x)+r(x)\left(f\left(\frac{x}{2}\right)-f(x)\right),$$

for every $f\in C_c^2(E)$. In order to give a many-to-one formula, we recall a theorem of [PR05]:

Theorem 5.5.2 (Sufficient condition for the existence of eigenelement)**.** *If there exist* $\underline{r},\bar{r}>0$ *such*

that

$$\underline{r} \leq r \leq \bar{r},$$

r is continuous and $r(x)$ is constant equal to r_∞ for x large enough, then there exist $V \in C^1(\mathbb{R}_+)$ and $\lambda_0 > 0$ such that

$$\mathcal{G}V = \lambda_0 V$$

and

$$\forall x \geq 0,\ c(1+x^k) \leq V(x) \leq C(1+x^k),$$

where C, c are two constant and $2^k = \frac{2r_\infty}{\lambda_0 + r_\infty}$.

So, we get a many-to-one formula with an auxiliary process generated by A be defined, for every $f \in C_c^2(E)$ and $x \in E$, by

$$Af(x) = f'(x) + r(x)\frac{2V(x/2)}{V(x)}\left(f(x/2) - f(x)\right).$$

Our main result gives the two following limit theorems.

Corollary 5.5.3 (Convergence of the empirical measure for a mitosis model)**.** *If there exist $\underline{r}, \bar{r} > 0$ such that*

$$\underline{r} \leq r \leq \bar{r},$$

r is continuous and $r(x)$ is constant equal to $\bar{r}$ for x large enough, then there exists a probability measure π such that, for any continuous and bounded function g, we have

$$\lim_{t\to+\infty} \frac{1}{N_t} \sum_{u\in\mathcal{V}_t} g(X_t^u) = \int g\, d\pi \text{ in probability.}$$

In particular for a constant rate r, π has Lebesgue density:

$$x \mapsto \frac{2r}{\prod_{n=1}^{+\infty}(1-2^{-n})} \sum_{n=0}^{+\infty} \left(\prod_{k=1}^{n} \frac{2}{1-2^k}\right) e^{-2^{n+1}rx}. \tag{5.12}$$

This explicit formula (5.12) is not new [Per07, PR05], but here, the empirical measure convergences in probability, while in the mentioned papers, the mean measure or the macroscopic process converges (see Theorem 5.1.2).

Proof of corollary 5.5.3. By Theorem 5.5.2, the mapping $x \mapsto V(x/2)/V(x)$ is upper and lower bounded. Thus, the auxiliary process is ergodic and admits a unique invariant law, as can be checked

using a suitable Foster-Lyapunov function [MT93, Theorem 6.1] (for instance, we can use $x \mapsto 1+x$). See also [GK10]. Using Theorem 5.1.1 we have the convergence on the set $\{W \neq 0\}$ where

$$W = \lim_{t\to\infty} \mathbf{Z}_t(V)e^{-\lambda_0 t} \text{ a.s..}$$

It rests to prove that $W > 0$ almost surely. We begin by proving that the martingale $(\mathbf{Z}_t(V)e^{-\lambda_0 t})_{t\geq 0}$ converges to W in L^1. Let $p > 1$, by the Burkholder-Davis-Gundy's inequality [DM80, Theorem 92 p.304], there exists $C > 0$ such that

$$\begin{aligned}\mathbb{E}\left[\left|\mathbf{Z}_t(V)e^{-\lambda_0 t} - \mathbf{Z}_0(V)\right|^p\right] &\leq C\mathbb{E}\left[\sum_{t\geq 0}\left|\mathbf{Z}_{t+}(V)e^{-\lambda_0 t} - \mathbf{Z}_{t-}(V)e^{-\lambda_0 t}\right|^p\right] \\ &\leq C\mathbb{E}\left[\sum_{u\in\mathcal{T}} e^{-\lambda_0 p\beta(u)}\left|2V\left(\frac{X^u_{\beta(u)-}}{2}\right) - V(X^u_{\beta(u)-})\right|^p\right]\end{aligned}$$

Now by Lemma 5.3.9, we have that

$$\mathbb{E}\left[\left|\mathbf{Z}_t(V)e^{-\lambda_0 t} - \mathbf{Z}_0(V)\right|^p\right] \leq \int_0^\infty \mathbb{E}[\mathbf{Z}_0(V)]e^{-(p-1)\lambda_0 s}\mathbb{E}\left[r(Y_s)\frac{|2V(Y_s/2) - V(Y_s)|^p}{V(Y_s)}\right] ds$$

Finally, using that r is bounded, the conclusion of Theorem 5.5.2 and that all moments of Y are bounded, we have that the martingale $(\mathbf{Z}_t(V)e^{-\lambda_0 t})_{t\geq 0}$ converges to W in L^1. We deduce that $\mathbb{E}[W] > 0$ and $\varrho = \mathbb{P}(W = 0) < 1$. But, conditioning to the time of the first division and taking the limit $t \to +\infty$ gives that $\varrho^2 = \varrho$. Finally $\varrho = 0$ and this ends the proof. Thee explicit formula is an application of [OKM96]. □

We can see that the assumptions of Theorem 5.5.2 are strong, and not necessary:

Corollary 5.5.4 (Convergence of the empirical measure when r is affine)**.** *If*

$$\forall x \geq 0, r(x) = ax + b,$$

where $a, b \geq 0$ and a or b is positive then there exists a measure π such that

$$\lim_{t\to+\infty} \frac{1}{N_t}\sum_{u\in\mathcal{V}_t} g(X^u_t) = \int g\, d\pi.$$

The convergence holds in probability and for any continuous function g on E such that $\forall x \in E,\ |g(x)| \leq C(1 + x)$.

Proof. If $r(x) = ax+b$ then $V(x) = x\frac{\sqrt{b^2+4a}-b}{2}+1$ is an eigenvector and $\frac{2a}{\sqrt{b^2+4a}-b}$ is its corresponding eigenvalue. Henceforth, this result is a direct application of Theorem 5.1.1 □

Remark 5.5.5 (Malthus parameter)**.** *We also deduce that*

$$\lim_{t\to+\infty} N_t\, e^{-\lambda_0 t} = W \int_E \frac{1}{V}\, d\pi,$$

where $\lambda_0 = \frac{2a}{\sqrt{b^2+4a}-b}$ *is the Malthus parameter (see Remark 5.3.7).*

Remark 5.5.6 (Estimation of r for the Escherichia coli cell)**.** *We can find some estimates of the division rate in the literature. An inverse problem was developed and applied with experimental data in [DMZ09](see also [Kub69]). More recently, [DHRR11] gives a nonparametric estimation of the division rate.*

5.5.2 Homogeneous case: moment and rate of convergence

When r is constant, the process is easier to study since the auxiliary process has already been studied [CMP10, LvL08, OKM96]. Here, we give the moments and a first approach to estimate the rate of convergence.

Lemma 5.5.7 (Moments of the empirical measure)**.** *For all* $m \in \mathbb{N}$*, and* $t \geq 0$*, we have*

$$\begin{aligned}\mathbb{E}[\mathbf{Z}_t(x^m)] &= \mathbb{E}\left[\sum_{u\in\mathcal{V}_t} (X_t^u)^m\right] \\ &= \int_0^{+\infty} e^{rt}\left[\frac{m!}{\prod_{i=1}^m \theta_i} + m!\sum_{i=1}^m\left(\sum_{k=0}^i \frac{x^k}{k!}\prod_{j=k,j\neq i}^m \frac{1}{\theta_j-\theta_i}\right)e^{-\theta_i t}\right]\mathbf{z}_0(dx),\end{aligned}$$

where $\theta_i = 2r\,(1-2^{-i})$*. In particular,*

$$\mathbb{E}[\mathbf{Z}_t(x)] = \mathbb{E}\left[\sum_{u\in\mathcal{V}_t^\mu} X_t^u\right] = e^{rt}\int_0^{+\infty} \frac{1}{r} - \left(\frac{1}{r}-x\right)e^{-rt}\,\mathbf{z}_0(dx),$$

and

$$\mathbb{E}[\mathbf{Z}_t(x^2)] = \mathbb{E}\left[\sum_{u\in\mathcal{V}_t}(X_t^u)^2\right]$$
$$= e^{rt}\int_0^{+\infty}\frac{4}{3r^2} + 2\left[e^{-rt}\left(\frac{-2}{r^2}+\frac{2x}{r}\right) + e^{-3rt/2}\left(\frac{4}{3r^2}-\frac{2x}{3r}+\frac{x^2}{2}\right)\right]\mathbf{z}_0(dx).$$

Proof. Since r is constant, we have $\mathcal{G}\mathbf{1} = r\mathbf{1}$, where $\mathbf{1}$ is the constant mapping, which is equal to 1. Lemma 5.3.4 gives

$$\frac{1}{\mathbb{E}[N_t]}\mathbb{E}\left[\sum_{u\in\mathcal{V}_t} f(X_t^u)\right] = \mathbb{E}[f(Y_t)],$$

for every continuous and bounded function f, where Y is generated by A be defined, for every $f \in C_c^2(E)$ and $x \in E$, by

$$Af(x) = f'(x) + 2r\left(f\left(\frac{x}{2}\right) - f(x)\right).$$

Finally, we complete the proof using [LvL08, Theorem 4]. □

Now, let us talk about the rate of convergence. To estimate the distance between two random measures, we will use the Wasserstein distance [Rac91, Vil09]:

Definition 5.5.8 (Wasserstein distance)**.** *Let μ_1 and μ_2 two finite measures on a Polish space (F, d_F), the Wasserstein distance between μ_1 and μ_2 is defined by*

$$W_{d_F}(\mu_1,\mu_2) = \inf\int_{F\times F} d_F(x_1,x_2)\Pi(dx_1,dx_2),$$

where the infimum runs over all the measures Π on $F\times F$ with marginals μ_1 and μ_2. In particular, if μ_1 and μ_2 are two probability measures, we have

$$W_{d_F}(\mu_1,\mu_2) = \inf\mathbb{E}[d_F(X_1,X_2)],$$

where the infimum runs over all two random variables X_1, X_2, which are distributed according to μ_1, μ_2.

So, if M_1, M_2 are two random measures then

$$W_d(\mathcal{L}(M_1),\mathcal{L}(M_2)) = \inf\ \mathbb{E}[d(M_1,M_2)],$$

where the infimum is taken over all the couples of random variables (M_1, M_2) such that $M_1 \sim \mathcal{L}(M_1)$ and $M_2 \sim \mathcal{L}(M_2)$, and d is a distance on the measures space. Here, we consider $d = W_{|\cdot|}$. It is the Wasserstein distance on $(E, |\cdot|)$. We have

Theorem 5.5.9 (Quantitative bounds)**.** *If r is constant, then we have, for all $t \geq 0$,*

$$W_{W_{|\cdot|}}\left(\mathcal{L}\left(\frac{\mathbf{Z}_t^x}{\mathbb{E}[N_t]}\right), \mathcal{L}\left(\frac{\mathbf{Z}_t^y}{\mathbb{E}[N_t]}\right)\right) \leq |x-y|e^{-rt},$$
$$W_{W_{|\cdot|}}\left(\mathcal{L}\left(\frac{\mathbf{Z}_t^x}{N_t}\right), \mathcal{L}\left(\frac{\mathbf{Z}_t^y}{N_t}\right)\right) \leq |x-y|\frac{rte^{-rt}}{1-e^{-rt}},$$

where $\mathbf{Z}^{\mathbf{x}}, \mathbf{Z}^{\mathbf{y}}$ are distributed as $\mathbf{Z}$ and start from δ_x and δ_y.

