



**HAL**  
open science

# Stochastic branching processes with selection

Brieuc Frénais

► **To cite this version:**

Brieuc Frénais. Stochastic branching processes with selection. Mathematics [math]. Université de Strasbourg, 2024. English. NNT : 2024STRAD033 . tel-04834834v3

**HAL Id: tel-04834834**

**<https://theses.hal.science/tel-04834834v3>**

Submitted on 30 Jan 2025

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

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



Distributed under a Creative Commons Attribution 4.0 International License

École doctorale Mathématiques, sciences de l'information et de l'ingénieur  
Institut de Recherche Mathématique Avancée

**THÈSE** présentée par :  
**Briec FRÉNAIS**

soutenue le /defended on : 20 décembre 2024

pour obtenir le grade de / to obtain the grade of :  
Docteur de l'Université de Strasbourg / Strasbourg University Doctor

Discipline/S spécialité / Discipline/specialty : Mathématiques

**Modèles stochastiques de  
branchement-sélection**

THÈSE dirigée par / DISSERTATION supervisor :

**BÉRARD Jean**

Professeur, Université de Strasbourg

RAPPORTEURS :

**BERESTYCKI Julien**

Associate Professor, University of Oxford

**VÉBER Amandine**

Directrice de recherche, Université Paris Cité

AUTRES MEMBRES DU JURY / OTHER MEMBERS OF THE JURY :

**LENÔTRE Lionel**

Maître de conférences, Université de Haute-Alsace

**HERRMANN Samuel**

Professeur, Université de Bourgogne

**LEMAN Hélène**

Chargée de recherche, ENS Lyon

**VILLEMONAIS Denis**

Professeur, Université de Strasbourg

## Avertissement au lecteur / Warning to the reader

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition des membres de la communauté universitaire. Il est soumis à la propriété intellectuelle de l'auteur. Cela implique une obligation de citation et de référencement lors de l'utilisation de ce document. D'autre part, toute contrefaçon, plagiat, reproduction ou représentation illicite encourt une poursuite pénale.

This document is the result of a long process approved by the jury and made available to members of the university community. It is subject to the intellectual property rights of its author. This implies an obligation to quote and reference when using this document. Furthermore, any infringement, plagiarism, unlawful reproduction or representation will be prosecuted.

### [Code de la Propriété Intellectuelle](#)

[Article L122-4](#) : Toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite. Il en est de même pour la traduction, l'adaptation ou la transformation, l'arrangement ou la reproduction par un art ou un procédé quelconque.

Any representation or reproduction in whole or in part without the consent of the author or his successors in title or assigns is unlawful. The same applies to translation, adaptation or transformation, arrangement or reproduction by any art or process whatsoever.

[Articles L335-1 à L335-9](#) : Dispositions pénales / Penal provisions.

## Licence attribuée par l'auteur / Licence attributed by the author

[*Veillez à ne conserver que la licence qui vous convient / Make sure you only keep the licence that suits you best*]



<https://creativecommons.org/licenses/?lang=fr-FR>

# Thèse

INSTITUT DE  
RECHERCHE  
MATHÉMATIQUE  
AVANCÉE

UMR 7501

Strasbourg

présentée pour obtenir le grade de docteur de  
l'Université de Strasbourg  
Spécialité MATHÉMATIQUES

**Brieuc Frénais**

**Modèles stochastiques de branchement-sélection**

Soutenue le 20 décembre 2024  
devant la commission d'examen

Samuel Herrmann, président du jury  
Jean Bérard, directeur de thèse  
Lionel Lenôtre, co-encadrant  
Julien Berestycki, rapporteur  
Amandine Véber, rapportrice  
Hélène Leman, examinatrice  
Denis Villemonais, examinateur

<https://irma.math.unistra.fr>



**Université**

de Strasbourg

# Modèles stochastiques de branchement-sélection

Brieuc Frénais



## Remerciements

Comment commencer ? C'est une question que je me suis posée à maintes reprises pendant ces quelques années, et probablement celle qui m'aura donné le plus de difficultés. Aujourd'hui encore, après avoir relu tant de fois ce manuscrit, je me la pose une dernière fois – avant la prochaine.

Mes premiers mots iront naturellement à ceux qui ont commencé ce voyage avec moi, il y a un peu plus de trois ans. Merci Jean, pour les innombrables heures passées à m'écouter, à débattre et à m'impressionner avec tes idées et ta compréhension de problèmes sur lesquels j'avais tant travaillé, à relire ces nombreuses pages avec tant de rigueur, et à me recentrer quand je m'égarais. Impossible de se lancer dans une aventure pareille sans un guide, et j'ai eu la chance de tomber sur le meilleur d'entre eux quand les aléas de la vie m'ont mené à Strasbourg. Merci Lionel pour m'avoir poussé à sortir de ma zone de confort en allant au CEMRACS, où j'ai appris tant de choses et élargi ma vision des mathématiques. Merci à tous les deux pour m'avoir fait découvrir le monde de la recherche, ce pour quoi je vous serai toujours reconnaissant.

Je remercie chaleureusement Julien et Amandine, qui ont accepté la lourde tâche de relire ce manuscrit, et dont les remarques me sont très précieuses. Merci notamment à Amandine pour les commentaires très précis mais aussi les conseils généraux que je ne manquerai pas de garder à l'esprit dans les années à venir. Merci également à Hélène, Samuel et Denis pour avoir accepté de faire partie de ce jury, et pour les discussions enrichissantes au détour d'un couloir ou entre deux exposés.

Un grand merci à Mathieu, qui a imaginé ce très beau projet qui nous a bien occupés à l'été 2022 – et dans les mois qui ont suivi. Merci à Sebastian et Amal pour les séances de programmation dans la fraîcheur du CIRM, ainsi qu'à Yuchen pour avoir eu le courage de rejoindre un projet très avancé et lui avoir donné encore plus de profondeur. Ce travail conclut un très bel été, avec de belles mathématiques mais aussi de belles rencontres : merci à Claire et Léo pour les débriefs, et aux collègues strasbourgeois Mickaël, Romane, Guillaume, Laurent et Emmanuel pour les parties de tarot, les randonnées, et les baignades dans les calanques.

Plus généralement, je souhaite remercier tous les membres de l'IRMA, qui font du labo un environnement de travail très agréable et vivant au quotidien : en particulier, merci à Jean, Jacques, Pierre-Olivier, Laure et Xiaolin pour les discussions animées après les séminaires, à Davide et Sylvain pour les nombreux échanges autour des cours de probabilités, et à Clémentine, Amaury, Pierre B., Raphaël, Michaël, Alexandru, Olivier, Charles, Guillaume K., Giuseppe et Florence pour les discussions au salon ou au détour d'un couloir.

J'ai eu la chance de découvrir un environnement complètement différent depuis le mois de septembre, et je suis en particulier très reconnaissant envers Anouk, sans laquelle ce début d'année aurait été très différent : merci pour les nombreux conseils et discussions sur les différents cours que l'on a en commun. Yannick, je suis très heureux de savoir que nos longues discussions pourront continuer, j'espère simplement qu'elles pourront avoir lieu un peu plus tôt dans la journée ! Un grand merci à Wahiba, Parisa, Dominik, Rémi, Sandie et Frédéric pour m'avoir si bien accueilli à l'Ecole des Mines, et à Pierrick, Nathan, Mabrouk, Yann et les autres pour m'avoir invités à participer à la vie doctorante ici (promis, je viendrai plus souvent au second semestre).

Je ne serais pas là sans tous ceux qui m'ont fait découvrir les mathématiques sous toutes leurs formes : je remercie tout particulièrement Pascale, Serge et François pour ces belles années à Clemenceau, et NicJui et son enthousiasme sans bornes, sans qui le M2 et la thèse n'auraient pas eu la même saveur.

J'en viens à ceux que j'ai côtoyés au quotidien pendant ces trois années. Les doctorants (et post-doctorants) de l'IRMA forment une magnifique communauté, et c'est un plaisir immense d'avoir partagé ces trois années avec vous. Désolé de ne pas être venu aussi souvent que prévu depuis le mois de septembre, je vous assure que ce n'est pas l'envie qui a manqué ! Entre ceux qui étaient là à mon arrivée et ceux qui sont arrivés depuis mon départ, vous êtes très nombreux, et j'espère que vous me pardonneriez de ne pas tous vous citer ici; je serai très heureux de vous remercier de vive voix quand je vous verrai (il se peut que je sois en ce moment même en train de filer vers vous, en croisant les doigts pour trouver les mots pour finir ce manuscrit avant que le train n'arrive). Je tiens tout de même à remercier Clarence, qui m'a tout enseigné en matière de présidence. Merci à Victoria, Quentin, Céline et Armand pour avoir organisé deux très belles conférences en montagne (ou plutôt de m'y avoir aidé, si j'en crois la page web ?). Thomas, ç'aura été un plaisir de partager avec toi, en plus du reste, la responsabilité de ce superbe séminaire et cet extraordinaire pot de thèse à venir. Merci Roméo, pour tous ces beaux moments partagés dans nos moult bureaux successifs et pour avoir su assumer avec un grand talent le rôle que nous t'avions confié. Enfin, merci Antoine et merci Victoria pour m'avoir (presque immédiatement) accueilli dans votre si belle région, merci pour votre gentillesse, votre écoute dans les moments difficiles, et merci de rayonner et de m'avoir offert tous ces merveilleux moments de bonheur.

Je dois dire que c'est très difficile de se figurer à l'avance ce que c'est que d'écrire ces pages; difficile d'anticiper le vertige que cet exercice appelle. C'est avec une immense gratitude que je repense à toutes les personnes qui m'ont accompagné depuis si longtemps. Arthur, Léo, "si longtemps" ne vous rend pas justice, car vous avez toujours été là. Impossible d'imaginer une vie sans vous. Nolwenn, désolé mais je me dois de te rappeler que les poudous n'existent pas; heureusement, toi, tu existes bel et bien, et je suis si reconnaissant de ton amitié sans faille depuis des années. Lisa, Emilien, merci pour ces belles soirées, ces magnifiques vacances, ces moments passés à discuter de tout sauf de maths. Morgane, merci pour ta gentillesse et ta folie, et pour être une si belle amie pour Clarisse et moi. Et merci aussi pour nous avoir fait rencontrer tes amis de longue date à notre arrivée à Strasbourg : Yoann, Clémentine, nos années strasbourgeoises n'auraient pas été les mêmes sans vous. Je crois que vous l'avez compris, je suis extrêmement heureux de voir votre petite famille se construire, et tellement reconnaissant de faire partie de votre vie.

Reste donc ceux qui sont là depuis le début : un grand merci à mes cousines et mes cousins, à mes oncles et tantes, à Aline et à Granny, pour avoir construit cette si belle famille dans laquelle j'ai pu grandir et évoluer. J'ai une pensée toute particulière pour Yannick, Grand-Père et Grand-Père Louis, avec qui j'aurais tant aimé partager ces moments. Je remercie également mes beaux-parents, avec qui nous passons toujours de merveilleux moments. Papa, Maman, je n'arrive pas à trouver les mots pour exprimer tout ce que je ressens; je crois que je vais abdiquer pour le moment et que cela se fera tout seul la semaine prochaine. Ju, merci pour tous les beaux moments passés ensemble. C'est très précieux d'avoir une petite sœur et de savoir qu'on peut compter sur elle, et qu'on pourra toujours le faire.

J'ai dit qu'il fallait un guide pour réussir à aller au bout de cette aventure, et j'espère que je vous ai fait comprendre à quel point je suis redevable à chacun d'entre vous. Ce que je n'ai pas encore dit, c'est qu'il faut trouver la force, celle de toujours se relancer quand on a l'impression de ne plus avancer. Cette force, Clarisse, c'est toi qui me l'as donnée. Jamais je ne pourrai te rendre ce que tu m'as offert en étant à mes côtés pendant ces trois années, et j'espère que tu as conscience de tout ce que tu as apporté à ma vie. Merci pour ta douceur, merci pour ta gentillesse, merci pour ton sourire, que tu m'offres au quotidien. Je n'aurais pas pu réussir sans toi. Merci.



## Résumé

L'objet central de cette thèse est un système de particules appelé  $N$ -processus de Markov branchant qui généralise le  $N$ -mouvement brownien branchant étudié par Maillard [Mai16] : on part de  $N$  particules aux positions initiales  $x^1, \dots, x^N$ ; chaque particule se déplace alors sur la droite réelle suivant la loi d'un processus markovien homogène en temps donné (on parle de processus sous-jacent), indépendamment des autres trajectoires. De plus, chaque particule branche à taux 1 : c'est-à-dire qu'après une durée aléatoire de loi exponentielle  $\mathcal{E}(1)$ , celle-ci se duplique pour donner naissance à deux nouvelles particules à la même position. Enfin, simultanément à l'apparition d'une nouvelle particule dans le système à chaque instant de branchement, on retire du système la particule dont la position est la plus basse. Cette étape de sélection permet de conserver à tout instant une taille de population égale à  $N$ .

Nos principaux résultats concernent l'existence d'une limite hydrodynamique pour ce système de particules, c'est-à-dire l'existence d'une limite pour la fonction de répartition empirique associée à ce processus, quand la taille de population  $N$  tend vers l'infini. Sous certaines hypothèses de régularité, nous démontrons que cette fonction de répartition empirique converge vers la fonction de répartition de la loi du processus sous-jacent conditionné à ne pas avoir franchi une certaine frontière, cette dernière étant caractérisée comme solution du problème inverse du premier temps de passage avec loi exponentielle  $\mathcal{E}(1)$  pour le processus sous-jacent. Notre stratégie de preuve utilise un couplage avec des processus auxiliaires dont la construction fait appel à une hypothèse de monotonie stochastique sur le processus sous-jacent. En lien avec cette preuve, nous établissons dans un contexte général l'existence d'un couplage préservant l'ordre, entre plusieurs trajectoires d'un même processus de Markov satisfaisant l'hypothèse de monotonie stochastique et la propriété de Feller.

Enfin, nous nous intéressons par ailleurs à un second problème, celui du transport optimal en champ moyen, avec un point de vue entièrement numérique : nous développons trois méthodes d'apprentissage profond fournissant des solutions approchées de ce problème et les mettons en œuvre dans différents cas tests.



# Table des matières

<b>Table des matières</b>	<b>vii</b>
<b>Introduction</b>	<b>1</b>
1 Limite hydrodynamique du $N$ -BMP	3
1.1 Contexte	3
1.2 Résultats principaux	6
1.3 Stratégie de preuve	9
1.3.1 L'approche de De Masi, Ferrari, Presutti et Soprano-Loto	10
1.3.2 Notre approche	12
1.4 Discussion	13
1.4.1 Limite hydrodynamique et é.d.p. pour les modèles de branchement-sélection	13
1.4.2 Problème inverse du premier temps de passage	15
2 Flot comonotone pour des processus de Feller sur la droite réelle	21
2.1 Hypothèses	22
2.2 Résultats	22
2.3 Discussion	24
3 Problème de transport optimal en champ moyen	24
3.1 Contrôle et transport optimal en champ moyen	24
3.2 Description des méthodes numériques	26
3.2.1 Méthode directe	26
3.2.2 Résolution d'une é.d.p. couplée	27
3.2.3 Méthode de lagrangien augmenté pour une formulation duale	27
4 Questions en suspens et perspectives	28
4.1 Limite hydrodynamique de systèmes de particules	28
4.2 Flot comonotone d'un processus markovien sur la droite réelle	33
4.3 Transport optimal en champ moyen	34
<b>I Hydrodynamic limit of the <math>N</math>-BMP</b>	<b>35</b>
I.1 Introduction	35
I.1.1 Statement of the main results	35
I.1.2 Discussion	37
I.1.2.1 Assumptions	37
I.1.2.2 Related works	38
I.1.3 Outline of the proof	39
I.1.4 Organization of the paper	39
I.2 Study of the $b$ -BMP	40
I.2.1 Hydrodynamic limit of the $b$ -BMP	40
I.2.2 Application to $\mathcal{X}^+$ and $\mathcal{X}^-$	45

I.3	Hydrodynamic limit of the $N$ -BMP	49
I.4	Convergence of the minimum of the $N$ -BMP	52
<b>II The comonotone flow of a stochastically monotone Feller process on the real line</b>		<b>61</b>
II.1	Introduction	61
II.1.1	Markov kernels	61
II.1.2	Markov and Feller semigroups	61
II.1.3	Consistent families	62
II.1.4	Stochastic orders	63
II.1.5	Statement of the main results	63
II.1.6	Discussion	64
II.1.7	Organization of the paper	65
II.2	Feller-related results	65
II.2.1	General properties	66
II.2.2	Extension of $(P_t)$ to $\mathbb{R}$	68
II.3	Proof of Theorem II.1	70
II.3.1	Discrete-time flow	70
II.3.2	Construction of the limit $R_t^{(n)}$	71
II.4	Proof of Theorem II.2	75
<b>III Deep Learning methods for Mean Field Optimal Transport problem</b>		<b>77</b>
III.1	Introduction	77
III.2	Definition of the problem	79
III.3	Numerical methods	80
III.3.1	Direct approach for the optimal control formulation	80
III.3.1.1	Problem Approximation	81
III.3.1.2	Description of the algorithm	82
III.3.2	Deep Galerkin Method for the PDE system	84
III.3.2.1	PDE system for MFOT	84
III.3.2.2	Description of the algorithm	86
III.3.3	Augmented Lagrangian Method with Deep Learning	87
III.3.3.1	Primal and dual problems	87
III.3.3.2	Description of the algorithm	90
III.4	Numerical experiments	91
III.4.1	Case 1: Linear Quadratic Problem	92
III.4.1.1	Description of the problem	92
III.4.1.2	Evaluation Metrics	92
III.4.1.3	Numerical results	94
III.4.2	Case 2: Transport with congestion effects	98
III.4.2.1	Description of the problem	98
III.4.2.2	Numerical results	98
III.4.3	Remarks on the choice of hyperparameters	99
III.5	Conclusion and future directions	101
<b>A Construction of a coupling by pasting</b>		<b>103</b>
A.1	Pasting of Markov processes	103
A.1.1	Construction by piecing out	103
A.1.2	Formal definition of some branching processes with selection	107
A.1.3	Pasting of pasted processes	110

A.1.3.1	Definition and notation for the pasting of pasted process	110
A.1.3.2	Commutativity of the construction	112
A.1.4	Distribution of the coordinate processes	117
A.2	Formal definition of the coupling	121
A.2.1	Construction of the coupling	121
A.2.1.1	State space and base process	121
A.2.1.2	Probability space and pasting times	123
A.2.1.3	Renewal distribution	124
A.2.2	Proof of Proposition 1.4	127
<b>B</b>	<b>Some results on the geometric distribution</b>	<b>133</b>
<b>C</b>	<b>Additional results related to Chapter II (comonotone flow)</b>	<b>135</b>
C.1	About Theorems 4 and 5 in <a href="#">KKO77</a>	135
C.1.1	About Theorem 4	135
C.1.2	About Theorem 5	136
C.2	Miscellaneous lemmas	137
C.3	Numerical illustrations	138
<b>D</b>	<b>Some more details on the MFOT problem</b>	<b>141</b>
D.1	Solution for LQ problem	141
D.2	Computation of $\mathcal{G}$	142
D.3	Loss plots	144
	<b>Index of notations for Chapters I and II</b>	<b>151</b>
	<b>Index of notations for Chapter III and Appendix D</b>	<b>153</b>
	<b>Index of notations for Appendices A and B</b>	<b>155</b>
	<b>Références</b>	<b>157</b>



# Introduction

Les modèles stochastiques constituent une approche mathématique largement utilisée pour modéliser et expliquer des phénomènes physiques, biologiques ou financiers. Dans le contexte de la dynamique des populations, l'introduction de l'aléa permet notamment de répondre à deux problèmes : d'une part, capturer des fluctuations inhérentes à la nature des phénomènes modélisés (comme par exemple l'appariage des individus au moment de la reproduction au sein d'une population donnée); et d'autre part, simplifier des modèles déterministes faisant intervenir un (trop) grand nombre de degrés de liberté.

Une classe particulièrement étudiée de modèles, décrivant l'évolution de la distribution de traits génétiques au sein d'une population, est celle introduite par Fleming et Viot [FV79] en 1979; certaines variantes permettent de prendre en compte des interactions spatiales entre individus [BEV10] ou de favoriser certains gènes [EVY20]. Dans cette thèse, nous nous intéressons à des modèles de particules, en temps et en espace continu, correspondant à une description microscopique de la dynamique de la population au cours du temps; la distribution des traits génétiques dans la population se présente ainsi comme une combinaison linéaire de masses de Dirac.