This result does not give a bound for $W_{W_{|\cdot|}}\left(\mathcal{L}\left(\mathbf{Z}_t/\mathbb{E}[N_t]\right), \mathcal{L}\left(W\pi\right)\right)$ or $W_{W_{|\cdot|}}\left(\mathcal{L}\left(\mathbf{Z}_t/N_t\right), \mathcal{L}\left(\pi\right)\right)$, where π is the limit measure of Corollary 5.5.3.

proof of Theorem 5.5.9. By homogeneity, we can see our branching measure $\mathbf{Z}$ as a process indexed by a Galton-Watson tree [BDMT11]. For our coupling, we take two processes indexed by the same tree. More precisely, as the branching time does not depend on the position, we can set the same times to our two processes. Let $\mathcal{T} = \bigcup_{n\in\mathbb{N}}\{1,2\}^n$ representing cells that have lived at a certain moment. Let $(d_u)_{u\in\mathcal{U}}$ be a family of i.i.d. exponential variables with mean $1/r$, which model the lifetimes. We build $\mathbf{Z}^x$ and $\mathbf{Z}^y$ by induction. First, for all $t \in [0, d_\emptyset)$, $X_t^\emptyset = x + t$ and $Y_t^\emptyset = y + t$. We set $\alpha(\emptyset) = 0$. Then, for all $u \in \mathcal{T}$ and $k \in \{1,2\}$, we set $\alpha(uk) = \alpha(u) + d_u$ and

$$\forall t \in [\alpha(uk), \alpha(uk) + d_{uk}),\ X_t^{uk} = \frac{1}{2}X^u_{\alpha(uk)-} + t - \alpha(uk)$$

and $Y_t^{uk} = Y^u_{\alpha(uk)-}/2 + t - \alpha(uk)$. Finally we have $\mathcal{V}_t = \{u \in \mathcal{T} \mid \alpha(u) \leq t < \alpha(u) + d_u\}$ and

$$\mathbf{Z}_t^x = \sum_{u\in\mathcal{V}_t}\delta_{X_t^u} \text{ and } \mathbf{Z}_t^y = \sum_{u\in\mathcal{V}_t}\delta_{Y_t^u}.$$

We observe that, for any cell u, the trajectories of X^u and Y^u are parallel (because they are linear). When a branching occurs, $\sum_{u\in\mathcal{V}_t}|X_t^u - Y_t^u|$ is constant. Hence, we easily deduce that

$$\sum_{u\in\mathcal{V}_t}|X_t^u - Y_t^u| = |x-y|.$$

Finally we have, for all $t \geq 0$,

$$\begin{aligned} W_{|\cdot|}(\mathbf{Z}_t^x, \mathbf{Z}_t^y) &\leq \sum_{u \in \mathcal{V}_t} |X_t^u - Y_t^u| \\ &\leq |x - y|. \end{aligned}$$

Dividing by $\mathbb{E}[N_t] = e^{-rt}$, we obtain the first bound. For the second bound, a similar computation gives

$$W_{W_{|\cdot|}}\left(\mathcal{L}\left(\frac{\mathbf{Z}_t^x}{N_t}\right), \mathcal{L}\left(\frac{\mathbf{Z}_t^y}{N_t}\right)\right) \leq \mathbb{E}\left[\frac{1}{N_t}\right] |x - y|.$$

The process $(N_t)_{t \geq 0}$ is know to be the Yule's process. It is geometrically distributed with parameter e^{-rt}, so we have

$$\mathbb{E}\left[\frac{1}{N_t}\right] = \frac{rte^{-rt}}{1 - e^{-rt}}.$$

It ends the proof. □

Remark 5.5.10 (Generalisation of Theorem 5.5.9)**.** *In the proof of Theorem 5.5.9, we only need that, for all $n \in \mathbb{N}^*$, $\theta \in [0, 1]$,$t \geq 0$, and $x, y \in E$*

$$\sum_{j=1}^{n} |F_j^{(k)}(X_t, \theta) - F_j^{(k)}(Y_t, \theta)| \leq |x - y|$$

where X, Y are generated by G and start respectively from x, y. For instance, we can consider that X is a continuous lévy process and the division is a sub-critical fragmentation; namely

$$\forall x \in E,\ \forall k \in \mathbb{N}^*, \forall j \leq k,\ F_j^{(k)}(x, \Theta) = \Theta_j^k x,$$

where $(\Theta_j^k)_{j,k}$ is a family of random variable verifying

$$\sum_{j=1}^{k} \Theta_j^k \leq 1 \ \textit{and}\ \forall j \in \{1, \ldots, k\},\ \ \Theta_j^k \in [0, 1].$$

Even if we do not find an explicit bound, we are able to prove a Wasserstein convergence.

Lemma 5.5.11 (Wasserstein convergence)**.** *Under the assumptions of Theorem 5.5.9, we have*

$$\lim_{t \to +\infty} W_{|\cdot|}\left(\frac{\mathbf{Z}_t}{N_t}, \pi\right) = 0 \textit{ in probability}.$$

Proof. As $x \mapsto 1+x$ is a Lyapounov function for the auxiliary process, we have

$$\lim_{t\to+\infty} \frac{\mathbf{Z}_t}{N_t}(f) = \pi(f) \text{ in probability},$$

for all function f such that $|f(x)| \leq C(1+x)$. The convergence also holds in distribution. By the Skorohod's Theorem, in another probability space, we have,

$$\lim_{t\to+\infty} \frac{\mathbf{Z}_t}{N_t}(f) = \pi(f) \text{ a.s.}$$

for all continuous bounded function and for $f(x) = x$. This convergence is equivalent to the Wasserstein convergence. Thus, by a classical argument of discreteness (Varadarajan Theorem type), we get,

$$\lim_{t\to+\infty} W_{|\cdot|}\left(\frac{\mathbf{Z}_t}{N_t}, \pi\right) = 0 \text{ a.s.}.$$

Hence, in our probability space we get that $\lim_{t\to+\infty} W_{|\cdot|}\left(\mathbf{Z}_t/N_t, \pi\right) = 0$ in distribution. And as the convergence is deterministic, we get the result. □

5.5.3 Asymmetric mitosis : Macroscopic approximation

Now, we do not assume that the division is symmetric. We assume that $F_1^{(2)}(x,\theta) = F^{-1}(\theta)x$ and $F_2^{(2)}(x,\theta) = (1-F^{-1}(\theta))x$. We recall that $F(x) = 1 - F(1-x)$. In this case, Equation (5.5) becomes

$$\partial_t n(t,x) + \partial_x n(t,x) + r(x)\, n(t,x) = 2\mathbb{E}[\frac{1}{\Theta} r(x/\Theta) n(t, x/\Theta)],$$

where $n(t,.)$ is the density of $\mathbf{X}_t$. In particular, we deduce that the following P.D.E. has a weak solution:

$$\partial_t n(t,x) + \partial_x n(t,x) + r(x) n(t,x) = \int_x^{+\infty} b(x,y) n(t,y) dy$$

where b verify the following properties:

$$b(x,y) \geq 0, b(x,y) = 0 \text{ for } y < x \tag{5.13}$$

$$\int_0^{+\infty} b(x,y)dx = 2r(y) \tag{5.14}$$

$$\int_0^{+\infty} xb(x,y)dx = yr(y) \tag{5.15}$$

$$b(x,y) = b(y-x,y). \tag{5.16}$$

This equation was studied in [Per07]. Here,

$$b(x,y) = \frac{2}{y} r(y) g(\frac{x}{y}), \tag{5.17}$$

where g is the weak density of F. We easily prove the equivalence between to verify (5.17) and (5.13 - 5.16). Our aim in this section is to describe the limit of the fluctuation process. It is defined by:

$$\forall t \in [0,T], \forall n \in \mathbb{N}^*,\ \eta_t^{(n)} = \sqrt{n}(\mathbf{X}_t^{(n)} - \mathbf{X}_t).$$

Theorem 5.5.12 (Central limit Theorem for asymmetric size-structured population)**.** *Let $T > 0$. Assume that $\eta_0^{(n)}$ converges in distribution and that*

$$\mathbb{E}\left[\sup_{n\geq 1} \int_E (1+x^2)\, \mathbf{X}_0^{(n)}(dx)\right] < +\infty. \tag{5.18}$$

Then the sequence $(\eta^{(n)})_{n\geq 1}$ converges in $\mathbb{D}([0,T], C^{-2,0})$ to the unique solution of the evolution equation: for all $f \in C^{2,0}$,

$$\begin{aligned} \eta_t(f) &= \eta_0(f) \\ &+ \int_0^t \int_0^{+\infty} f'(x) + r(x) \left(\int_0^1 f(qx) + f((1-q)x) F(dq) - f(x) \right) \eta_s(dx) ds \\ &+ \widetilde{\mathbf{M}}_t(f), \end{aligned} \tag{5.19}$$

where $\widetilde{\mathbf{M}}(f)$ is a martingale and a Gaussian process with bracket:

$$\langle \widetilde{\mathbf{M}}(f) \rangle_t = \int_0^t \int_0^{+\infty} 2f'(x)f(x) + 2r(x) \int_0^1 (f(qx) - f(x))^2 F(dq)\, \mathbf{X}_s(dx) ds.$$

And $C^{2,0}$ is the set of C^2 functions, such that f, f', f'' vanish to zero when x tends to infinity. $C^{-2,0}$ is its dual space.

Lemma 5.4.1 gives

$$\forall\, t \geq 0,\ \eta_t^{(n)} = \eta_0^{(n)} + \widetilde{\mathbf{V}}_t^{(n)} + \widetilde{\mathbf{M}}_t^{(n)},$$

where for any $f \in C_c^2(E)$,

$$\widetilde{\mathbf{V}}_t^{(n)}(f) = \int_0^t \int_0^{+\infty} f'(x) + r(x) \left(\int_0^1 f(qx) + f((1-q)x) F(dq) - f(x) \right) \eta_s^{(n)}(dx) ds,$$

and $\widetilde{\mathbf{M}}^{(n)}$ is a martingale with bracket:

$$\langle \widetilde{\mathbf{M}}^{(n)}(f)\rangle_t = \int_0^t \int_0^{+\infty} 2r(x) \int_0^1 (f(qx) - f(x))^2 F(dq)\, \mathbf{X}_s^{(n)}(dx)\, ds. \tag{5.20}$$

As the set of signed measure is not metrizable, we can not adapt the proof of Theorem 5.1.2. Following [MŐ98, Tra06], we consider $\eta^{(n)}$ as an operator in a Sobolev space, and use the Hilbertian properties of this space to prove tightness. See for instance [Mét84] for condition to prove tightness on Hilbert spaces. Let us explain the Sobolev space that we will use. Let $p > 0$ and $j \in \mathbb{N}$. The set $W^{j,p}$ is the closure of C_c^∞, which is the set of functions of class C^∞ from $\mathbb{R}_+$ into $\mathbb{R}$ with compact support, embedded with the following norm:

$$\forall f \in W^{j,p},\ \|f\|^2_{W^{j,p}} = \sum_{k=0}^{j} \int_0^\infty \left(\frac{f^{(k)}(x)}{1+x^p}\right)^2 dx.$$

The set $W^{j,p}$ is an Hilbert space and we denote by $W^{-j,p}$ its dual space. Let $C^{j,p}$ be the space of function f of class C^j such that:

$$\forall k \le j,\ \lim_{x\to+\infty} \frac{f^{(k)}(x)}{1+x^p} = 0.$$

We embed it with the following norm:

$$\forall f \in C^{j,p},\ \|f\|_{C^{j,p}} = \sum_{k=0}^{j} \sup_{x\ge 0} \frac{f^{(k)}(x)}{1+x^p}.$$

The set $C^{j,p}$ is also a Banach space and we denote by $C^{-j,p}$ its dual space. These spaces verify the following continuous injection [MŐ98, Ada75]:

$$C^{j,p} \subset W^{j,p+1} \text{ and } W^{1+j,p} \subset C^{j,p}. \tag{5.21}$$

Or equivalently, if for every function f, we have

$$\|f\|_{W^{j,p+1}} \le C\|f\|_{C^{j,p}} \text{ and } \|f\|_{C^{j,p}} \le C\|f\|_{W^{j+1,p}}.$$

The first embedding/inequality prove that the tightness in $W^{j,p+1}$ implies the tighness in $C^{j,p}$. The second is useful for some upper bounds. For instance, we have

Lemma 5.5.13. *If $(e_k)_{k\geq 1}$ is a basis of $W^{2,1}$ then we have, for all $k \geq 0$ and $x \in E$,*

$$\sum_{k\geq 1} e_k(x)^2 \leq C(1+x^2).$$

Proof. $\delta_x : f \mapsto f(x)$ is an operator on $W^{2,1}$. We have, for all $f \in W^{2,1}$,

$$|\delta_x f| \leq (1+x)\|f\|_{C^{0,1}} \leq C(1+x)\|f\|_{W^{1,1}} \leq C(1+x)\|f\|_{W^{2,1}}$$