Les modèles que nous étudions se rattachent à une classe de systèmes de particules dont l'étude a été initiée par Brunet et Derrida [BD97], en lien avec la propagation de front d'onde pour les solutions de l'équation de Fisher-Kolmogorov-Petrovsky-Piskunov<sup>1</sup> (F-KPP)

$$\partial_t u = \frac{1}{2} \partial_x^2 u + u(1 - u). \quad (1)$$

Le premier modèle proposé par Brunet et Derrida consiste en une population de  $N$  particules sur l'axe  $\mathbb{Z}$  renouvelée (en temps discret) *via* l'échantillonnage de paires de particules, permettant de maintenir une population de taille constante tout en favorisant les positions les plus hautes. Un second modèle, plus proche de celui étudié dans cette thèse, est la *N-marche aléatoire branchante* [BDMM06, BG10] : à chaque étape, chacune des  $N$  particules est remplacée par un nombre (déterministe ou aléatoire) de particules obtenues en effectuant un pas de marche aléatoire à partir de sa position; on ne garde ensuite que les  $N$  particules les plus hautes pour former la génération suivante.

Une variante en temps et en espace continu de la *N-marche aléatoire branchante* est le *N-mouvement brownien branchant* (abrégé en *N-BBM* dans la suite pour suivre la terminologie anglophone de "Branching Brownian Motion") étudié par Maillard [Mai16]. Dans ce modèle, on part de  $N$  particules aux positions initiales  $x^1, \dots, x^N$ . Chaque particule se déplace alors sur la droite réelle suivant un mouvement brownien, indépendamment des autres trajectoires. De plus, chaque particule branche à taux 1 : c'est-à-dire qu'après une durée aléatoire de loi exponentielle  $\mathcal{E}(1)$ , celle-ci se duplique pour donner naissance à deux nouvelles particules à la

---

<sup>1</sup>Cette équation a été introduite indépendamment par Fisher [Fis37] et Kolmogorov, Petrovsky et Piskunov [KPP37] en 1937. Dans sa version la plus générale, le terme non-linéaire  $u(1 - u)$  peut être remplacé par  $F(u)$ , où  $F$  est une fonction de classe  $C^1$  telle que  $F(0) = F(1) = 0$ ,  $F(h) > 0$  pour  $h \in (0, 1)$ , et le maximum de  $F'$  sur  $[0, 1]$  est strictement positif et atteint uniquement en 0.)

même position (on parle de branchement *dyadique local*). Enfin, simultanément à l'apparition d'une nouvelle particule dans le système à chaque instant de branchement, on retire du système la particule<sup>2</sup> dont la position est la plus basse<sup>3</sup> (dans la suite, on dira aussi que l'on tue cette particule). Cette étape de sélection permet de conserver à tout instant une taille de population égale à  $N$ . En l'absence de mécanisme de sélection dans le modèle précédent, on obtient le *mouvement brownien branchant* (BBM), abondamment étudié [Bra78, LS92, HHK06, ABB513] et connu depuis les travaux de McKean [McK75] pour être lui aussi lié à l'équation de F-KPP (I).

L'objet central de cette thèse est un système de particules appelé  *$N$ -processus de Markov branchant* ( $N$ -BMP<sup>4</sup>), qui généralise le  $N$ -BBM : plutôt que de suivre des trajectoires browniennes standard, les particules suivent des trajectoires d'un processus markovien homogène en temps sur  $\mathbb{R}$ , noté  $(X_t)_{t \geq 0}$  (on parlera dans la suite du *processus sous-jacent*); on applique les mêmes règles de branchement et de sélection que pour le  $N$ -BBM. Ce modèle se base donc sur une description individuelle – ou microscopique – de la dynamique de la population, et utilise des règles de sélection pour réguler la taille de la population. Cependant, pour se faire une idée globale du comportement de la population, on souhaite repasser du point de vue microscopique au point de vue macroscopique, d'où la question de la *limite hydrodynamique* : quel est le comportement de la distribution empirique des particules quand la taille de la population tend vers l'infini ? C'est cette question centrale qui est à l'origine des travaux qui suivent.

Ce manuscrit s'articule comme suit. Dans la suite de ce chapitre, on présente les principaux résultats obtenus dans le cadre de cette thèse et la manière dont ils s'intègrent dans la littérature mathématique. Les chapitres qui suivent, sous forme d'articles (rédigés en anglais), contiennent la version détaillée des travaux effectués. Dans le Chapitre II, nous étudions la limite hydrodynamique du  $N$ -BMP. Nous établissons sous certaines hypothèses la convergence de la fonction de répartition empirique du  $N$ -BMP vers une limite déterministe, qui s'interprète comme la fonction de répartition du processus  $(X_t)_{t \geq 0}$  conditionné à rester au-dessus d'une certaine frontière. Pour démontrer ce résultat, nous construisons un couplage du  $N$ -BMP avec un processus auxiliaire plus simple à analyser. Dans le Chapitre II<sup>5</sup>, nous montrons l'existence de couplage de plusieurs versions du processus  $(X_t)_{t \geq 0}$  préservant l'ordre, sous des hypothèses très générales. Ce résultat intervient de manière cruciale dans la construction du couplage du  $N$ -BMP avec le processus auxiliaire. Le Chapitre III<sup>6</sup> est consacré à l'étude du problème de transport optimal en champ moyen. Ce travail a été réalisé indépendamment des autres durant l'été 2022 et les mois qui ont suivi dans le cadre du CEMRACS, en collaboration avec Sebastian Baudelet, Mathieu Laurière, Amal Machtalay et Yuchen Zhu. Les chapitres annexes contiennent des détails techniques que nous avons choisi de reporter pour plus de clarté : l'Annexe A contient des détails techniques sur la construction du couplage utilisé dans le Chapitre II; l'Annexe B contient un résultat de grandes déviations pour la loi géométrique; l'Annexe C contient des détails techniques et des illustrations numériques autour du flot comonotone; et l'Annexe D apporte des précisions sur les travaux réalisés pendant le CEMRACS 2022.

<sup>2</sup>S'il y a plusieurs particules à la position la plus basse, on choisit uniformément l'indice de celle qui est retirée.

<sup>3</sup>Notre convention est de considérer (et représenter) le temps en abscisse et la position sur l'axe réel en ordonnée; une particule à la position  $x$  est donc plus haute qu'une particule à la position  $y$  si  $x \geq y$ .

<sup>4</sup>Pour " $N$ -Branching Markov Process".

<sup>5</sup>Le Chapitre II et les Annexes A et B reprennent [BF23a].

<sup>6</sup>Le Chapitre III et l'Annexe C reprennent [BF23b].

<sup>7</sup>Le Chapitre III et l'Annexe D reprennent [BFL<sup>+</sup>23].



# 1 Limite hydrodynamique du $N$ -BMP

Les résultats du [Chapitre II](#), résumés dans cette partie, étendent ceux obtenus par De Masi, Ferrari, Presutti et Soprano-Loto [\[DMFPSL19a\]](#) sur la limite hydrodynamique du  $N$ -BBM. Les résultats pré-existants à nos travaux sont décrits dans la [Section 1.1](#). Nos résultats sont présentés dans la [Section 1.2](#), et leur preuve est esquissée dans la [Section 1.3](#). Enfin, nous discutons en détails les hypothèses faites, en lien avec la littérature mathématique, dans la [Section 1.4](#).

## 1.1 Contexte

Une description du  $N$ -BBM équivalente à celle donnée au paragraphe précédent est de demander que les  $N$  particules suivent des trajectoires (browniennes) indépendantes, et qu'à taux  $N-1$ , la particule la plus basse saute à la position de l'une des  $N-1$  autres, choisie uniformément, symbolisant le branchement de l'une et la mort de l'autre. Cela permet notamment de suivre la position des particules au cours du temps, et l'on peut donc noter, pour  $t \geq 0$ ,  $\mathcal{X}_t = (X_t^1, \dots, X_t^N)$  la position des  $N$  particules composant ce processus à l'instant  $t$ . On caractérise la loi empirique

$$\frac{1}{N} \sum_{i=1}^N \delta_{X_t^i}$$

de ce nuage de points par la fonction de répartition empirique définie par

$$\forall r \in \mathbb{R}, \forall t \geq 0, \quad F^N(t, r) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{X_t^i \leq r\}}.$$

Une manière d'aborder la question de la limite hydrodynamique du  $N$ -BBM est donc d'étudier la limite, lorsque  $N$  tend vers l'infini, de cette fonction de répartition empirique. Dans ce cadre, De Masi Ferrari, Presutti et Soprano-Loto ont démontré le résultat suivant.

**Théorème A** ([\[DMFPSL19a\]](#)<sup>8</sup>, Théorème 1). *Il existe une fonction  $\psi : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$  telle que pour  $t \geq 0$ ,  $\psi(t, \cdot)$  est une densité de probabilité sur  $\mathbb{R}$  et pour tout réel  $r$ ,  $F^N(t, r)$  converge presque sûrement et dans  $L^1$  vers  $\int_{-\infty}^r \psi(t, x) dx$  quand  $N$  tend vers l'infini.*

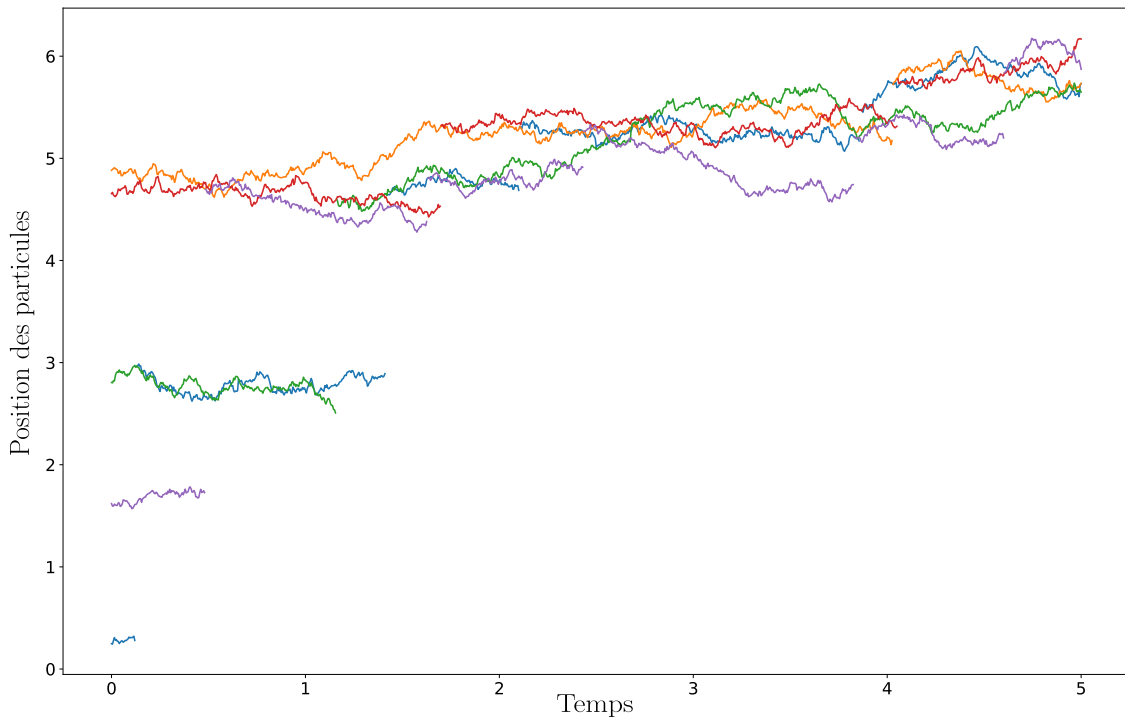
De plus, cette limite a été caractérisée comme solution du *problème à frontière libre* suivant, associé à une équation de réaction-diffusion : trouver une fonction  $u : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$  et une fonction  $b : (0, +\infty) \rightarrow \mathbb{R}$  (appelée *frontière*) continue telles que  $u$  est de classe [\[9\]](#)  $C^{1,2}$  sur  $\mathcal{U}^b := \{(t, x) \in (0, +\infty) \times \mathbb{R}, x > b_t\}$  et vérifie

$$\begin{cases} \partial_t u = \frac{1}{2} \partial_x^2 u + u, & \text{pour } t > 0, x > b_t, \\ u(t, x) = 0, & \text{pour } t > 0, x \leq b_t, \\ \int_{b_t}^{+\infty} u(t, x) dx = 1, & \text{pour } t > 0. \end{cases} \quad (2)$$

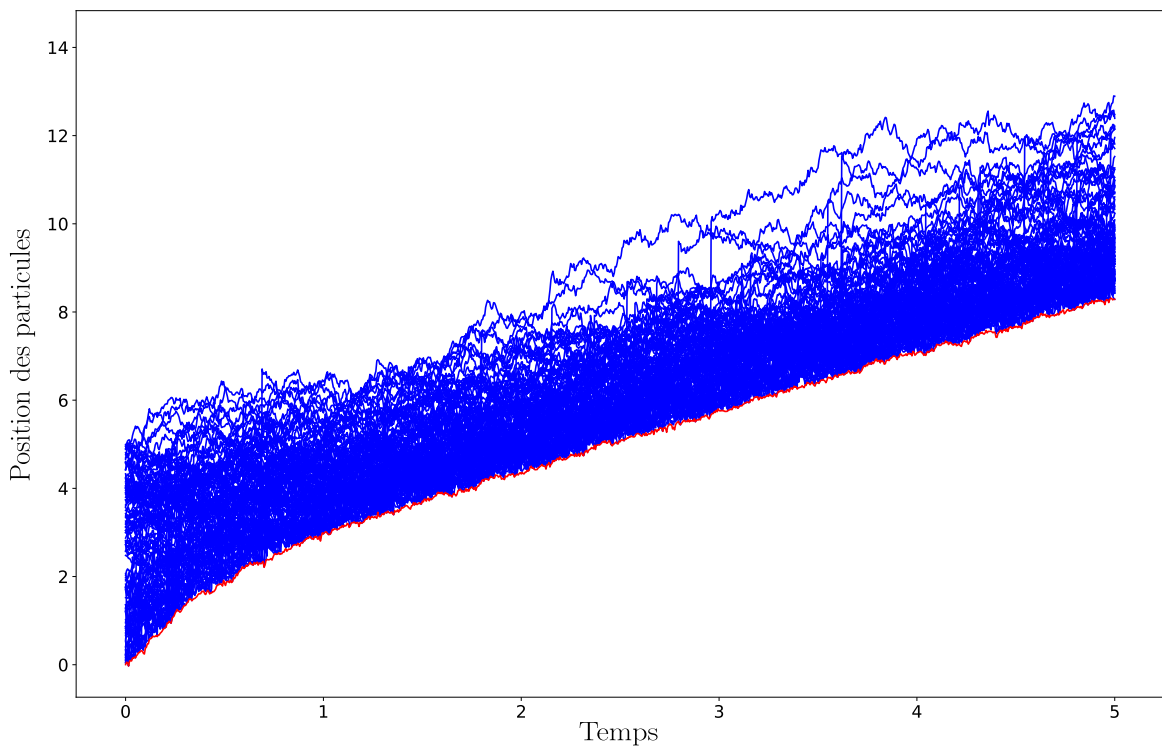
Ce problème a été étudié en deux temps par Lee [\[Lee20\]](#) (existence d'une solution locale en temps) puis Berestycki, Brunet et Penington [\[BBP19\]](#) (existence d'une unique solution globale), ce qui permet d'arriver au résultat suivant.

<sup>8</sup>Voir aussi [\[DMFPSL19b\]](#), Théorème 3.1

<sup>9</sup>Si  $u$  est une fonction de deux variables, on dit que  $u$  est de classe  $C^{1,2}$  si elle est de classe  $C^1$  en la première variable (temps) et de classe  $C^2$  en la seconde (espace).



(a) Une réalisation d'un 5-BBM avec une loi de départ une uniforme sur l'intervalle  $[0, 5]$



(b) Une réalisation d'un 1000-BBM avec une loi de départ une uniforme sur l'intervalle  $[0, 5]$

Figure 1

**Théorème B** ([BBP19], Corollary 2.1 & [DMFPSL19a], Théorème 2). *La limite hydrodynamique du  $N$ -BBM partant de  $N$  particules identiquement et indépendamment distribuées selon une loi initiale à densité  $u_0$  sur  $\mathbb{R}$  est l'unique solution  $u$  du problème à frontière libre (2) avec condition initiale  $u(0, x) = u_0(x)$ , au sens où la fonction  $\psi$  donnée au Théorème A coïncide sur  $\mathbb{R}_+ \times \mathbb{R}$  avec  $u$ .*

Ce problème à frontière libre admet<sup>[10]</sup> l'interprétation probabiliste suivante : supposons que pour une densité initiale  $u_0$  donnée, il existe une frontière  $b : (0, +\infty) \rightarrow \mathbb{R}$  et une fonction  $u : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$  résolvant (2) avec donnée initiale  $u_0$ ; alors  $u(t, \cdot)$  est la densité d'un mouvement brownien issu de la loi initiale à densité  $u_0$  et conditionné à rester au-dessus de la frontière  $b$  avant l'instant  $t$ , avec la condition que le temps d'atteinte de cette frontière est de loi exponentielle  $\mathcal{E}(1)$ . La question de l'existence de telles frontières et l'origine du lien avec la loi exponentielle sont détaillées dans les Sections 1.3.2 et 1.4.2.

Dans [DMFPSL19a, DMFPSL19b, BBNP22], les auteurs étudient le  $N$ -BBM et deux généralisations de celui-ci :

- le  $N$ -BBM avec *branchement non-local* : la seule différence avec le  $N$ -BBM (local) est que la position de la particule qui apparaît dans le système à chaque instant de branchement n'est pas égale à celle de la particule branchante, mais présente un décalage aléatoire par rapport à celle-ci,
- les *abeilles browniennes* : on considère cette fois un BBM en dimension  $d \geq 2$ , où l'on tue à chaque instant de branchement la particule la plus éloignée de l'origine de  $\mathbb{R}^d$ .

Les auteurs démontrent dans chaque cas l'existence d'une limite hydrodynamique au système de particules considéré, et la relient dans un second temps au problème à frontière libre correspondant, pour lequel l'existence d'une solution est étudiée indépendamment [Lee20, BBD18, BBP19, Lee18, BBNP21]. Notons que les preuves d'existence de limite hydrodynamique ne semblent pas permettre d'obtenir directement l'existence de solution pour le problème à frontière libre : il faudrait démontrer que la limite hydrodynamique présente une frontière et que celle-ci est suffisamment régulière, pour ensuite revenir à l'interprétation probabiliste et en déduire que c'est une solution du problème à frontière libre.

Mentionnons les travaux d'Atar [Ata23] sur la limite hydrodynamique d'une classe de modèles de particules avec branchement, injection et sélection : le déplacement des particules se fait selon des processus de diffusion sur la droite réelle à coefficients réguliers; les particules branchent à taux constant; et des particules sont ajoutées et retirées du système à des instants et des positions aléatoires. Il démontre alors sous certaines hypothèses sur les lois d'injection et de sélection<sup>[11]</sup> que ce processus admet une limite hydrodynamique qui est l'unique solution d'une formulation faible d'un problème aux dérivées partielles généralisant le problème (2). Ce problème ne s'interprète pas toujours comme une formulation faible d'un problème à frontière libre, mais dans le cas où le mouvement des particules est brownien, où l'on ne considère aucune injection et où l'on fait coïncider les instants de sélection avec les instants de branchement, on retrouve le résultat du Théorème A : le processus étudié par Atar est dans ce cas-là le  $N$ -BBM, et la solution du problème à frontière libre (2) donnée par le Théorème B coïncide avec l'unique solution de la formulation faible donnée par le Théorème 2.5 dans [Ata23].

<sup>10</sup>Ce résultat est démontré par Berestycki et Tough [BT24] (Théorème B.1.), voir également [CCCS11] (Théorème 1, point 4, et Proposition 1).

<sup>11</sup>Voir [Ata23], Définition 2.2 et Hypothèse 2.3. Le résultat mentionné est le Théorème 2.5.