But, by Parseval's identity we get,

$$\|\delta_x\|^2_{W^{-2,1}} = \sum_{k\geq 1} e_k(x)^2,$$

which completes the proof. □

We introduce the trace $\left(\langle\langle \widetilde{\mathbf{M}}^{(n)}\rangle\rangle_t\right)_{t\geq 0}$ of $\left(\widetilde{\mathbf{M}}^{(n)}_t\right)_{t\geq 0}$. It is defined such that

$$\left(\|\widetilde{\mathbf{M}}^{(n)}_t\|^2_{W^{-2,1}} - \langle\langle \widetilde{\mathbf{M}}^{(n)}\rangle\rangle_t\right)_{t\geq 0}$$

is a local martingale. Then since

$$\left\|\widetilde{\mathbf{M}}^{(n)}_t\right\|^2_{W^{-2,1}} = \sum_{k\geq 1} \widetilde{\mathbf{M}}^{(n)}_t(e_k),$$

where $(e_k)_{k\geq 1}$ is a basis of $W^{2,1}$. Then by (5.20), we get

$$\langle\langle \widetilde{\mathbf{M}}^{(n)}\rangle\rangle_t = \sum_{k\geq 1} \int_0^t \int_0^{+\infty} 2r(x) \int_0^1 (e_k(qx) - e_k(x))^2 F(dq)\mathbf{X}^{(n)}_s(dx)ds.$$

Now, we first prove the tightness of $(\eta^{(n)})_{n\geq 1}$ then Theorem 5.5.12

Lemma 5.5.14. *$(\eta^n)_{n\geq 1}$ is tight in $\mathbb{D}([0,T], W^{-2,1})$*

Proof. By [JM86, Theorem 2.2.2] and [JM86, Theorem 2.3.2] (see also [Mǿ98, Lemma C]), it is enough to prove

1. $\mathbb{E}\left[\sup_{s\leq t}\|\eta^n_s\|^2_{W^{-2,1}}\right] < +\infty$,
2. $\forall n \in \mathbb{N},\ \forall \varepsilon, \rho > 0,\ \exists \delta > 0$ such that for each stopping times S_n bounded by T

$$\limsup_{n\to+\infty} \sup_{0\leq u\leq \delta} \mathbb{P}\left(\left\|\widetilde{\mathbf{V}}^{(n)}_{S_n+u} - \widetilde{\mathbf{V}}_{S_n}\right\|_{W^{-2,1}} \geq \eta\right) \leq \varepsilon,$$

$$\limsup_{n\to+\infty}\sup_{0\le u\le\delta}\mathbb{P}\left(\left|\left\langle\left\langle\widetilde{\mathbf{M}}^{(n)}\right\rangle\right\rangle_{S_n+u}-\left\langle\left\langle\widetilde{\mathbf{M}}^{(n)}\right\rangle\right\rangle_{S_n}\right|\ge\eta\right)\le\varepsilon.$$

For the first point, using lemma 5.5.1, there exists $C>0$ such that

$$\begin{aligned}\sum_{k\ge1}\left\langle\widetilde{\mathbf{M}}_t^{(n)}(e_k)\right\rangle &\le\int_0^t 2\bar{r}\int_0^1 2\sum_{k\ge1}e_k^2(qx)+2\sum_{k\ge1}e_k^2(x)\,F(dq)\mathbf{X}_s^{(n)}(dx)ds\\ &\le C\mathbf{X}_0^{(n)}(1+x).\end{aligned}$$

Then, since

$$\left\|\widetilde{\mathbf{M}}_t^{(n)}\right\|_{W^{-2,1}}^2=\sum_{k\ge1}\left(\widetilde{\mathbf{M}}_t^{(n)}(e_k)\right)^2,$$

Doob's inequality and (5.18) gives

$$\mathbb{E}\left[\sup_{t\in[0,t]}\left\|\widetilde{\mathbf{M}}_t^{(n)}\right\|_{W^{-2,1}}^2\right]\le C',$$

where $C'>0$. Then there exits $C''>0$ such that

$$\left\|\eta_t^{(n)}\right\|_{W^{-2,1}}^2\le\left\|\eta_0^{(n)}\right\|_{W^{-2,1}}^2+\left\|\widetilde{\mathbf{V}}_t^{(n)}\right\|_{W^{-2,1}}^2+\left\|\widetilde{\mathbf{M}}_t^{(n)}\right\|_{W^{-2,1}}^2\le C''+\left\|\widetilde{\mathbf{V}}_t^{(n)}\right\|_{W^{-2,1}}^2.$$

And as

$$\left\|\widetilde{\mathbf{V}}_t^{(n)}\right\|_{W^{-2,1}}^2\le C\int_0^t\sup_{w\le s}\left\|\eta_s^{(n)}\right\|_{W^{-2,1}}^2ds,$$

the Gronwall Lemma gives

$$\mathbb{E}\left[\sup_{s\le t}\left\|\eta_s^{(n)}\right\|_{W^{-2,1}}^2\right]\le K,$$

for a certain constant K. Finally for the second point, we have

$$\begin{aligned}\mathbb{E}\left[\left\|\widetilde{\mathbf{V}}_{S_n+u}^{(n)}-\widetilde{\mathbf{V}}_{S_n}^{(n)}\right\|_{W^{-2,1}}\right]&\le\mathbb{E}\left[K'\int_{S_n}^{S_n+u}\sup_{s\le T}\left\|\eta_s^{(n)}\right\|_{W^{-2,1}}^2\right]\\ &\le K''u.\end{aligned}$$

Here K',K'' are two constants. Using the Markov-Chebyshev inequality, we prove the Aldous condition. We similarly prove that $\langle\langle\widetilde{\mathbf{M}}^{(n)}\rangle\rangle$ verifies the Aldous condition. We deduce that $(\eta^{(n)})_{n\ge1}$ is tight. □

Proof of Theorem 5.5.12. Let $\widetilde{\mathbf{M}}$ be a continuous Gaussian process with quadratic variation verifying,

for every $f \in C^{2,0}$ ($\subset W^{2,1}$) and $t \in [0,T]$,

$$\langle \widetilde{\mathbf{M}}(f) \rangle_t = \sum_{k \geq 1} \int_0^t \int_0^{+\infty} 2r(x) \int_0^1 (f(qx) - f(x))^2 F(dq) \mathbf{X}_s(dx).$$

Since there exists C_f such that

$$\forall f \in C^{2,0}, \ \sup_{t \in [0,T]} |\widetilde{\mathbf{M}}^{(n)}(f)| \leq \frac{C_f}{\sqrt{n}},$$

and $\langle \widetilde{\mathbf{M}}^{(n)} \rangle_t$ converges in law to $\langle \widetilde{\mathbf{M}} \rangle_t$, then by [JS03, Theorem 3.11 p.473], $\widetilde{\mathbf{M}}^{(n)}(f)$ converges to $\widetilde{\mathbf{M}}(f)$ in distribution, as n tends to ∞.

By Lemma 5.5.14 and (5.21) , the sequence $(\eta^{(n)})_{n \geq 1}$ is also tight in $C^{-2,0}$. Let η be an accumulation point. Since its martingale part $\widetilde{\mathbf{M}}$ in its Doob's decomposition is almost surely continuous, then η is also almost surely continuous. Hence, η is a solution of (5.19). Using Gronwall's inequality, we obtain the uniqueness of this equation, in $C([0,T], C^{-2,0})$, up to a Gaussian white noise $\widetilde{\mathbf{M}}$. We deduce the announced result. □

5.6 Another two examples

5.6.1 Space-structured population model

Here, we study an example which can models the cells localisation. One cell moves following a diffusion on $E \subset \mathbb{R}^d$, $d \geq 1$, and when it dies, its offspring is localised at the same place. Hence, in all this section the branching is local; that is

$$\forall k \geq 0, \forall j \leq k, \forall x \in E, \forall \theta \in [0,1], \ F_j^{(k)}(x, \theta) = x.$$

5.6.1.1 Branching Ornstein Uhlenbeck

In this subsection, we consider the model of [EHK10, Example 10]. Assume that $E = \mathbb{R}^d$ and G is given by

$$Gf(x) = \frac{1}{2}\sigma^2 \Delta f(x) - gx.\nabla f(x),$$

for every $f \in C_c^2(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$, where $d \in \mathbb{N}^*$ and $\sigma, g > 0$. Also assume that the division is dyadic, that is $p_2 = 1$, with rate

$$r(x) = bx^2 + a,$$

where $a, b \geq 0$ and a or b is not null. Here $x^2 = \|x\|^2 = x.x$. If $g > \sqrt{2b}$ then we add the following notations:

$$\Gamma = \frac{g - \sqrt{g^2 - 2b\sigma^2}}{2\sigma^2} \text{ and } \alpha = \sqrt{g^2 - 2b\sigma^2}.$$

We also denote by π_∞ the Gaussian measure whose density is defined by

$$x \mapsto \left(\frac{\alpha}{\pi\sigma^2}\right) \exp\left(-\frac{\alpha}{\sigma^2}x^2\right).$$

From our main theorem, we deduce

Corollary 5.6.1 (Limit theorem for an branching Ornstein-Uhlenbeck process)**.** *If $g > \sigma\sqrt{2b}$ and $X_0^\emptyset = x \in \mathbb{R}^d$ then, for any continuous and bounded f, we have*

$$\lim_{t\to+\infty} \frac{1}{N_t} \sum_{u\in\mathcal{V}_t} f(X_t^u) = \frac{\int_{\mathbb{R}^d} f(y)e^{\Gamma y^2}\pi_\infty(dy)}{\int_{\mathbb{R}^d} e^{\Gamma y^2}\pi_\infty(dy)},$$

in probability. In particular,

$$\mathbb{E}[N_t] = e^{\lambda t+\Gamma x^2}\left(\frac{\alpha}{\pi\sigma^2}\right)\int_{\mathbb{R}^d} e^{-\Gamma y^2}\exp\left(-\frac{\alpha(y - xe^{-\alpha t/\sigma^2})^2}{\sigma^2(1 - e^{-2\alpha t/\sigma^2})}\right)dy,$$

where $\lambda = \frac{g-\sqrt{g^2-2b\sigma^2}}{2} + a$ is the Malthus parameter.

Proof. If $V : x \mapsto e^{\lambda x^2}$ then it is an eigenvector of $\mathcal{G}$, which is defined for every $f \in C_c^2(E)$ by

$$\mathcal{G}f(x) = Gf(x) + r(x)f(x).$$

We conclude the proof using Theorem 5.1.1 and Lemma 5.3.4. □

Remark 5.6.2 (Another eigenelement)**.** *Note that if $V_2 : x \mapsto e^{\lambda_2 x^2}$ then it is an eigenvector of $\mathcal{G}$, associated to the eigenvalue*

$$\lambda_2 = \frac{g + \sqrt{g^2 - 2b\sigma^2}}{2} + a.$$

But in this case, the auxiliary process is not ergodic and we are not able to deduce any convergence from our main theorem.

5.6.1.2 General case

Let us assume that G is the generator of a diffusive Markov process. If the state space E is bounded then we can find sufficient conditions to the eigenproblem in [Pin95, section 3] and [Pin95,

Theorem 5.5]. For instance, under some assumptions, we have

$$\lambda_0 = \lim_{t\to+\infty} \ln\left(\sup_{x\in E} \mathbb{E}\left[N_t \mid X_0^{\emptyset} = x\right]\right).$$

If E is not bounded then we can see [HS96, RS78]. This example is developed in [EHK10]. They prove a strong law of large number, which is close to Theorem 5.1.1.

5.6.2 Self-similar fragmentation

Self-similar mass fragmentation processes are characterised by

- the index of self-similarity $\alpha \in \mathbb{R}$;
- a so-called dislocation measure ν on $\mathcal{S} = \{s = (s_i)_{i\in\mathbb{N}} \mid \lim_{i\to+\infty} s_i = 0, 1 \geq s_j \geq s_i \geq 0, \forall j \leq i\}$ which satisfies

$$\nu(1,0,0,..) = 0 \text{ and } \int_{\mathcal{S}} (1-s)\nu(ds) < +\infty.$$

If $\nu(\mathcal{S}) < +\infty$ then the dynamics is as follows:

- a block of mass x remains unchanged for exponential periods of time with parameter $x^\alpha \nu(\mathcal{S})$;
- a block of mass x dislocates into a mass partition xs, where $s \in \mathcal{S}$, at rate $\nu(ds)$;
- there are finitely many dislocations over any finite time horizon.

The last point is not verified when $\nu(\mathcal{S}) = +\infty$. In this case, there is a countably infinite number of dislocations over any finite time horizon. So, when $\nu(\mathcal{S}) < +\infty$, our setting capture this model with the following parameters:

$$G = 0,\ r(x) = x^\alpha \nu(\mathcal{S}),$$

and for every continuous and bounded function f,

$$\int_0^1 \sum_{k\geq 0} p_k(x) \sum_{j=1}^{k} f(F_j^k x, \theta) d\theta = \int_{\mathcal{S}} \sum_{i\geq 0} f(s_i x) \frac{\nu(ds)}{\nu(\mathcal{S})}.$$

Hence, in this case we have

$$\mathcal{G}f(x) = x^\alpha \nu(\mathcal{S}) \sum_{i\geq 0} \left(\int_{\mathcal{S}} \sum_{i\geq 0} f(s_i x) \frac{\nu(ds)}{\nu(\mathcal{S})} - f(x) \right),$$

for every continuous and bounded f, and $V : x \mapsto x^p$ is an eigenevector. See [Ber06] for further details.