## 1.2 Résultats principaux

Nous avons cherché à montrer l'existence d'une limite hydrodynamique pour un  $N$ -BMP dont les particules ont des trajectoires plus générales qu'un mouvement brownien standard. Considérons donc un processus de Markov<sup>[2]</sup> (homogène en temps)  $(X_t)_{t \geq 0}$  sur la droite réelle, défini sur un espace filtré  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0})$  muni d'une famille de mesures de probabilité  $(\mathbb{Q}_x)_{x \in \mathbb{R}}$  vérifiant  $\mathbb{Q}_x(X_0 = x) = 1$ . On fait de plus l'hypothèse classique que la filtration  $(\mathcal{F}_t)_{t \geq 0}$  est complète et continue à droite. On considère une mesure de probabilité  $\mu_0$  sur  $\mathbb{R}$ , et l'on note  $\mathbb{Q}_{\mu_0} = \int_{\mathbb{R}} \mathbb{Q}_x \mu_0(dx)$  la mesure classique sous laquelle  $X_0$  est de loi  $\mu_0$ . On note  $(P_t)_{t \geq 0}$  le semi-groupe de  $(X_t)_{t \geq 0}$ , défini sur l'espace  $\mathcal{B}_b$  des fonctions mesurables bornées de  $\mathbb{R}$  dans  $\mathbb{R}$  par

$$\forall f \in \mathcal{B}_b \quad \forall t \geq 0 \quad \forall x \in \mathbb{R} \quad P_t f(x) = \mathbb{E}_x(f(X_t)),$$

où  $\mathbb{E}_x$  désigne l'espérance sous  $\mathbb{Q}_x$ . On note aussi  $(P_t(x, \cdot))_{t \geq 0, x \in \mathbb{R}}$  les noyaux de transition associés, au sens où, pour  $t \geq 0$  et  $x \in \mathbb{R}$ ,  $P_t(x, \cdot)$  est la loi de  $X_t$  sous  $\mathbb{Q}_x$ . On a en particulier

$$P_t f(x) = \int_{\mathbb{R}} f(y) P_t(x, dy).$$

Dans la suite, nous appellerons *frontière* toute fonction  $b : (0, +\infty) \rightarrow \overline{\mathbb{R}}$  semi-continue supérieurement. Si  $b$  est une frontière, on note  $b_0 = \overline{\lim}_{t \rightarrow 0^+} b_t \in \overline{\mathbb{R}}$ , et l'on confond  $b$  avec son extension en 0 (qui est semi-continue supérieurement sur  $[0, +\infty)$ ). Pour une frontière  $b$  donnée, on définit le temps de passage à cette frontière par

$$\tau_b = \inf\{s > 0, X_s < b_s\},$$

qui est un temps d'arrêt<sup>[3]</sup>.

Rappelons l'interprétation probabiliste du problème à frontière libre <sup>(2)</sup>: la fonction  $u$ , qui est aussi la limite hydrodynamique du  $N$ -BBM, est la densité de la loi du mouvement brownien conditionné à rester au-dessus d'une frontière qui est atteinte en un temps de loi exponentielle. Cette observation nous a poussés à faire, dans le cadre général, l'hypothèse qu'une telle frontière existe, fournissant le candidat pour la limite hydrodynamique. Plus précisément, nous partons de l'hypothèse qu'il existe une frontière  $b$  (qui dépend de  $\mu_0$ ) telle que :

(H1)  $\tau_b$  est de loi exponentielle  $\mathcal{E}(1)$  sous  $\mathbb{Q}_{\mu_0}$  : pour  $t > 0$ , on a  $\mathbb{Q}_{\mu_0}(\tau_b > t) = e^{-t}$ .

Grâce aux travaux récents de Klump et Savov <sup>[KS23]</sup>, ceci est vérifié sous des hypothèses assez générales pour des diffusions et des processus de Lévy<sup>[4]</sup>. Nous renvoyons à la [Section 1.4.2](#) pour plus de détails à propos de cette hypothèse.

Nous faisons également l'hypothèse suivante de régularité sur le processus sous-jacent, classique pour les processus en temps continu :

(H2)  $X$  est un *processus de Feller* : son semi-groupe est fortement continu et envoie l'espace  $\mathcal{C}_0$  des fonctions continues de  $\mathbb{R}$  dans  $\mathbb{R}$  qui tendent vers 0 en  $\pm\infty$  dans lui-même<sup>[5]</sup>.

Enfin, un élément-clé du couplage utilisé dans notre preuve est l'hypothèse suivante, qui s'applique par exemple aux processus markoviens à trajectoires continues ainsi qu'aux processus de Lévy :

<sup>12</sup>Nous prenons comme base le cadre du chapitre III de <sup>[RY99]</sup>.

<sup>13</sup>On utilise ici le fait que la filtration est complète et continue à droite; voir [Section A.2.1.2](#)

<sup>14</sup>Voir par exemple <sup>[Ber96]</sup>.

<sup>15</sup>Voir <sup>[RY99]</sup>, §III.2.

(H3)  $X$  est *stochastiquement monotone*<sup>16</sup>, c'est-à-dire que si  $x \leq y$  sont deux réels et  $t \geq 0$ , on a  $P_t(x, \cdot) \leq_{\text{st}} P_t(y, \cdot)$ , où  $\leq_{\text{st}}$  désigne l'*ordre stochastique*<sup>17</sup> sur l'ensemble des mesures de probabilité sur  $\mathbb{R}$ .

On introduit la fonction de répartition  $F$  du processus  $(X_t)_{t \geq 0}$  conditionné à rester au-dessus de la frontière :

$$\forall t \geq 0 \forall r \in \mathbb{R} \quad F(t, r) = \mathbb{Q}_{\mu_0}(X_t \leq r | \tau_b > t). \quad (3)$$

Nous avons alors montré que la fonction de répartition  $F$  décrit la limite hydrodynamique du  $N$ -BMP.

**Théorème 1.** *Fixons  $T > 0$  et supposons que les hypothèses (H1), (H2) et (H3) sont vérifiées. Alors pour tout  $t \in [0, T]$ ,  $\sup_{r \in \mathbb{R}} |F^N(t, r) - F(t, r)|$  converge presque sûrement vers 0 quand  $N$  tend vers l'infini. Plus précisément, pour tout  $\beta \in (0, \frac{1}{2})$ , il existe des constantes  $c_1, c_2 > 0$  dépendant uniquement de  $\beta$  et de  $T$  telles que pour tout entier  $N \geq 2$ , on a*

$$\mathbb{P} \left( \sup_{r \in \mathbb{R}} |F^N(t, r) - F(t, r)| > N^{-\beta} \right) \leq c_2 e^{-c_1 N^{\frac{1-2\beta}{3}}}. \quad (4)$$

Une seconde question naturelle est de retrouver la frontière  $b$  grâce au système de particules : au vu du [Théorème 1](#), on peut s'attendre à ce que sous des hypothèses de régularité convenables, quand  $N$  est grand, la position de la particule la plus basse soit proche de la frontière. Cette question a été étudiée dans [DMFPSL19b](#) dans le cas du mouvement brownien avec un branchement non-local, ainsi que par Berestycki, Brunet, Nolen et Penington dans le cas des abeilles browniennes [BBNP22](#). Cette fois, le processus admet encore une limite hydrodynamique, la densité limite s'exprimant à nouveau comme solution d'un problème à frontière libre faisant intervenir une fonction continue  $\mathbf{b} : (0, +\infty) \rightarrow \mathbb{R}_+$ , sous la forme suivante :

$$\begin{cases} \partial_t u = \frac{1}{2} \Delta u + u, & \text{pour } t > 0, \|x\| < \mathbf{b}_t, \\ u(t, x) = 0, & \text{pour } t > 0, \|x\| \geq \mathbf{b}_t, \\ \int_{\{\|x\| < \mathbf{b}_t\}} u(t, x) dx = 1, & \text{pour } t > 0. \end{cases} \quad (5)$$

Les auteurs montrent alors la convergence de la distance maximale entre une particule et l'origine converge quand  $N$  tend vers l'infini, à l'instant  $t \geq 0$ , vers  $\mathbf{b}_t$ .

En plus des hypothèses précédentes, nous ferons l'hypothèse de régularité suivante sur le processus sous-jacent  $(X_t)_{t \geq 0}$  :

(H4) pour tous réels  $x > y$  et  $\kappa > 0$ , on a

$$\mathbb{Q}_x(\tau_y \leq t) \underset{t \rightarrow 0}{=} o(t^\kappa), \quad (6)$$

où  $\tau_y = \inf\{s > 0, X_s < y\}$ .

---

<sup>16</sup>Cette définition est équivalente à ce que pour toute fonction mesurable bornée croissante  $f$  et  $t \geq 0$ , la fonction  $P_t f$  soit elle-même croissante, c'est-à-dire que l'on ait pour  $x \leq y$

$$\mathbb{E}_x(f(X_t)) \leq \mathbb{E}_y(f(X_t)).$$

<sup>17</sup>Rappelons (voir [MS02](#)) que pour deux mesures de probabilité  $\mu$  et  $\nu$  sur  $\mathbb{R}$ , on a la relation  $\mu \leq_{\text{st}} \nu$  si, et seulement si, pour tout réel  $r$ , on a

$$\mu((-\infty, r]) \geq \nu((-\infty, r]).$$

Cette hypothèse est connue dans la littérature mathématique sous l'appellation d'*estimée de Ray*. Elle est vraie pour tout processus de Feller à trajectoires continues (voir [LM12], Section 3.8, sur une idée de Ray [Ray56]). Enfin, nous faisons cette dernière hypothèse de régularité sur la frontière :

(H5) la frontière  $b$  est continue sur  $(0, +\infty)$ .

Notons, pour  $t \geq 0$ ,  $m_t^N = \min_{1 \leq i \leq N} X_t^i$  la position de la particule la plus basse du  $N$ -BMP à l'instant  $t \geq 0$ .

**Théorème 2.** *Supposons que les hypothèses (H1-5) sont vérifiées. Alors, presque sûrement, la trajectoire  $m^N$  de la position de la particule la plus basse converge vers la frontière  $b$ , uniformément en temps sur tout intervalle  $[t_0, T]$ , avec  $0 < t_0 < T$ . Plus précisément, il existe pour tous  $\varepsilon, \kappa > 0$  une constante  $C > 0$  (dépendant de  $\varepsilon, \kappa, T$  mais également de  $X$ ) telle que pour tout entier  $N \geq 2$ , on a*

$$\mathbb{P} \left( \sup_{t \in [t_0, T]} |m_t^N - b_t| > \varepsilon \right) \leq CN^{-\kappa}. \quad (7)$$

Enfin, le résultat est encore vrai avec  $t_0 = 0$  si  $b$  est continue en 0.

On note que le mouvement brownien satisfait les hypothèses<sup>[18]</sup> (H1-5), et les Théorèmes [1] et [2] s'appliquent donc au cas du  $N$ -BBM.