Chapter 6

Asymptotic estimates for the largest individual in a mitosis model

6.1 Introduction and statement of result

In this note, we consider a growth-fragmentation model, in continuous time, to represent dividing cells. The dynamics is as follows. The size X_t of a cell, at time t, evolves according to the following stochastic differential equation

$$\forall t \geq 0, \quad dX_t = (\mu + \sigma^2)X_t dt + \sigma\sqrt{2}X_t dB_t, \tag{6.1}$$

where $\mu \in \mathbb{R}$ and $\sigma \geq 0$. If σ is equal to zero then the evolution is deterministic. At rate $r > 0$, each cell splits into two offspring whose size is worth half of that of their mother. This model is entirely determined by the parameters μ, σ and r. The parameters μ and σ depend on the ability of cells to ingest a common nutrient. The parameter r is the division rate. We assume that is is constant. More generally, instead of size of cell, this model can represent some biological content which grows in the cells and is shared when the cells divide (for example proteins, nutriments, energy or parasite). The process we study is a Markov process on Galton Watson trees and [BDMT11, Clo11] give asymptotic results, under an ergodicity assumption, which is not fulfilled here. The article [Clo11] also shows that the empirical process converges, when the size of the population tends to infinity, to the following

physiologically structured equation:

$$\forall t \geq 0,\ \forall x \geq 0,\ \partial_t n(t,x) + (\mu + \sigma^2)\partial_x(xn(t,x)) + rn(t,x) = 4rn(t,2x) + \sigma^2 \Delta n(t,x).$$

This type of equation has been recently studied in [DHKR12, DPZ09, Per07], for instance. From a probabilistic point of view, the behaviour of one cell is known. It is the exponential of a Lévy process and thus, using classical results [Ber96], we deduce its long time behavior. There is a duality: either the process explodes or it vanishes. Closely related, [BT11] studied a population of infected cell. Between the division, the parasites grow following a Feller diffusion. The main difference between our model and theirs is the fact that, in our model, the size of the cells can not be null even if it tends to zero. Here, we are interested by the properties of the largest individual. This question has been intensively studied in the special case of Branching Brownian motion [Bra78, HHK06, McK75, Rob11] and fragmentation processes [BHK11, Ber04, Ber06]. Due to the spacial motion (exponential growth) and the non-local branching mechanism (the offspring do not appear at the position of their mother), the mathematical study is different. There is a competition between the exponential growth of size, between the division, and the multiplicative decreasing, at the division. One of purpose of our main results is to highlight this competition.

Let us give now a qualitative description of our model, which is rigorously defined in the next section. We consider a continuous time Yule tree, which is, a tree where each branch lives during an independent exponential time of mean $1/r$ and then splits into two new branches. We denote by N_t the size of the living population set V_t, at time $t \geq 0$, and by $(X_t^u)_{u \in V_t}$ the size of the individual $u \in V_t$. We aim at determining how the maximum $\overline{X}$ evolves in the cell population:

$$\forall t \geq 0,\ \overline{X}_t = \max_{u \in V_t} X_t^u.$$

Now, we can state our main results:

Theorem 6.1.1 (Asymptotic estimates of the size of the largest individual)**.** *We have*

$$\lim_{t \to +\infty} \frac{1}{t} \ln \overline{X}_t = \inf_{\alpha \geq 0} \mu + \sigma^2 \alpha + r \frac{2e^{-\alpha \ln(2)} - 1}{\alpha}.$$

The infimum in the right hand side of the last expression is attained in a unique $\alpha \geq 0$. This theorem gives the long time behavior of the extremal particle. To compare, we also give the mean behavior of the population:

Theorem 6.1.2 (Mean behavior of the population)**.** *We have the following duality:*

(i) if $\mu < 2r\ln(2)$ *then for all* $\varepsilon > 0$*, we have*

$$\lim_{t\to+\infty} \frac{\textit{card}\left\{u \in V_t \mid X_t^u \geq \varepsilon\right\}}{N_t} = 0 \textit{ a.s.}$$

(ii) If $\mu > 2r\ln(2)$ *then for every* $0 \leq \kappa < g - 2r\ln(2)$

$$\mathbb{P}\left(\limsup_{t\to\infty} \frac{\textit{card}\{u \in V_t : X_t^u \geq e^{\kappa t}\}}{N_t} > 0\right) = 1 \tag{6.2}$$

These two theorems give sharp estimates. Nevertheless, this model is too simple and it will be efficient to generalize the first one in one of the following generalization:

– the parameter σ is null but r is not constant as in [DHKR12];
– the growth between two divisions is linear instead of exponential as in [Clo11];
– the growth between two divisions is described by a Feller diffusion as in [BT11].

Unfortunately, even if r can be constant in the two last models, these setting are very different to ours. Our model is relatively close to branching random walk contrary to these models.

Outline: in the next section, we give more formal definitions based on measure-valued processes and state the asymptotic behavior for geometric Brownian motion with multiplicative jumps. This gives the asymptotic behavior of the size in a cell line. We introduce, in Section 6.2.2, the scaling property, and, in Section 6.2.3, some martingales. Theorem 6.1.1 is proved in Section 6.3 and Theorem 6.1.2 in Section 6.4. In the last section, we give some comments about the link between the fluctuation of the extremal particle and a F-KPP type equation.

6.2 Preliminaries

To describe the cell population and label its nodes, we use the Ulam-Harris-Neveu notation. Let

$$\mathcal{T} = \bigcup_{n\geq 0} \{1,2\}^n,$$

be the set of labels, where $\{1,2\}^0 = \{\emptyset\}$. The first cell is labelled by $\emptyset$ and when the cell $u = (u_1,\ldots,u_n) = u_1\ldots u_n \in \mathcal{T}$ dies, it divides in two offspring labelled by $u0 = (u_1,\ldots,u_n,0)$ and $u1 = (u_1,\ldots,u_n,1)$. Let $V_t \subset \mathcal{U}$ be the set of cells alive at time t and $N_t = \text{card}(V_t)$ be the number of cells alive at this time. The process $(N_t)_{t\geq 0}$ is well known and it is called the Yule process. We have:

Lemma 6.2.1 (Properties of the Yule process)**.** *For all $t \geq 0$, N_t is a geometric random variable, with parameter e^{-rt}, and the following limit holds almost surely and in L^2*

$$\lim_{t\to+\infty} \frac{N_t}{\mathbb{E}[N_t]} = W,$$

where W is exponential with mean 1

Proof. See [AN04, Har02]. □

For $u \in V_t$, we denote by X_t^u the size the cell u at time t and by $\mathbf{Z}$ the empirical measure which describes the population; namely

$$\forall t \geq 0,\ \mathbf{Z}_t = \sum_{u\in V_t} \delta_{X_t^u}.$$

It is a measure-valued Markov process. Let $C_c^{2,1}(\mathbb{R}_+^*, \mathbb{R}_+)$ be the set of function $f : (t, x) \mapsto f_t(x)$, that are C^1 with bounded derivative in their first component and, for all $t \geq 0$, f_t is C^2 with compact support. By construction, the process $(\mathbf{Z}_t)_{t\geq 0}$ verifies the following stochastic differential equation:

$$\mathbf{Z}_t(f_t) = \mathbf{Z}_0(f_0) + \int_0^t \int_0^\infty \partial_s f_s(x) + \mathcal{L} f_s(x) \mathbf{Z}_s(dx)ds \tag{6.3}$$

$$+ \int_0^t \sum_{u\in V_s} \sigma\sqrt{2} B_s \partial_x f_s(x) dB_s^{(u)} \tag{6.4}$$

$$+ \int_0^t \int_{\mathcal{U}} \mathbf{1}_{u\in V_{s-}} \left(2 f_s\left(\frac{x}{2}\right) - f_s(x)\right) \mathcal{N}(ds, du),$$

for every function $f \in C_c^{2,1}(\mathbb{R}_+^*, \mathbb{R}_+)$, where $(B^{(u)})_{u\in\mathcal{T}}$ is a family of independent Brownian motions, $\mathcal{N}(ds, du)$ is a Poisson point measure of intensity $r\ ds\ n(du)$ and

$$\mathcal{L} f(x) = (\mu + \sigma^2) x f'(x) + \sigma^2 x^2 f''(x),$$

for any $f \in C_c^{2,1}(\mathbb{R}_+^*, \mathbb{R}_+)$ and $x \in \mathbb{R}_+$. Here, ds is the Lebesgue measure on $\mathbb{R}_+$ and $n(du)$ the counting measure on $\mathcal{U}$. Since r is constant, if $\mathbf{Z}_0 = \delta_x$ with $x \geq 0$, then there is no explosion and there exists a unique solution defined on $\mathbb{R}_+$ [IW89]. In particular we have

$$\mathbb{E}[\mathbf{Z}_t(f_t)] = \mathbb{E}[\mathbf{Z}_0(f_0)] + \int_0^t \mathbb{E}\left[\mathbf{Z_s}\left[\partial_s f_s + \mathcal{G} f_s\right]\right] ds, \tag{6.5}$$

where

$$\mathcal{G} f(x) = \mathcal{L} f(x) + r\left(2f\left(\frac{x}{2}\right) - f(x)\right),$$

for every $f \in C_c^2(\mathbb{R}_+^*)$, where $f \in C_c^2(\mathbb{R}_+^*)$ is the set of C^2 functions with compact support.

6.2.1 Behavior of one cell line

We are interested in the evolution of the size in a cell line. This means that at each division, we only keep one cell and consider its size. This process $(X_t)_{t\geq 0}$ is distributed as the exponential of a Lévy process. More precisely, we have, for all $t \geq 0$, $X_t = e^{L_t}$, where

$$L_t = \sqrt{2}\sigma B_t + \mu t - N_t \ln(2),$$

$(N_t)_{t\geq 0}$ is a Poisson process with intensity r and $(B_t)_{t\geq 0}$ is a standard Brownian motion. With this representation, we deduce that

$$\mathbb{E}[X_t] = e^{\mu t}\mathbb{E}\left[\frac{1}{2^{N_t}}\right]\mathbb{E}\left[e^{\sqrt{2}\sigma B_t}\right] = e^{t(\sigma^2+\mu-r/2)}$$

So, the size of one cell turn over $e^{t(\sigma^2+\mu-r/2)}$. Furthermore we easily prove that $\left(\frac{X_t}{\mathbb{E}[X_t]}\right)_{t\geq 0}$ is a positive martingale, thus it converges, almost surely, to a non-negative variable χ. Notice also that

$$\forall t \geq 0,\ \mathbb{E}\left[\left(\frac{X_t}{\mathbb{E}[X_t]}\right)^2\right] = e^{t(2\sigma^2+r/4)},$$

and there is no an L^2-convergence. Using classical result on Lévy processes, we have

Lemma 6.2.2 (Behavior of one cell line)**.** *If $\mu < r\ln(2)$ then*

$$\lim_{t\to+\infty} X_t = 0 \text{ a.s.}$$

If $\mu > r\ln(2)$ then

$$\lim_{t\to+\infty} X_t = +\infty \text{ a.s.}$$

If $\mu = r\ln(2)$ then X oscillates between 0 and $+\infty$.

Proof. it is an application to [Ber96, corollary 2 p.190] applied on L.

□

This lemma gives that if $r/2 < \mu < r\ln(2)$ then $X_t \to 0$ and $\mathbb{E}[X_t] \to +\infty$. In particular, in this case, $\chi = 0$ a.s.