Dans ce cadre, le Théorème [1] permet d'améliorer la borne obtenue<sup>[19]</sup> dans [DMFPSL19b] : le résultat<sup>[20]</sup> assure l'existence pour tous  $\beta \in (0, \frac{1}{2})$  de constantes  $c'_2, c' > 0$  telles que pour tous  $N$  et  $n$  supérieurs à des quantités dépendant encore uniquement de  $\beta$  et  $T$ , on a, pour  $t_k = k2^{-n}T$ ,

$$\mathbb{P} \left( \max_{1 \leq k \leq 2^n} \sup_{r \in \mathbb{R}} |F^N(r, t_k) - F(r, t_k)| > N^{-\beta} \right) \leq c'_2 2^n e^{-c' 2^n} N^{-\frac{1}{6}}.$$

Le Théorème [1] donne, pour tout  $\beta \in (0, \frac{1}{2})$ ,

$$\mathbb{P} \left( \max_{1 \leq k \leq 2^n} \sup_{r \in \mathbb{R}} |F^N(r, t_k) - F(r, t_k)| > N^{-\beta} \right) \leq c_2 2^n e^{-c_1 N^{\frac{1-2\beta}{3}}},$$

et nous obtenons donc la possibilité de considérer des valeurs de  $\beta$  jusqu'à l'échelle attendue des fluctuations ( $\beta = \frac{1}{2}$ ), ainsi qu'une décroissance plus rapide de la borne puisqu'elle est exponentielle en une puissance de  $N$ , plutôt que polynomiale. Précisons également que les constantes  $c_1$  et  $c_2$  sont explicites et universelles, au sens où elles ne dépendent pas de la loi initiale  $\mu_0$  ou de la loi du processus sous-jacent  $(X_t)_{t \geq 0}$ , à partir du moment où celles-ci vérifient les hypothèses (H1-3).

Un résultat quantitatif assez proche est donné dans [BBNP22] dans le cas des abeilles browniennes<sup>[21]</sup>. Dans notre contexte, cela correspond au cas du  $N$ -processus de Bessel branchant<sup>[22]</sup> : en

<sup>18</sup>C'est un processus de Feller (voir [Kal02] Théorème 19.10) à trajectoires continues, vérifiant donc les hypothèses (H2-4); l'existence de la frontière (hypothèse (H1)) est connue depuis [CCCS11]. Pour la continuité de cette frontière (hypothèse (H5)), elle est démontrée dans [CCS22] mais s'obtient également par le fait que cette frontière est aussi celle qui intervient dans le problème [2], dont on sait qu'elle est continue [BBP19].

<sup>19</sup>A noter que les auteurs de [DMFPSL19b] abordent également la possibilité du branchement non-local, que nous n'avons pas étudiée.

<sup>20</sup>Voir [DMFPSL19b], Théorème 3.1.

<sup>21</sup>Voir [BBNP22], Lemme 4.9.

<sup>22</sup>Le processus de Bessel est un processus de Feller à trajectoires continues sur  $[0, +\infty)$ . Bien que le Théorème [1] ait été formulé et démontré pour des processus de Feller sur  $\mathbb{R}$ , nous sommes convaincus que la même preuve s'applique au cas de processus sur  $[0, +\infty)$ , même si nous n'avons pas rédigé l'argument de manière détaillée. L'existence de la frontière découle encore de [CCCS11], et sa régularité de [CCS22] et de [BBNP22].

effet, dans le processus des abeilles browniennes, la norme euclidienne de chaque particule suit un processus de Bessel, si bien que l'évolution de ces  $N$  normes euclidiennes se décrit exactement comme un  $N$ -BMP, à ceci près que l'on tue la particule la plus haute plutôt que la plus basse, à chaque instant de branchement. Dans ce contexte<sup>23</sup>, le [Théorème 1](#) représente une amélioration de la vitesse de convergence obtenue précédemment.

Pour la question de la position de la particule la plus basse, le seul résultat existant à notre connaissance est celui obtenu pour les abeilles browniennes : Berestycki, Brunet, Nolen et Penington montrent<sup>24</sup> que si  $M_t^N$  désigne la norme de la particule la plus éloignée de l'origine, et  $\mathfrak{b}_t$  la frontière à nouveau, il existe un réel  $\kappa_1 > 1$  tel que pour tout  $\varepsilon > 0$ ,  $0 < t_0 < T$ , et  $N$  assez grand, on a

$$\mathbb{P} \left( \sup_{t \in [t_0, T]} M_t^N - \mathfrak{b}_t > \varepsilon \right) \leq N^{-\kappa_1}.$$

Ils démontrent et utilisent l'autre inégalité, à  $t$  fixé, dans la section 5 de [\[BBNP22\]](#), avec un autre exposant  $\kappa_2 > 1$  :

$$\mathbb{P} (M_t^N - \mathfrak{b}_t < -\varepsilon) \leq N^{-\kappa_2}.$$

Le [Théorème 2](#), dans sa version étendue au  $N$ -processus de Bessel branchant, fournit un contrôle sur  $\sup_{t \in [t_0, T]} |M_t^N - \mathfrak{b}_t|$ , et représente une amélioration de la vitesse de convergence de la borne à toute puissance négative de  $N$ .

Notons que c'est l'utilisation de l'estimée de Ray qui entraîne l'apparition de la puissance négative de  $N$  dans la majoration [\(7\)](#). Un meilleur contrôle du membre de gauche dans [\(6\)](#) permet d'obtenir un meilleur contrôle dans [\(7\)](#) : pour le mouvement brownien par exemple, ou encore dans le cas du processus de Bessel, il existe une constante  $C' > 0$  telle que pour  $t$  suffisamment petit et  $x > y$ , on a

$$\mathbb{Q}_x(\tau_y \leq t) \leq C' e^{-\frac{|x-y|^2}{4t}},$$

de sorte que l'on peut à nouveau avoir une décroissance exponentielle du même type que [\(4\)](#) pour le membre de gauche de [\(7\)](#).

Dans le cas plus général où le processus sous-jacent est une diffusion à coefficients réguliers<sup>25</sup>, les résultats d'Atar [\[Ata23\]](#) entraînent l'existence d'une limite hydrodynamique au  $N$ -BMP, mais ne fournissent pas de contrôle quantitatif sur la convergence du type de celui que nous obtenons dans le [Théorème 1](#). La question de la position de la particule la plus basse étudiée au [Théorème 2](#) n'est pas non plus abordée dans [\[Ata23\]](#).

### 1.3 Stratégie de preuve

L'idée générale utilisée pour démontrer l'existence d'une limite hydrodynamique, mise en œuvre de manière différente pour le [Théorème A](#) et pour le [Théorème 1](#), est d'encadrer la fonction de répartition empirique du  $N$ -BMP par deux fonctions auxiliaires (appelées en général barrières supérieure et inférieure). Nous comparons ici la méthode utilisée dans [\[DMFPSL19a\]](#) pour obtenir le [Théorème A](#) avec celle que nous utilisons dans le [Chapitre 1](#) pour démontrer les [Théorèmes 1](#) et [2](#).

<sup>23</sup>Le processus obtenu est donc l'opposé d'un  $N$ -BMP pour lequel les particules sont dirigées par l'opposé d'un processus de Bessel.

<sup>24</sup>Voir [\[BBNP22\]](#), Proposition 1.6.

<sup>25</sup>Ces processus sont encore des processus de Feller à trajectoires continues pour lesquelles l'existence et la continuité de la frontière sont démontrées dans [\[CCCS11\]](#), [\[CCS22\]](#).

### 1.3.1 L'approche de De Masi, Ferrari, Presutti et Soprano-Loto

Les preuves données dans [DMFPSL19a], puis [DMFPSL19b, BBNP22], se basent sur une comparaison avec deux processus auxiliaires dans lesquels la sélection est effectuée à des instants déterministes régulièrement espacés, les particules se déplaçant et branchant librement entre ces instants.

Comme pour le  $N$ -BBM, chacun des deux processus auxiliaires possède initialement  $N$  particules dont les positions à l'instant  $t = 0$  forment une famille i.i.d. selon une loi de densité  $u_0$ . Pour un pas de temps  $\delta > 0$  fixé, on les construit ensuite de la manière suivante. Sur chaque intervalle de temps déterministe  $(k\delta, (k+1)\delta)$ , on laisse les processus évoluer comme des BBM, c'est-à-dire que les particules se déplacent et branchent. La taille de la population peut donc changer au cours du temps, et l'on peut avoir plus de  $N$  particules. Pour le premier processus auxiliaire, on procède alors par *post-sélection*<sup>26</sup> : on retire du système les particules dans l'ordre de leur position à l'instant  $(k+1)\delta$ , en partant de la plus basse, jusqu'à obtenir  $N$  particules. Pour le second processus, on procède par *pré-sélection*<sup>27</sup> : on conserve uniquement les familles<sup>28</sup> de particules issues des individus les plus hauts à l'instant  $k\delta$ , de manière à être le plus proche possible de  $N$  particules sans dépasser ce seuil.

On peut ensuite coupler ces processus auxiliaires avec le  $N$ -BBM de manière à encadrer la population de particules du  $N$ -BBM par celle des deux processus auxiliaires.

Etudier cette construction alternant le comportement du BBM avec une sélection à un instant déterministe est plus simple que d'étudier directement le  $N$ -BBM, dans lequel l'étape de sélection introduit une dépendance complexe entre les trajectoires. En particulier, on dispose de la formule many-to-one<sup>29</sup>, qui assure que si  $(B_t^i)_{1 \leq i \leq N_t}$  est un BBM issu d'une particule située en  $x$  à l'instant 0, on a

$$\mathbb{E}_x \left( \sum_{i=1}^{N_t} f(B_{[0,t]}^i) \right) = e^t \mathbb{E}_x (f(B_{[0,t]})),$$

où  $N_t$  désigne le nombre de particules présentes à la date  $t \geq 0$ ,  $B_t^i$  la position de la  $i$ -ème particule à la date  $t$ ,  $B_{[0,t]}$  désignant la trajectoire de  $B$  entre 0 et  $t$  d'un mouvement brownien standard issu de  $x$ , et  $f$  est une fonction mesurable de l'espace  $\mathcal{D}([0, t], \mathbb{R})$  des fonctions càdlàg dans  $\mathbb{R}$ .