6.2.2 Scaling property and application to the study of the size of the largest individual

Our model is close to branching random walks, branching Brownian motion or fragmentation processes. One of the main common features is the following scaling property:

Lemma 6.2.3 (Scaling property)**.** *We have*

$$\mathcal{L}aw((\mathbf{Z}_t)_{t\geq 0} \mid \mathbf{Z}_0 = \delta_x) = \mathcal{L}aw((\mathbf{Z}_t^{(x)})_{t\geq 0} \mid \mathbf{Z}_0 = \delta_1),$$

where for every $t \geq 0$*,*

$$\mathbf{Z}_t^{(x)} = \sum_{u\in V_t} \delta_{x\times X_t^u}.$$

Proof. This comes directly from the construction, the stochastic differential equation (6.3) or the generator (6.5). □

As a consequence, most of the time, we shall assume, without less of generality, that $\mathbf{Z}_0 = \delta_1$. So, we add the following notation: for $x > 0$, we define $P_x(A) = \mathbb{P}(A|\mathbf{Z}_0 = \delta_x)$, for all measurable set A, and denote by $\mathbb{E}_x$ the corresponding expectation. As application of the previous lemma, we have the following result, adapted from the study of branching Brownian motion [Rob09, Theorem 1.7]:

Lemma 6.2.4 (Law of 0-1 for the largest individual)**.** *For every* $\lambda \in \mathbb{R}$*,*

$$\mathbf{p} = \mathbb{P}_1\left(\liminf_{t\to+\infty} \frac{\ln(\overline{X}_t)}{t} < \lambda\right) \in \{0, 1\}.$$

Proof. Let us denote by $\mathcal{F}_s = \sigma\{(X_t^u)_{u\in V_t}, t \leq s\}$. For all $s \geq 0$, we have

$$\begin{aligned}
P(s) &= \mathbb{P}_1\left(\liminf \frac{\ln(\max_{u\in V_t} X_t^u)}{t} < \lambda \mid \mathcal{F}_s\right) \\
&= \mathbb{P}_1\left(\bigcap_{u\in V_s}\left\{\liminf \frac{\ln(\max_{v\in uV_t} X_t^v)}{t} < \lambda\right\} \mid \mathcal{F}_s\right) \\
&= \prod_{u\in V_s} \mathbb{P}_{X_s^u}\left(\liminf \frac{\ln(\max_{v\in V_t} X_t^v)}{t} < \lambda\right) \\
&= \mathbb{P}_1\left(\liminf \frac{\ln(\max_{v\in V_t} X_t^v)}{t} < \lambda\right)^{N_s} \\
&= \mathbf{p}^{N_s}
\end{aligned}$$

where $uV_t = \{v \in V_t \mid \exists w \in \mathcal{U},\ v = uw\}$ and the passage between the third and the fourth line comes from Lemma 6.2.3. Thus $P(s)$ converges to 0 or 1 following that $\mathbf{p} < 1$ or $\mathbf{p} = 1$. However we have $\mathbb{E}[P(s)] = \mathbf{p}$ and so $\mathbf{p} \in \{0, 1\}$. □

Immediately, we deduce that there exists $\lambda_* \in \mathbb{R}$ such that if $\lambda > \lambda_*$ then

$$\liminf_{t\to+\infty} \overline{X}_t e^{-\lambda t} = 0,$$

and if $\lambda < \lambda_*$ then

$$\lim_{t\to+\infty} \overline{X}_t e^{-\lambda t} = +\infty.$$

With the same method, we prove that there exists $\lambda^* \in \mathbb{R}$ such that if $\lambda > \lambda^*$ then

$$\lim_{t\to+\infty} \overline{X}_t e^{-\lambda t} = 0,$$

and if $\lambda < \lambda^*$ then

$$\limsup_{t\to+\infty} \overline{X}_t e^{-\lambda t} = +\infty.$$

then we see that it is impossible that $\lambda_* > \lambda^*$, and we have two possibilities: either $\lambda_* = \lambda^*$ and then $\overline{X}_t \approx e^{\lambda t}$; either $\lambda_* < \lambda^*$ and then, if $\lambda < \lambda_*$ then we have

$$\lim_{t\to+\infty} e^{-\lambda t} = +\infty$$

if $\lambda > \lambda^*$ then

$$\lim_{t\to+\infty} e^{-\lambda t} = 0$$

and if $\lambda_* < \lambda < \lambda^*$ we have

$$\limsup_{t\to+\infty} e^{-\lambda t} = +\infty \text{ and } \liminf_{t\to+\infty} e^{-\lambda t} = 0.$$

We shall prove, with another method and without to use this lemma, that $\lambda_* = \lambda^*$.

6.2.3 Additive martingales

In all this section, we shall suppose that $X_0^{\emptyset} = 1$. We begin by remarking that

$$\mathcal{G} g_\alpha = \lambda_\alpha g_\alpha, \tag{6.6}$$

where

$$\lambda_\alpha = \mu\alpha + \sigma^2\alpha^2 + r(2^{1-\alpha} - 1),$$

$g_\alpha : x \mapsto x^\alpha$ and $\mathcal{G}$ is defined in the expression (6.5). In others words, $(g_\alpha, \lambda_\alpha)_{\alpha\in\mathbb{R}}$ is a family of eigenelements of $\mathcal{G}$. As a consequence, we have

Lemma 6.2.5 (A family of martingales)**.** *Let $\alpha \in \mathbb{R}$ and let us define, for every $\alpha \in \mathbb{R}$ and $t \geq 0$,*

$$M_t^{(\alpha)} = e^{-\lambda_\alpha t} \sum_{u\in V_t} (X_t^u)^\alpha = e^{-\lambda_\alpha t} \mathbf{Z}_t(g_\alpha).$$

It is a positive martingale which converges to a random variable $M_\infty^{(\alpha)}$.

Proof. Using (6.6), we see that the expectation of $(M_t^{(\alpha)})_{t\geq 0}$ is constant. Thanks to the branching property, this gives the martingale property. See [Clo11, Lemma 3.2], for further details. Finally, it is positive and then converges almost surely to a random variable $M_\infty^{(\alpha)}$. □

The mapping $\alpha \mapsto \lambda_\alpha$ is similar to the function $-\kappa$ introduced in [Ber06, Chapter 1] for the fragmentation processes. In this book, it is proved that κ can explode and is always a continuous strictly increasing function on the set where it does not explode. Due to the increasing of the size between the divisions, our situation is different to [Ber06]. Indeed, we have the following elementary lemma

Lemma 6.2.6 (Variation of $\alpha \mapsto \lambda_\alpha$)**.** *There exists a unique $\alpha^* \in \mathbb{R}$ such that the function $\alpha \mapsto \lambda_\alpha$ is convex on $\mathbb{R}$, decreasing on $(-\infty, \alpha^*)$ and increasing on $(\alpha^*, +\infty)$. Furthermore,*

- *if $\mu - 2r\ln(2) < 0$ then $\alpha^* > 0$,*
- *if $\mu - 2r\ln(2) > 0$ then $\alpha^* < 0$.*

In particular, thanks to Theorem 6.1.2, we have two situations:

- if $\mu > 2r\ln(2)$ then the sizes of cells explode and $\alpha^* > 0$.
- if $\mu < 2r\ln(2)$ then the sizes of cells vanish and $\alpha^* < 0$. Furthermore, λ is positive and increasing on $\mathbb{R}_+$.

Now we give the behavior of another important function:

Lemma 6.2.7 (Variation of $\alpha \mapsto p\lambda_\alpha - \lambda_{p\alpha}$)**.** *There exists $p^* > 1$ such that, for every $p \in (1, p^*]$, the function $\beta^{(p)} : \alpha \mapsto p\lambda_\alpha - \lambda_{p\alpha}$ is convex and decreasing on $(0, +\infty)$. Furthermore,*

$$\beta^{(p)}(0) = r(p-1) \text{ and } \lim_{\alpha\to+\infty} \beta^{(p)}(\alpha) = r(1-p).$$

In particular, there exists $\alpha^{(p)} > 0$ *such that*

$$\forall \alpha \in (0, \alpha^{(p)}),\ \beta^{(p)}(\alpha) > 0.$$

The proofs of the two previous lemmas are straightforward, but we deduce that

Lemma 6.2.8 (L^p-converge of the additive martingales)**.** *Let* $p \in (1, p^*]$ *and* $\alpha \in (0, \alpha^{(p)})$*, the martingale* $M_t^{(\alpha)}$ *converges also in* L^p *and* $M_\infty^{(\alpha)} > 0$ *a.s.*

Proof. In order to prove that the convergence holds in L^p, we shall use the Burkholder-Davis-Gundy inequality [DM80, Theorem 92 p. 304]. For every continuous time martingale $(M_t)_{t\geq 0}$, with finite variation and which verifies $M_0 = 0$, and $p \in (1, 2]$, there exists a universal constant $C_p > 0$, such that

$$\mathbb{E}\left[\sup_{t\geq 0} |M_t|^p\right] \leq C_p \mathbb{E}\left[\sum_{t\geq 0} |M_{t+} - M_{t-}|^p\right].$$

Since $\mathbf{Z}$ is continuous between the division times, we deduce

$$\mathbb{E}\left[\sup_{t\geq 0} \left|M_t^{(\alpha)} - M_0^{(\alpha)}\right|^p\right] \leq C_p \mathbb{E}\left[\sum_{u\in\mathcal{T}} e^{-\lambda_\alpha d(u)p} \left|X_{d(u)}^u\right|^{\alpha p} \left|2^{1-\alpha} - 1\right|^p\right].$$

Now, using the fact that the lifetime of one individual is exponentially distributed, we have

$$\begin{aligned}
\sum_{u\in\mathcal{T}} \mathbb{E}\left[e^{-\lambda_\alpha d(u)p} \left|X_{d(u)}^u\right|^{\alpha p}\right] &= \sum_{u\in\mathcal{T}} \mathbb{E}\left[\int_{b(u)}^{d(u)} e^{-\lambda_\alpha tp} (X_t^u)^{\alpha p} dt\right] \\
&= \int_0^\infty \mathbb{E}\left[\sum_{u\in V_t} e^{-\lambda_\alpha tp} (X_t^u)^{\alpha p}\right] dt \\
&= \int_0^\infty e^{-t(p\lambda_\alpha - \lambda_{\alpha p})} dt \\
&< +\infty.
\end{aligned}$$

In the last expression, $b(u)$ and $d(u)$ design the birth and the death times of the individual u. Finally, let us check that $M_\infty^{(\alpha)} > 0$ almost surely. Write $\varrho = \mathbb{P}\left(M_\infty^{(\alpha)} = 0\right)$; the fact that $\mathbb{E}\left[M_\infty^{(\alpha)}\right] = \mathbb{E}\left[M_0^{(\alpha)}\right] = 1$ ensures that $\varrho < 1$. On the other hand, an application to the branching property and the L^1 convergence yields

$$M_\infty^{(\alpha)} \stackrel{d}{=} e^{-\lambda_\alpha \tau}\left(\frac{\widetilde{X}_\tau^{(1)}}{2}\widetilde{M}_\infty^{(1)} + \frac{\widetilde{X}_\tau^{(2)}}{2}\widetilde{M}_\infty^{(2)}\right),$$

where $\widetilde{M}_\infty^{(1)}, \widetilde{M}_\infty^{(2)}$ are two independent variables distributed as $M_\infty^{(\alpha)}$, $(\widetilde{X}_t^{(1)})_{t\geq 0}, (\widetilde{X}_t^{(2)})_{t\geq 0}$ are two

independent diffusive processes generated by (6.3) and τ is an independent random variable, exponentially distributed, with mean $1/r$. This gives $\varrho^2 = \varrho$ and ends the proof. □

Let us add the following notations:

$$\alpha^{(1)} = \lim_{p \to 1} \alpha^{(p)}.$$

Trivially, we have

Corollary 6.2.9. *Let $\alpha \in (0, \alpha^{(1)})$, the martingale $M_t^{(\alpha)}$ converges also in L^1 and $M_\infty^{(\alpha)} > 0$ a.s.*

6.3 Asymptotic of the largest individual

In this section, we prove Theorem 6.1.1.

6.3.1 Exponential increasing for the size of the largest individual

Lemma 6.3.1 (Variation of $\gamma : \alpha \mapsto \lambda_\alpha/\alpha$)**.** *There exists $\overline{\alpha} > 0$ such that γ is decreasing on $(0, \overline{\alpha})$ and increasing on $(\overline{\alpha}, +\infty)$. Furthermore,*

$$\frac{\lambda_{\overline{\alpha}}}{\overline{\alpha}} = \lambda'_{\overline{\alpha}}.$$

Proof. Since $\alpha \mapsto \lambda_\alpha$ is convex, we deduce the variation of $\alpha \mapsto \alpha\lambda'_\alpha - \lambda_\alpha$, and then the variation of γ. The equality holds because $\overline{\alpha}$ is a local minimum. □

Lemma 6.3.2 (Comparison of constants)**.** *We have*

$$\overline{\alpha} \leq \alpha^{(1)}$$

Proof. Firstly, for any $\alpha > 0$, we have

$$\lim_{p \to 1} \partial_p \beta^{(p)}(\alpha) = \begin{cases} \lambda_\alpha - \alpha\lambda'_\alpha < 0, & \text{if } \alpha < \overline{\alpha} \\ \lambda_\alpha - \alpha\lambda'_\alpha > 0, & \text{if } \alpha > \overline{\alpha}. \end{cases}$$

Now, for any $p \in (1, 2]$, we have

$$\beta^{(p)}(\alpha^{(p)}) = 0.$$

Differencing in p and taking the limit as $p \to 1$, we deduce

$$\partial_p \alpha^{(p)} \partial_\alpha \beta^{(p)}(\alpha^{(p)}) + \partial_p \beta^{(p)}(\alpha^{(p)}) = 0$$

Finally, as the mapping $p \mapsto \alpha^{(p)}$ is increasing and the mapping $\alpha \mapsto \beta^p(\alpha)$ is decreasing, taking the limit as $p \to 1$, we deduce the announced result. □

6.3.2 Proof of Theorem 6.1.1

For every $\alpha \in (0, \alpha^{(1)})$, we have

$$e^{-\lambda_\alpha t} \overline{X}_t^\alpha \leq e^{-\lambda_\alpha t} \sum_{u \in V_t} (X_t^u)^\alpha$$

and the right-hand side remains bounded as t tends to infinity. Hence

$$\limsup_{t \to +\infty} \frac{1}{t} \ln \overline{X}_t \leq \frac{\lambda_\alpha}{\alpha}.$$

And optimizing over α yields

$$\limsup_{t \to +\infty} \frac{1}{t} \ln \overline{X}_t \leq \frac{\lambda_{\overline{\alpha}}}{\overline{\alpha}}.$$

On the other hand, for every $\alpha \in (0, \alpha^{(1)})$ and $\varepsilon > 0$ sufficiently small, we have the lower bound

$$e^{-\lambda_\alpha t} \sum_{u \in V_t} (X_t^u)^\alpha \leq \overline{X}_t^\varepsilon e^{-\lambda_\alpha t} \sum_{u \in V_t} (X_t^u)^{\alpha - \varepsilon}.$$

We know that both limits

$$\lim_{t \to +\infty} e^{-\lambda_\alpha t} \sum_{u \in V_t} (X_t^u)^\alpha \text{ and } e^{-\lambda_{\alpha-\varepsilon} t} \sum_{u \in V_t} (X_t^u)^{\alpha - \varepsilon}$$

are finite and strictly positive a.s. and we deduce that

$$\liminf_{t \to +\infty} \frac{1}{t} \ln \overline{X}_t \geq \frac{\lambda_\alpha - \lambda_{\alpha-\varepsilon}}{\varepsilon}.$$

We take the limit of the right-hand side as $\varepsilon \to 0$ and then taking the limit $\alpha \to \overline{\alpha}$ to conclude that

$$\liminf_{t \to +\infty} \frac{1}{t} \ln \overline{X}_t \geq \lambda'_{\overline{\alpha}}.$$

6.4 Mean behavior of the population

In this section, we prove Theorem 6.1.2.

6.4.1 Many-to-one formula and auxiliary process

Using the martingales introduced in section 6.2.3, we can use a Girsanov-type transformation and deduce several many-to-one formula. In particular, the case $\alpha = 0$ leads to:

Lemma 6.4.1 (Many-to-one formula)**.** *For every $t \geq 0$ and $x \geq 0$, we have*

$$\mathbb{E}_x \left[\sum_{u \in V_t} f(X_t^u) \right] = \mathbb{E}[f(Y_t)],$$

where $(Y_t)_{t \geq 0}$ is an auxiliary Markov process, starting from $Y_0 = x$, and generated by

$$\mathcal{A}f(x) = \sigma^2 f''(x) + (\mu + \sigma^2) x f'(x) + 2r \left(f\left(\frac{x}{2}\right) - f(x) \right),$$

for every $f \in C_c^2(\mathbb{R}_+^*)$.

Proof. See [BDMT11, Proposition 3.3] or [Clo11, Lemma 3.3]. □

Note that the behavior of the auxiliary process is different to the behavior of one cell line. The resulting jump rate $2r$ is equal to the original rate r times the number of offspring. This is heuristically explained by the fact that when one chooses an individual uniformly in the population, at a certain time, an individual belonging to a lineage with more generations or with prolific ancestors is more likely to be chosen. Except for this difference, their behavior is similar and we can write

$$\forall t \geq 0, Y_t = \exp\left(\mu t - N_t \ln(2) + \sigma \sqrt{2} B_t \right),$$

where $(B_t)_{t \geq 0}$ is a standard Brownian motion and $(N_t)_{t \geq 0}$ is an independent Poisson process, with intensity $2r$, and we deduce the analogue of Lemma 6.2.2 :

Lemma 6.4.2 (Limit of the auxiliary process)**.** *If $\mu > 2r \ln(2)$ then*

$$\lim_{t \to +\infty} Y_t = +\infty \text{ a.s.}.$$

If $\mu < 2r \ln(2)$ then

$$\lim_{t \to +\infty} Y_t = 0 \text{ a.s.}.$$

and if $\mu = 2r\ln(2)$ *then* Y *oscillates*

6.4.2 Proof of Theorem 6.1.2

Let us begin by proving (i) and let us define $V_t^\varepsilon = \{u \in V_t \mid X_t^u \geq \varepsilon\}$ and $N_t^\varepsilon = \text{card}V_t^\varepsilon$. By the Lemma 6.4.2 and Lemma 6.4.1, we have

$$\lim_{t\to+\infty} \mathbb{E}\left[\frac{N^\varepsilon}{\mathbb{E}[N_t]}\right] = \lim_{t\to+\infty} \mathbb{P}\left(Y_t \geq \varepsilon\right) = 0$$

and as $N_t/\mathbb{E}[N_t]$ converges a.s. to an exponential variable, see Lemma 6.2.1, we have the convergence of $N^\varepsilon/\mathbb{E}[N_t]$ to 0 in probability. Now, it remains to show that the convergence holds a.s. Denoting by $V_{t,s}(u)$ the set of cells alive at time $t+s$ and whose ancestor at time t is the cell $u \in V_t$, we have

$$\frac{N_{t+s}^\varepsilon}{N_{t+s}} \leq \frac{\sum_{u\in V_t^\varepsilon} \text{card}V_{s,t}(u)}{\sum_{u\in V_t} \text{card}V_{s,t}(u)}.$$

Let us add the notation $\Upsilon_s^u = \text{card}V_{s,t}(u)$. Conditionally to $\mathcal{F}_t = \sigma\{\sum_{u\in V_t} \delta_{X_r^u} \mid r \leq t\}$, $(\Upsilon^u)_{u\in V_t}$ is a family of i.i.d. Yule process, hence

$$\begin{aligned}
\sup_{s\geq 0} \frac{N_{t+s}^\varepsilon}{N_{t+s}} &\leq \frac{\sum_{u\in V_t^\varepsilon} \sup_{s\geq 0} \Upsilon_s^u e^{-rs}}{\sum_{u\in V_t} \inf_{s\geq 0} \Upsilon_s^u e^{-rs}} \\
&\leq \frac{\sum_{u\in V_t^\varepsilon} \sup_{s\geq 0} \Upsilon_s^u e^{-rs}}{\sum_{u\in V_t} \inf_{s\geq 0} \Upsilon_s^u e^{-rs}} \\
&\leq \frac{N_t^\varepsilon}{N_t} \times \frac{1}{N_t^\varepsilon} \sum_{u\in V_t^\varepsilon} \sup_{s\geq 0} \Upsilon_s^u e^{-rs} \times \frac{N_t}{\sum_{u\in V_t} \inf_{s\geq 0} \Upsilon_s^u e^{-rs}}.
\end{aligned}$$

By the usual law of large number, the second and third terms converges a.s., thus we deduce that

$$\lim_{t\to+\infty} \sup_{s\geq 0} \frac{N_{t+s}^\varepsilon}{N_{t+s}} = 0 \text{ in probability.}$$

Then, we deduce the a.s. convergence of N_t^ε/N_t using the following standard argument

$$\{\exists \delta > 0, \forall t \geq 0, \exists s \geq 0, N_{t+s}^\varepsilon/N_{t+s} \geq 2\delta\} \subset \bigcup_{\delta>0} \bigcap_{t\geq 0} \{\sup N_{t+s}^\varepsilon/N_{t+s} \geq 2\delta\} \text{ a.s.}$$

where the probability of the right hand side event is equal to 0. This ends the proof. Let us now prove (ii). As $g > 2r\ln(2) \Rightarrow g > r\ln(2)$, Lemma 6.2.2 gives that the size of each cells go to infinity. Thus there always exists a random time T such that $V_t^1 = \{u \in V_t \mid X_t^u \geq 1\}$ is not empty when $t = T$.

By branching properties

$$\mathbb{P}\left(\limsup_{t\to} \text{card}V_t^1 = +\infty\right) = 1.$$

Then, for every $n \in \mathbb{N}^*$, the stopping time T_n, defined by

$$T_n = \inf\{t \geq 0 \mid \text{card}V_t^1 \geq n\},$$

for every $n \geq 1$, is finite a.s. Denoting by $V_t(j)$ the set of cells alive at time t which are issue to the cell j, we have

$$\frac{\text{card}\{i \in V_{T_n+t} : X^i_{T_n+t} \geq \exp(\kappa t)\}}{N_{T_n+t}} \geq \sum_{j\in V_{T_n}} \frac{\text{card}V_t(j)}{N_{T_n+t}} \frac{\text{card}\{j \in V_t(j) : X_t^i(j) \geq \exp(\kappa t)\}}{\#V_t(j)}.$$

Letting $t \to \infty$ in this inequality and $\text{card}V_{T_n} \geq n$ gives

$$\begin{aligned}&\mathbb{P}\left(\limsup_{t\to\infty} \frac{\text{card}\{u \in V_t : X_t^u \geq \exp(\kappa t)\}}{N_t} = 0\right)\\ &\leq \mathbb{P}_1\left(\limsup_{t\to\infty} \frac{\text{card}\{u \in V_t : X_t^u \geq \exp(\kappa t)\}}{N_t} = 0\right)^n.\end{aligned}$$

This ensures that

$$\mathbb{P}\left(\limsup_{t\to\infty} \frac{\text{card}\{u \in V_t : X_t^u \geq \exp(\kappa t)\}}{N_t} = 0\right) \in \{0, 1\}.$$

But, if there exists $\kappa \in [0, g - 2r\log(2))$ such that:

$$\mathbb{P}\left(\limsup_{t\to\infty} \frac{\text{card}\{u \in V_t : X_t^u \geq \exp(\kappa t)\}}{N_t} > 0\right) = 0, \tag{6.7}$$

then $\lim_{t\to+\infty} \text{card}\{u \in V_t : X_t^u \geq \exp(\kappa t)\}/N_t = 0$, in probability. Since $N_t/\mathbb{E}(N_t)$ converges in probability, we have

$$\lim_{t\to+\infty} \frac{\text{card}\{i \in V_t : X_t^i \geq \exp(\kappa t)\}}{\mathbb{E}(N_t)} = 0, \quad \text{in probability.}$$

Moreover,

$$\frac{\text{card}\{u \in V_t : X_t^u \geq \exp(\kappa t)\}}{\mathbb{E}(N_t)} \leq \frac{N_t}{\mathbb{E}(N_t)},$$

which is bounded in L^2. Then card$\{u \in V_t : X_t^u \geq \exp(\kappa t)\}/\mathbb{E}(N_t)$ is uniformly integrable and the convergence in probability of (6.4.2) implies the L^1 convergence. Thus,

$$\lim_{t\to+\infty} \mathbb{P}(Y_t \geq \exp(\kappa t)) = 0,$$

which is in contradiction with Lemma 6.4.2. It ends the proof.

6.5 Fluctuation of the largest individual via deterministic method

Kolmogorov *et al* proved that the extremal particle in a standard branching Brownian motion sits near $\sqrt{2}t$ at time t. Higher order corrections to this result were given by [Bra78]. Theirs proof are based on the study of the Fisher-Kolmogorov-Petrovski-Piscounov equation:

$$\partial_t u = \partial_{xx} u + u^2 - u.$$

More precisely, they give

$$u(t, m(t) + x) \to w(x),$$

where

$$m(t) = \sqrt{2}t - \frac{3}{2\sqrt{2}} \ln(t) + O(1).$$

Recently this equation was generalized in the case of fragmentation process [BHK11]. In the following, we shall give the analogous of the F-KPP equation in our setting. Let u_M be the cumulative function of the law of $\overline{X}$; that is defined by

$$u_M(t, x) = \mathbb{P}\left[\max_{u\in V_t} X_t^u \geq 1 \mid X_0^\emptyset = x\right], \tag{6.8}$$

or equivalently,

$$u_M(t, x) = \mathbb{E}\left[\prod_{u\in V_t} \mathbf{1}_{(-\infty,1]}(X_t^u) \mid X_0^\emptyset = x\right].$$

The study of u is then connected with the non-linear behavior of the measure $\prod_{u\in V_t} \delta_{X_t^u}$ which is different to the linear problems associated to the measure $\sum_{u\in V_t} \delta_{X_t^u}$ as in [BDMT11, Clo11].