On peut alors étudier la limite hydrodynamique des deux processus auxiliaires décrits précédemment. A cette fin, on introduit, pour  $t \geq 0$  et  $m \geq 0$ , les opérateurs de diffusion  $G_t$  et de troncature  $C_m$  sur  $\mathcal{B}_b$ , définis par

$$\forall f \in \mathcal{B}_b \quad \forall x \in \mathbb{R} \quad \begin{cases} G_t f(x) = \int_{\mathbb{R}} f(y) \frac{e^{-\frac{(x-y)^2}{2t}}}{\sqrt{2\pi t}} dy, \\ C_m f(x) = f(x) \mathbf{1}_{\{\int_x^{+\infty} f(y) dy \leq m\}}. \end{cases}$$

Heuristiquement, laisser le processus évoluer selon un BBM sur l'intervalle de temps  $(k\delta, (k+1)\delta)$  revient, à la limite hydrodynamique, à appliquer l'opérateur  $e^\delta G_\delta$  à la densité de particules à l'instant  $k\delta$ , l'exponentielle correspondant à la croissance de masse due à la formule many-to-one, et l'opérateur de diffusion au déplacement brownien des particules. D'autre part, effectuer une sélection revient à appliquer un opérateur de troncature ( $C_1$  dans le cas de la post-sélection,

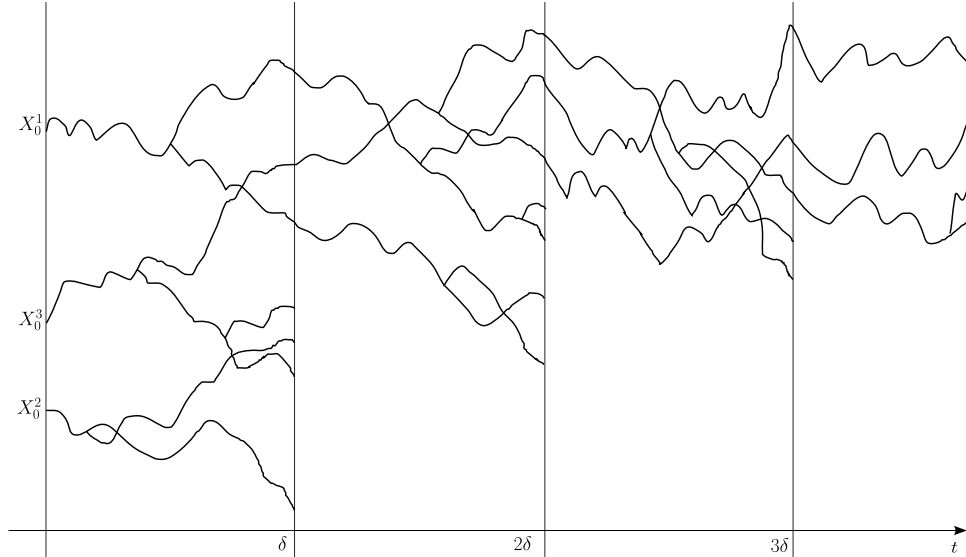
<sup>26</sup>Voir Figure 2a

<sup>27</sup>Voir Figure 2b

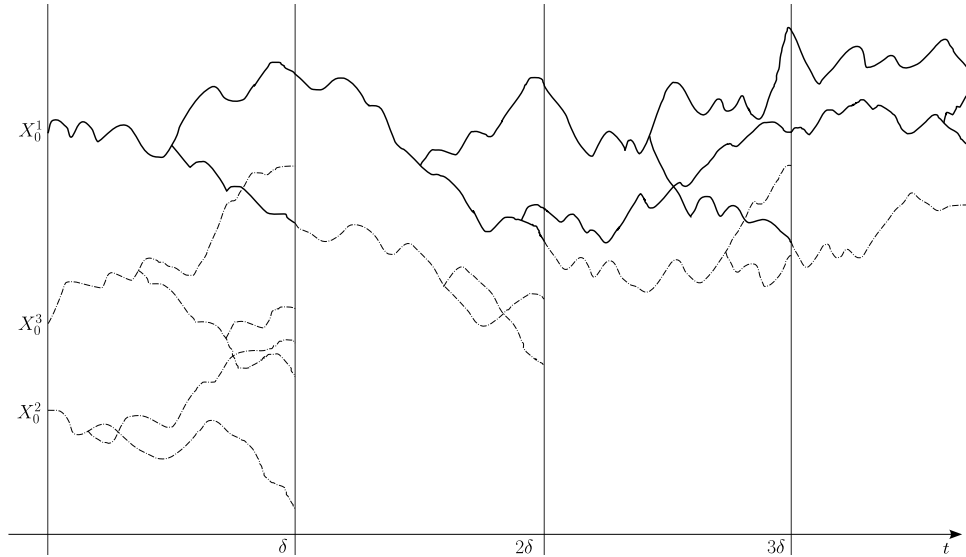
<sup>28</sup>On appelle famille l'ensemble des particules issues par branchement successifs d'une même particule à l'instant  $k\delta$ .

<sup>29</sup>Voir [HR17].





(a) Processus obtenu par post-sélection : on ne garde que les  $N$  particules les plus hautes à la fin de chaque intervalle de temps de taille  $\delta$ .



(b) Processus obtenu par pré-sélection : on ne garde que les familles issues des particules les plus hautes au début de chaque intervalle de temps de taille  $\delta$ , de manière à avoir au plus  $N$  particules à la fin de l'intervalle de temps; la taille de la population évolue de manière aléatoire entre 0 et  $N$ .

Figure 2: Processus auxiliaires utilisés dans [\[DMFPSL19a\]](#) pour la comparaison avec un 3-BBM.

$C_{e^{-\delta}}$  dans le cas de la pré-sélection). Finalement, la limite hydrodynamique des processus auxiliaires s'obtient en itérant ces opérateurs : la limite hydrodynamique de celui construit par post-sélection est, à l'instant  $k\delta$ , la loi de densité  $(C_1 e^\delta G_\delta)^k u_0$ . De même, celle du processus construit par pré-sélection est  $(e^\delta G_\delta C_{e^{-\delta}})^k u_0$ .

Les auteurs montrent alors que ces deux fonctions de densité sont proches quand  $\delta$  tend vers 0 et  $k\delta$  tend vers  $t$ , et ils développent un argument de compacité et de monotonie pour montrer que ces deux fonctions admettent une même limite; cette fonction est la fonction  $\psi$  (voir [Théorème A](#)).

Le reste de la preuve du [Théorème A](#) se déroule donc comme suit : comme le  $N$ -BBM est

encadré par les barrières stochastiques, dont les limites hydrodynamiques à  $\delta > 0$  fixé ont la même limite quand  $\delta$  tend vers 0, et que lui-même ne dépend pas de  $\delta$ , sa limite hydrodynamique est effectivement cette fonction  $\psi$ .

### 1.3.2 Notre approche

Notre approche diffère de celle présentée dans le paragraphe précédent par le fait qu'elle repose fortement sur l'hypothèse *a priori* d'existence de la frontière  $b$  atteinte en un temps de loi exponentielle (hypothèse (H1)) : nous comparons alors le  $N$ -BMP avec un processus auxiliaire que nous appelons *b-processus de Markov branchant* ( $b$ -BMP), qui est un processus de Markov branchant dont les particules sont tuées quand elles passent sous la frontière  $b$ . Ce processus est une variante du BBM avec absorption, étudié en détail dans le cas où la frontière est linéaire depuis [Kes78]. Comme pour le  $N$ -BMP, les particules sont soumises à un branchement dyadique local et se déplacent indépendamment, à partir de leur naissance, selon la loi du processus sous-jacent. En revanche, la règle de sélection est différente : on tue chaque particule au premier instant où elle passe strictement sous la frontière  $b$ . La taille de la population varie donc avec le temps, et le processus peut même s'éteindre (voir [HHK06, Mai13, BBS15] et les références qui s'y trouvent pour des travaux plus récents dans le cas du BBM avec absorption le long d'une frontière linéaire). Cependant, la formule many-to-one<sup>30</sup> assure qu'il y a en moyenne une compensation exacte entre le phénomène de branchement et le fait de tuer les particules qui touchent la frontière. En effet, la taille moyenne d'un BMP sans sélection ayant  $N$  particules à l'instant 0 est  $\mathbb{E}_{\mu_0}(N_t) = N e^t$ , et pour une particule donnée, la probabilité de ne pas avoir dépassé la frontière est, sous l'hypothèse (H1),  $\mathbb{P}_{\mu_0}(\tau_b > t) = e^{-t}$ , de sorte que le nombre moyen de particules présentes à l'instant  $t$  est exactement égal à  $N$ . On s'attend donc à ce que le  $b$ -BMP et le  $N$ -BMP se comportent de façon similaire quand le nombre de particules est suffisamment grand.

La comparaison entre le  $N$ -BMP et le  $b$ -BMP passe alors par un couplage entre le  $N$ -BMP et deux versions du  $b$ -BMP qui l'encadrent. Plus précisément, on construit un couplage entre un  $N$ -BMP  $\mathcal{X}$  et des  $b$ -BMP  $\mathcal{X}^+$  et  $\mathcal{X}^-$  partant respectivement de  $\lceil N(1 + \delta) \rceil$  et  $\lfloor N(1 - \delta) \rfloor$  particules, tel que pour tout  $t \geq 0$  et tout  $r \in \mathbb{R}$ , on a avec forte probabilité

$$\#(\mathcal{X}_t^- \cap [r, +\infty)) \leq \#(\mathcal{X}_t \cap [r, +\infty)) \leq \#(\mathcal{X}_t^+ \cap [r, +\infty)), \quad (8)$$

ce qui assure un encadrement de la fonction de répartition empirique du  $N$ -BMP par celles des  $b$ -BMP : si l'on note  $N_t^\pm$  la taille de  $\mathcal{X}_t^\pm$ , et

$$F^{N,\pm}(t, r) = \frac{1}{N} \sum_{i=1}^{N_t^\pm} \mathbf{1}_{\{X_t^{i,\pm} \leq r\}}$$

leurs fonctions de répartitions normalisées, [8] donne

$$\frac{\lfloor N(1 - \delta) \rfloor}{N} - F^{N,-}(t, r) \leq 1 - F^N(t, r) \leq \frac{\lceil N(1 + \delta) \rceil}{N} - F^{N,+}(t, r). \quad (9)$$

L'intérêt de cet encadrement est que le  $b$ -BMP se révèle plus simple à étudier (voir Section L.2) que le  $N$ -BMP, dont l'étude est rendue difficile par les interactions qui ont lieu entre les particules au moment de la sélection. En effet, le  $b$ -BMP partant de  $n$  particules est constitué de  $n$  copies **indépendantes** de  $b$ -BMP n'ayant qu'une seule particule à l'instant initial. Nous montrons alors la Proposition L.1, qui affirme que la limite hydrodynamique du  $b$ -BMP est la loi du processus

<sup>30</sup>Cette formule est également valable pour le  $N$ -BMP, voir [HR17].