Lemma 6.5.1. *Let $f \in C_c^2(\mathbb{R}_+^*)$, if u_f is defined by*

$$u_f(t, x) = \mathbb{E}\left[\prod_{u\in V_t} f(X_t^u) \mid X_0^\emptyset = x\right],$$

for every $t \geq 0$ and $x \geq 0$, then u_f is solution to

$$\begin{cases} \partial_t u_f(t,x) = \sigma^2 x^2 \partial_{xx} u_f(t,x) + (\mu+\sigma^2) x \partial_x u_f(t,x) + r\left(u_f^2\left(t, \frac{x}{2}\right) - u_f(t,x)\right) \\ u_f(0,x) = f(x) \end{cases}$$

Proof. The proof is usual, we may split the expectation into two pieces, according to weither the original particle splits at some $T \leq t$ or not, and obtain

$$u_f(t,x) = P_t f(x) e^{-rt} + \int_0^t P_{t-s}\left(u_f^2\left(s, \frac{\cdot}{2}\right)\right)(x)\, r e^{-r(t-s)} ds,$$

where P_t denote the semigroup associated to the vector field $y' = gy$. Now, a differentiation produces the system. □

Remark 6.5.2 (More general setting)**.** *In the general model of [BDMT11, Clo11], this equation is given by*

$$\partial_t u_f(t,x) = G u_f(t,x) + r \sum_k p_k \int_0^1 \left(\prod_{i=1}^k u_f\left(t, F_i^{(k)}(x,\theta)\right) - u_f(t,x)\right) d\theta,$$

with the notation of chapter 5. In particular, if A is the generator of the Brownian motion, the division is dyadic and local then we recover the classical F-KPP equation.

Finally, note that Lemma 6.2.3 gives

Lemma 6.5.3 (Rescaling properties)**.** *For all $x \geq 0, M \geq 0$ and $t \geq 0$, we have*

$$u_M(t,x) = u_1\left(t, \frac{x}{M}\right)$$

It will be interesting to generalize the result of [BHK11, Bra78, Rob11] for this equation.

Bibliography

[ABBS11] E. Aïdékon, J. Berestycki, É. Brunet, and Z. Shi. The branching Brownian motion seen from its tip. *ArXiv e-prints*, April 2011. 1, 8, 119

[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*, volume 10 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2000. With a preface by Dominique Bakry and Michel Ledoux. 15, 16

[Ada75] Robert A. Adams. *Sobolev spaces*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, 1975. Pure and Applied Mathematics, Vol. 65. 152

[AN04] K. B. Athreya and P. E. Ney. *Branching processes*. Dover Publications Inc., Mineola, NY, 2004. Reprint of the 1972 original [Springer, New York; MR0373040]. 6, 114, 129, 162

[BCG08] Dominique Bakry, Patrick Cattiaux, and Arnaud Guillin. Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.*, 254(3):727–759, 2008. 2

[BCG+11] J.-B. Bardet, A. Christen, A. Guillin, F. Malrieu, and P.-A. Zitt. Total variation estimates for the TCP process. *ArXiv e-prints*, December 2011. 35, 51, 54, 58

[BCnG12] Daniel Balagué, José Alfredo Cañizo, and Pierre Gabriel. Fine asymptotics of profiles and relaxation to equilibrium for growth-fragmentation equations with variable drift rates. Marsh 2012. 7, 114, 122

[BDMT11] Vincent Bansaye, Jean-François Delmas, Laurence Marsalle, and Viet Chi Tran. Limit theorems for Markov processes indexed by continuous time Galton-Watson trees. *Ann. Appl. Probab.*, 21(6):2263–2314, 2011. 7, 35, 57, 110, 113, 116, 119, 120, 121, 123, 129, 130, 134, 148, 159, 170, 173, 174

[BÉ85] D. Bakry and Michel Émery. Diffusions hypercontractives. In *Séminaire de probabilités, XIX, 1983/84*, volume 1123 of *Lecture Notes in Math.*, pages 177–206. Springer, Berlin, 1985. 2, 58

[Bec07] J. Bect. *Processus de Markov diffusifs par morceaux : outils analytiques et numériques*. PhD thesis, Universités Paris-sud XI, 2007. 35, 36

[Ber96] Jean Bertoin. *Lévy processes*, volume 121 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1996. 160, 163

[Ber04] Jean Bertoin. On small masses in self-similar fragmentations. *Stochastic Process. Appl.*, 109(1):13–22, 2004. 1, 160

[Ber06] Jean Bertoin. *Random fragmentation and coagulation processes*, volume 102 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006. 1, 8, 113, 114, 119, 157, 160, 166

[BGM10] Jean-Baptiste Bardet, Hélène Guérin, and Florent Malrieu. Long time behavior of diffusions with Markov switching. *ALEA Lat. Am. J. Probab. Math. Stat.*, 7:151–170, 2010. 69, 79, 102, 103

[BH12] Y. Bakhtin and T. Hurth. Invariant densities for dynamical systems with random switching. *Nonlinearity*, 25:2937–2952, October 2012. 69, 76, 101

[BHK11] J. Berestycki, S. C. Harris, and A. E. Kyprianou. Traveling waves and homogeneous fragmentation. *Ann. Appl. Probab.*, 21(5):1749–1794, 2011. 160, 173, 174

[Bil99] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication. 124, 135

[BKKP05] Onno Boxma, Haya Kaspi, Offer Kella, and David Perry. On/off storage systems with state-dependent input, output, and switching rates. *Probab. Engrg. Inform. Sci.*, 19(1):1–14, 2005. 69

[BLMZ12a] M. Benaïm, S. Le Borgne, F. Malrieu, and P.-A. Zitt. On the stability of planar randomly switched systems. *ArXiv e-prints*, April 2012. 4, 103, 104

[BLMZ12b] M. Benaïm, S. Le Borgne, F. Malrieu, and P.-A. Zitt. Qualitative properties of certain piecewise deterministic Markov processes. *ArXiv e-prints*, April 2012. 69, 76, 101

[BLMZ12c] M. Benaïm, S. Le Borgne, F. Malrieu, and P.-A. Zitt. Quantitative ergodicity for some switched dynamical systems. *ArXiv e-prints*, April 2012. 5, 28, 30, 31, 69, 70, 73, 74, 76, 79, 86

[Bra78] Maury D. Bramson. Maximal displacement of branching Brownian motion. *Comm. Pure Appl. Math.*, 31(5):531–581, 1978. 160, 173, 174

[BS83] F. A. Berezin and M. A. Shubin. *Uravnenie Shredingera [The Schrödinger equation]*. Moskov. Gos. Univ., Moscow, 1983. 48, 112

[BT11] Vincent Bansaye and Viet Chi Tran. Branching Feller diffusion for cell division with parasite infection. *ALEA Lat. Am. J. Probab. Math. Stat.*, 8:95–127, 2011. 8, 35, 55, 61, 119, 121, 129, 130, 160, 161

[But12] O. Butkovsky. Subgeometric rates of convergence of Markov processes in the Wasserstein metric. *ArXiv e-prints*, November 2012. 34

[BY05] Jean Bertoin and Marc Yor. Exponential functionals of Lévy processes. *Probab. Surv.*, 2:191–212 (electronic), 2005. 61

[CD08] O. L. V. Costa and F. Dufour. Stability and ergodicity of piecewise deterministic Markov processes. *SIAM J. Control Optim.*, 47(2):1053–1077, 2008. 35, 36, 69

[CDMR11] A. Crudu, A. Debussche, A. Muller, and O. Radulescu. Convergence of stochastic gene networks to hybrid piecewise deterministic processes. *ArXiv e-prints*, January 2011. 69

[CDPP09] Pietro Caputo, Paolo Dai Pra, and Gustavo Posta. Convex entropy decay via the Bochner-Bakry-Emery approach. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(3):734–753, 2009. 2

[CGZ10] P. Cattiaux, A. Guillin, and P.-A. Zitt. Poincar\'e inequalities and hitting times. *ArXiv e-prints*, December 2010. 67

[CH] Bertrand Cloez and Martin Hairer. Exponential ergodicity for markov processes with random switching. 9

[Cha04] Djalil Chafaï. Entropies, convexity, and functional inequalities: on Φ-entropies and Φ-Sobolev inequalities. *J. Math. Kyoto Univ.*, 44(2):325–363, 2004. 15

[Che96] Mufa Chen. Estimation of spectral gap for Markov chains. *Acta Math. Sinica (N.S.)*, 12(4):337–360, 1996. A Chinese summary appears in Acta Math. Sinica **4**0 (1997), no. 1, 159. 63

[Che04] Mu-Fa Chen. *From Markov chains to non-equilibrium particle systems*. World Scientific Publishing Co. Inc., River Edge, NJ, second edition, 2004. 19, 23, 63, 64, 66

[CJ10] D. Chafai and A. Joulin. Intertwining and commutation relations for birth-death processes. *ArXiv e-prints*, November 2010. 25, 37, 47

[Clo11] B. Cloez. Limit theorems for some branching measure-valued processes. *ArXiv e-prints*, June 2011. 9, 35, 49, 57, 159, 161, 166, 170, 173, 174

[Clo12] B. Cloez. Wasserstein decay of one dimensional jump-diffusions. *ArXiv e-prints*, February 2012. 8, 72, 99

[CMMS12] P. Collet, S. Martinez, S. Meleard, and J. San Martin. Stochastic models for a chemostat and long time behavior. *ArXiv e-prints*, June 2012. 69

[CMP10] Djalil Chafaï, Florent Malrieu, and Katy Paroux. On the long time behavior of the TCP window size process. *Stochastic Process. Appl.*, 120(8):1518–1534, 2010. 35, 44, 58, 59, 124, 142, 146

[CPY01] Philippe Carmona, Frédérique Petit, and Marc Yor. Exponential functionals of Lévy processes. In *Lévy processes*, pages 41–55. Birkhäuser Boston, Boston, MA, 2001. 61

[Dav93] M. H. A. Davis. *Markov models and optimization*, volume 49 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1993. 35, 36, 59

[DF99] Persi Diaconis and David Freedman. Iterated random functions. *SIAM Rev.*, 41(1):45–76, 1999. 69

[DFG09] Randal Douc, Gersende Fort, and Arnaud Guillin. Subgeometric rates of convergence of f-ergodic strong Markov processes. *Stochastic Process. Appl.*, 119(3):897–923, 2009. 33

[DG09] M. Doumic-Jauffret and P. Gabriel. Eigenelements of a General Aggregation-Fragmentation Model. *ArXiv e-prints*, July 2009. 7, 114, 122

[DHKR12] M. Doumic, M. Hoffmann, N. Krell, and L. Robert. Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *ArXiv e-prints*, October 2012. 160, 161

[DHRR11] M. Doumic Jauffret, M. Hoffmann, P. Reynaud-Bouret, and V. Rivoirard. Nonparametric estimation of the division rate of a size-structured population. *ArXiv e-prints*, March 2011. 146

[DM80] Claude Dellacherie and Paul-André Meyer. *Probabilités et potentiel. Chapitres V à VIII*, volume 1385 of *Actualités Scientifiques et Industrielles [Current Scientific and Industrial Topics]*. Hermann, Paris, revised edition, 1980. Théorie des martingales. [Martingale theory]. 145, 167

[DM10] Jean-François Delmas and Laurence Marsalle. Detection of cellular aging in a Galton-Watson process. *Stochastic Process. Appl.*, 120(12):2495–2519, 2010. 119

[DMZ09] M. Doumic Jauffret, P. Maia, and J. P. Zubelli. On the Calibration of a Size-Structured Population Model from Experimental Data. *ArXiv e-prints*, September 2009. 146

[Doo57] J. L. Doob. Conditional Brownian motion and the boundary limits of harmonic functions. *Bull. Soc. Math. France*, 85:431–458, 1957. 109

[DPZ09] Marie Doumic, Benoît Perthame, and Jorge P. Zubelli. Numerical solution of an inverse problem in size-structured population dynamics. *Inverse Problems*, 25(4):045008, 25, 2009. 160

[dSY05] Benoîte de Saporta and Jian-Feng Yao. Tail of a linear diffusion with Markov switching. *Ann. Appl. Probab.*, 15(1B):992–1018, 2005. 69

[Ebe11] Andreas Eberle. Reflection coupling and Wasserstein contractivity without convexity. *C. R. Math. Acad. Sci. Paris*, 349(19-20):1101–1104, 2011. 26, 99

[EHK10] János Engländer, Simon C. Harris, and Andreas E. Kyprianou. Strong law of large numbers for branching diffusions. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(1):279–298, 2010. 1, 110, 113, 119, 155, 157

[EM12] Matthias Erbar and Jan Maas. Ricci Curvature of Finite Markov Chains via Convexity of the Entropy. *Arch. Ration. Mech. Anal.*, 206(3):997–1038, 2012. 22

[Eng09] János Engländer. Law of large numbers for superdiffusions: the non-ergodic case. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(1):1–6, 2009. 110

[EW06] János Engländer and Anita Winter. Law of large numbers for a class of superdiffusions. *Ann. Inst. H. Poincaré Probab. Statist.*, 42(2):171–185, 2006. 110, 119, 121, 122, 129

[FGM10] J. Fontbona, H. Guérin, and F. Malrieu. Quantitative estimates for the long time behavior of an ergodic variant of the telegraph process. *ArXiv e-prints*, June 2010. 69

[FM04] Nicolas Fournier and Sylvie Méléard. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Ann. Appl. Probab.*, 14(4):1880–1919, 2004. 119, 124, 137, 142

[GG96] Charles M. Goldie and Rudolf Grübel. Perpetuities with thin tails. *Adv. in Appl. Probab.*, 28(2):463–480, 1996. 69

[GIY04] Xavier Guyon, Serge Iovleff, and Jian-Feng Yao. Linear diffusion with stationary switching regime. *ESAIM Probab. Stat.*, 8:25–35, 2004. 69

[GK10] Ilie Grigorescu and Min Kang. Steady state and scaling limit for a traffic congestion model. *ESAIM Probab. Stat.*, 14:271–285, 2010. 35, 145

[GRST12] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Displacement convexity of entropy and related inequalities on graphs. *ArXiv e-prints*, July 2012. 22

[GRZ04] Fabrice Guillemin, Philippe Robert, and Bert Zwart. AIMD algorithms and exponential functionals. *Ann. Appl. Probab.*, 14(1):90–117, 2004. 35, 59, 124, 142

[GT11] A. Genadot and M. Thieullen. Averaging for a Fully-Coupled Piecewise Deterministic Markov Process in Infinite Dimension. *ArXiv e-prints*, September 2011. 69

[Hai11] Martin Hairer. On Malliavin's proof of Hörmander's theorem. *Bull. Sci. Math.*, 135(6-7):650–666, 2011. 100

[Har02] Theodore E. Harris. *The theory of branching processes*. Dover Phoenix Editions. Dover Publications Inc., Mineola, NY, 2002. Corrected reprint of the 1963 original [Springer, Berlin; MR0163361 (29 #664)]. 6, 162

[HHK06] J. W. Harris, S. C. Harris, and A. E. Kyprianou. Further probabilistic analysis of the Fisher-Kolmogorov-Petrovskii-Piscounov equation: one sided travelling-waves. *Ann. Inst. H. Poincaré Probab. Statist.*, 42(1):125–145, 2006. 160

[HM08] Martin Hairer and Jonathan C. Mattingly. Spectral gaps in Wasserstein distances and the 2D stochastic Navier-Stokes equations. *Ann. Probab.*, 36(6):2050–2091, 2008. 87, 88

[HM10] Martin Hairer and Andrew J. Majda. A simple framework to justify linear response theory. *Nonlinearity*, 23(4):909–922, 2010. 87, 88

[HM11] Martin Hairer and Jonathan C. Mattingly. Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, volume 63 of *Progr. Probab.*, pages 109–117. Birkhäuser/Springer Basel AG, Basel, 2011. 2, 83, 84

[HMS11] M. Hairer, J. C. Mattingly, and M. Scheutzow. Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. *Probab. Theory Related Fields*, 149(1-2):223–259, 2011. 5, 30, 31, 34, 87, 88

[HS96] P. D. Hislop and I. M. Sigal. *Introduction to spectral theory*, volume 113 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1996. With applications to Schrödinger operators. 112, 114, 157

[HSV11] M. Hairer, A. Stuart, and S. Vollmer. Spectral Gaps for a Metropolis-Hastings Algorithm in Infinite Dimensions. *ArXiv e-prints*, December 2011. 24, 33, 63, 66, 72

[HW96] S. C. Harris and D. Williams. Large deviations and martingales for a typed branching diffusion. I. *Astérisque*, (236):133–154, 1996. Hommage à P. A. Meyer et J. Neveu. 119, 121

[IW89] Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, second edition, 1989. 61, 126, 162

[JM86] A. Joffe and M. Métivier. Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Adv. in Appl. Probab.*, 18(1):20–65, 1986. 124, 153

[JMW12] Benjamin Jourdain, Sylvie Méléard, and Wojbor A. Woyczynski. Lévy flights in evolutionary ecology. *J. Math. Biol.*, 65(4):677–707, 2012. 139

[Jou07] Aldéric Joulin. Poisson-type deviation inequalities for curved continuous-time Markov chains. *Bernoulli*, 13(3):782–798, 2007. 23, 24, 38, 40, 63, 72

[Jou09] Aldéric Joulin. A new Poisson-type deviation inequality for Markov jump processes with positive Wasserstein curvature. *Bernoulli*, 15(2):532–549, 2009. 23, 24, 72

[JS03] Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003. 61, 126, 155

[Kal02] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002. 24

[Kol11] Vassili N. Kolokoltsov. *Markov processes, semigroups and generators*, volume 38 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 2011. 3, 35, 37, 44, 128

[Kub69] H. E. Kubitschek. Growth during the bacterial cell cycle: Analysis of cell size distribution. *Biophysical Journal*, 1969. 146

[Las04] Günter Last. Ergodicity properties of stress release, repairable system and workload models. *Adv. in Appl. Probab.*, 36(2):471–498, 2004. 35, 36

[LP09] Philippe Laurençot and Benoit Perthame. Exponential decay for the growth-fragmentation/cell-division equation. *Commun. Math. Sci.*, 7(2):503–510, 2009. 7, 35, 58, 119, 121, 124

[LvL08] Andreas H. Löpker and Johan S. H. van Leeuwaarden. Transient moments of the TCP window size process. *J. Appl. Probab.*, 45(1):163–175, 2008. 35, 61, 124, 142, 146, 147

[MÓ98] Sylvie Méléard. Convergence of the fluctuations for interacting diffusions with jumps associated with Boltzmann equations. *Stochastics Stochastics Rep.*, 63(3-4):195–225, 1998. 152, 153

[McK75] H. P. McKean. Application of Brownian motion to the equation of Kolmogorov-Petrovskii-Piskunov. *Comm. Pure Appl. Math.*, 28(3):323–331, 1975. 160

[Mét84] Michel Métivier. Convergence faible et principe d'invariance pour des martingales à valeurs dans des espaces de Sobolev. *Ann. Inst. H. Poincaré Probab. Statist.*, 20(4):329–348, 1984. 152

[Mic06] Philippe Michel. Existence of a solution to the cell division eigenproblem. *Math. Models Methods Appl. Sci.*, 16(7, suppl.):1125–1153, 2006. 122

[MR93] Sylvie Méléard and Sylvie Roelly. Sur les convergences étroite ou vague de processus à valeurs mesures. *C. R. Acad. Sci. Paris Sér. I Math.*, 317(8):785–788, 1993. 139

[MT93] Sean P. Meyn and R. L. Tweedie. Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.*, 25(3):518–548, 1993. 2, 84, 145

[MT06] Florent Malrieu and Denis Talay. Concentration inequalities for Euler schemes. In *Monte Carlo and quasi-Monte Carlo methods 2004*, pages 355–371. Springer, Berlin, 2006. 45, 48

[MT12] Sylvie Méléard and Viet Chi Tran. Slow and fast scales for superprocess limits of age-structured populations. *Stochastic Process. Appl.*, 122(1):250–276, 2012. 124, 139

[MV11] S. Méléard and D. Villemonais. Quasi-stationary distributions and population processes. *ArXiv e-prints*, December 2011. 49

[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. 124, 142, 145, 146

[Oll09] Yann Ollivier. Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.*, 256(3):810–864, 2009. 72

[Oll10] Yann Ollivier. A survey of Ricci curvature for metric spaces and Markov chains. In *Probabilistic approach to geometry*, volume 57 of *Adv. Stud. Pure Math.*, pages 343–381. Math. Soc. Japan, Tokyo, 2010. 60, 63, 66, 72

[OV10] Y. Ollivier and C. Villani. A curved Brunn–Minkowski inequality on the discrete hypercube. *ArXiv e-prints*, November 2010. 22

[Per07] Benoît Perthame. *Transport equations in biology*. Frontiers in Mathematics. Birkhäuser Verlag, Basel, 2007. 7, 114, 119, 121, 124, 144, 151, 160

[Pin95] Ross G. Pinsky. *Positive harmonic functions and diffusion*, volume 45 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995. 49, 112, 114, 122, 129, 156

[PR05] Benoît Perthame and Lenya Ryzhik. Exponential decay for the fragmentation or cell-division equation. *J. Differential Equations*, 210(1):155–177, 2005. 119, 121, 124, 143, 144

[PTW12] Khashayar Pakdaman, Michèle Thieullen, and Gilles Wainrib. Asymptotic expansion and central limit theorem for multiscale piecewise-deterministic Markov processes. *Stochastic Process. Appl.*, 122(6):2292–2318, 2012. 69

[PZ07] Benoît Perthame and Jorge P. Zubelli. On the inverse problem for a size-structured population model. *Inverse Problems*, 23(3):1037–1052, 2007. 7

[Rac91] Svetlozar T. Rachev. *Probability metrics and the stability of stochastic models*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1991. 19, 147

[RC86] Sylvie Roelly-Coppoletta. A criterion of convergence of measure-valued processes: application to measure branching processes. *Stochastics*, 17(1-2):43–65, 1986. 124, 137

[Rio98] Emmanuel Rio. Distances minimales et distances idéales. *C. R. Acad. Sci. Paris Sér. I Math.*, 326(9):1127–1130, 1998. 22

[Rob09] Matthew Roberts. *Spine changes of measure and branching diffusions*. PhD thesis, University of Bath, 2009. 164

[Rob11] M. I. Roberts. A simple path to asymptotics for the frontier of a branching Brownian motion. *ArXiv e-prints*, June 2011. 160, 174

[RS78] Michael Reed and Barry Simon. *Methods of modern mathematical physics. IV. Analysis of operators*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1978. 112, 114, 157

[RT00] G. O. Roberts and R. L. Tweedie. Rates of convergence of stochastically monotone and continuous time Markov models. *J. Appl. Probab.*, 37(2):359–373, 2000. 35, 36

[Shi11] Zhan Shi. Random walks and trees. In *X Symposium on Probability and Stochastic Processes and the First Joint Meeting France-Mexico of Probability*, volume 31 of *ESAIM Proc.*, pages 1–39. EDP Sci., Les Ulis, 2011. 8

[Smi68] Walter L. Smith. Necessary conditions for almost sure extinction of a branching process with random environment. *Ann. Math. Statist*, 39:2136–2140, 1968. 69

[Str03] Daniel W. Stroock. *Markov processes from K. Itô's perspective*, volume 155 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 2003. 37

[Tra06] V.C. Tran. *Modèles particulaires stochastiques pour des problèmes d'évolution adaptative et pour l'approximation de solutions statistiques*. PhD thesis, Universités Paris X - Nanterre, 2006. http://tel.archives-ouvertes.fr/tel-00125100. 119, 124, 152

[Vey12] L. Veysseire. Coarse Ricci curvature for continuous-time Markov processes. *ArXiv e-prints*, February 2012. 66

[Vil09] Cédric Villani. *Optimal transport, old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. 19, 20, 38, 70, 147

[vLLO09] J. S. H. van Leeuwaarden, A. H. Löpker, and T. J. Ott. TCP and iso-stationary transformations. *Queueing Syst.*, 63(1-4):459–475, 2009. 35, 59

[vRS05] Max-K. von Renesse and Karl-Theodor Sturm. Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.*, 58(7):923–940, 2005. 22, 63

[Wan03] Feng-Yu Wang. Functional inequalities for the decay of sub-Markov semigroups. *Potential Anal.*, 18(1):1–23, 2003. 66

[Wu10] Liming Wu. Transportation inequalities for stochastic differential equations of pure jumps. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(2):465–479, 2010. 36, 38

[Zei12] Ofer Zeitouni. Branching random walks and gaussian fields. Notes for Lectures, 2012. 8