$(X_t)_{t \geq 0}$  conditionné à ce que la trajectoire ne passe pas sous la frontière  $b$ . On peut alors passer à la limite quand  $N$  tend vers l'infini dans [9], puisque les processus  $\mathcal{X}^\pm$  ont, à  $\delta$  fixé, chacun une limite hydrodynamique; plus précisément, la normalisation assure que  $F^{N,\pm}(t, r)$  converge vers  $(1 \pm \delta)F(t, r)$ . En faisant alors tendre  $\delta$  vers 0, on obtient la convergence de  $F^N(t, r)$ , et donc la limite hydrodynamique du  $N$ -BMP.

Pour construire le couplage de manière à assurer l'ordre [8] entre ses différentes composantes, on a besoin des ingrédients suivants :

- partir d'une configuration initiale respectant le nombre de particules pour  $\mathcal{X}$ ,  $\mathcal{X}^+$ , et  $\mathcal{X}^-$ , ainsi que [8],
- appairer chaque particule de  $\mathcal{X}^-$  avec une particule de  $\mathcal{X}$ , et chaque particule de  $\mathcal{X}$  avec une particule de  $\mathcal{X}^+$ , et attribuer à chaque paire des trajectoires respectant la loi du processus sous-jacent et qui ne se croisent pas; nous expliquons en détail dans la Section 2 comment l'hypothèse (H3) (monotonie stochastique) permet de réaliser cela en toute généralité,
- prescrire aux instants de branchement et de sélection la manière de reformer les paires affectées par une mort ou une naissance.

Cette construction n'est possible que tant que la taille de  $\mathcal{X}^+$  est au moins égale à  $N$  et celle de  $\mathcal{X}^-$  au plus égale à  $N$ . Grâce à la marge obtenue en débutant respectivement avec  $\lceil N(1 + \delta) \rceil$  et  $\lfloor N(1 - \delta) \rfloor$  particules et au contrôle de l'évolution de la taille du  $b$ -BMP, on vérifie que cette condition est respectée avec forte probabilité jusqu'à tout horizon de temps fixé. La construction du couplage fait l'objet de la Proposition I.4, et nous renvoyons à sa démonstration (Section I.3) pour plus de détails.

Notons enfin qu'un outil technique important pour construire ce couplage de manière complètement rigoureuse est le recollement de processus markoviens : on décrit le comportement du couplage entre les instants de branchements et de sélection de particules, et l'on prescrit également la manière dont se déroulent ces deux phénomènes. Cela revient à recoller des morceaux de trajectoires markoviennes pour créer un nouveau processus markovien. Nous utilisons dans ce but la construction d'Ikeda, Nagasawa et Watanabe [INW68b]; malgré le fait qu'elle ait des applications diverses liées à la création de nouveaux processus à partir de processus donnés (comme les processus de branchement, qui sont la motivation originale de cette construction [INW68a], les mélanges de processus [GH69] ou la renaissance de processus, comme étudiée par Meyer [Mey75]), cette construction semble avoir été peu reprise, avant notamment une généralisation par Werner [Wer21]. Nous avons donc adapté cette construction au contexte qui nous intéressait et démontré dans l'Annexe A plusieurs résultats qui permettent de prouver que le couplage a bien les propriétés requises (voir Proposition A.4).

## 1.4 Discussion

### 1.4.1 Limite hydrodynamique et é.d.p. pour les modèles de branchement-sélection

Pour de nombreux modèles stochastiques particuliers décrivant l'évolution de populations, il a été démontré ou conjecturé l'existence d'une limite hydrodynamique solution d'un problème aux dérivées partielles.

Les problèmes à frontière libre comme [2] permettent de faire le lien avec des systèmes de particules soumises à des dynamiques de sélection des particules extrémales : dans [CDMGPI6], Carinci, De Masi, Giardinà et Presutti considèrent un système de  $N$  mouvements browniens indépendants dans lequel, aux instants de saut d'un processus de Poisson d'intensité  $\lambda N$ , la particule la plus haute saute à la position 0, et démontrent le lien avec le problème de Stefan<sup>31</sup>

<sup>31</sup>Ce type de problèmes est bien documenté, voir par exemple [Cas05].

suivant : on cherche une fonction  $u$  de classe  $C^{1,2}$  sur  $(0, +\infty) \times \mathbb{R}$  et une fonction  $b : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  continue telles que

$$\begin{cases} \partial_t u = \frac{1}{2} \partial_x^2 u, & t \geq 0, x \in [0, b_t], \\ \partial_x u(t, 0) = \partial_x u(t, b_t) = -2\lambda, & t \geq 0, \\ u(t, b_t) = 0, & t \geq 0 \end{cases} \quad (10)$$

Ici, la frontière libre est déterminée par une condition de flux, puisque l'on demande que les flux entrant et sortant se compensent, traduisant le fait que, dans le système de particules, l'injection de masse à l'origine compense la perte de masse en haut du nuage de points.

L'étude du cas avec branchement non-local a été initiée par Durrett et Remenik [DR11] pour des particules dont la position ne change pas entre leur instant de naissance et leur instant de mort. Une récente généralisation a été effectuée par Atar [Ata24], qui considère des particules vivant dans  $\mathbb{R}^d$  et pour lesquelles la sélection se fait par une fonction d'adaptation assez générale<sup>32</sup>. Dans [DMFPSL19b], les auteurs étudient le cas du  $N$ -BBM avec branchement non-local. Pour chacun de ces modèles, la limite hydrodynamique fait intervenir un problème à frontière libre, le terme de croissance de masse étant remplacé par une convolution avec la mesure  $K$  qui détermine le décalage de chaque nouvelle particule par rapport à son parent : on obtient par exemple pour le cas brownien l'équation

$$\partial_t u = \frac{1}{2} \partial_x^2 u + K * u, \quad \text{pour } x > b_t.$$

D'autres modèles du même type possèdent une limite hydrodynamique caractérisée comme solution d'é.d.p. faisant intervenir leurs règles spécifiques de branchement et de sélection (voir par exemple [ABBP19, GJM20, GSL21]).

Revenons pour finir sur les travaux d'Atar [Ata23]. Il considère un modèle dans lequel les trajectoires des particules sont décrites par des é.d.s. à coefficients réguliers et bornés; les particules sont soumises à une dynamique de branchement dyadique local, ainsi qu'à un mécanisme d'injection et de sélection. Ce modèle inclue<sup>33</sup> notamment le  $N$ -BBM [DMFPSL19a] et le modèle présenté dans [CDMGPI6]. Atar démontre que ce modèle admet une limite hydrodynamique décrite par une formulation faible qui généralise les problèmes à frontière libre observés jusqu'ici. Il montre également que dans les cas où la formulation forte du problème à frontière libre associé admet une solution (donc en particulier dans le cas du  $N$ -BBM [BBP19]), celle-ci coïncide avec la solution de la formulation faible. La méthode d'Atar repose sur les trois points suivants :

- montrer que la formulation faible ne peut donner lieu qu'à au plus une solution,
- montrer que la famille de lois empiriques du système de particules, indicée par le nombre de particules, est tendue, et admet donc des sous-limites,
- enfin, montrer grâce à un argument de martingale utilisant la formule d'Itô que toutes ces sous-limites sont nécessairement solution de la formulation faible.

L'une des clés de cette approche est donc d'obtenir une formulation faible pour laquelle on peut démontrer l'unicité des solutions. Pour une formulation forte faisant explicitement intervenir

<sup>32</sup>Contrairement aux cas traités jusque-là : à part l'identité dans  $\mathbb{R}$ , le cas de fonctions symétriques décroissantes est étudié dans [Bec19]; cela correspond en prenant l'opposée de la norme euclidienne au cas des abeilles browniennes [BBNP22] en dimension supérieure.

<sup>33</sup>Dans le modèle d'Atar, ne prescrire aucune injection et synchroniser les instants de sélection avec les instants de branchement donne le  $N$ -BBM; demander un taux de branchement nul et synchroniser les instants de sélection avec les instants d'injection donne le modèle de [CDMGPI6].

une frontière libre, établir l'unicité implique souvent de pouvoir connaître *a priori* la régularité de cette frontière [DMFP15]. Or, comme nous le verrons dans la section suivante, cette question est déjà assez peu traitée dans le cas du problème à frontière libre [2] et son extension au modèle étudié par Atar, avec injection et sélection de particules selon des règles moins simples que celles choisies pour le  $N$ -BMP, semble encore moins facile d'accès.

#### 1.4.2 Problème inverse du premier temps de passage

L'hypothèse centrale (H1) pour notre méthode de preuve est l'existence, partant d'une loi initiale  $\mu_0$  pour le processus  $(X_t)_{t \geq 0}$ , d'une frontière<sup>34</sup>  $b : (0, +\infty) \rightarrow \overline{\mathbb{R}}$  dont le temps de passage  $\tau_b$  suit la loi exponentielle  $\mathcal{E}(1)$ .

Cette question est traitée dans la littérature mathématique sous l'appellation de *problème inverse du premier temps de passage*, et semble remonter à une question de Shiryaev dans les années 1970 sur la possibilité de réaliser une loi sur  $(0, +\infty)$  donnée comme temps d'arrêt d'un mouvement brownien. La question a ensuite évolué pour arriver à la formulation avec un temps de passage d'une frontière.

Le problème direct consiste, quant à lui, à déterminer la loi du temps de passage d'une frontière donnée. Dans le cas du mouvement brownien, l'existence d'une densité pour une frontière continument dérivable a été démontrée par Strassen [Str67], et des équations implicites sur cette densité sont connues<sup>35</sup>, avec notamment l'équation obtenue par Peskir<sup>36</sup> : notons

$$\Psi(x) = \int_x^{+\infty} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} dy$$

la fonction de survie de la loi normale standard; on a alors pour tous  $x \leq b_t$

$$\Psi\left(\frac{x}{\sqrt{t}}\right) = \int_0^t \Psi\left(\frac{x - b_s}{\sqrt{t - s}}\right) \mathbb{P}(\tau_b \in ds).$$

En revanche, même pour le mouvement brownien, on obtient des formules explicites seulement dans certains cas simples comme les frontières linéaires (formule de Bachelier-Lévy, voir par exemple [ZS09]) et quadratiques [Sal88]. Diverses méthodes d'approximation ou de simulation de cette loi ont également été proposées [WP97, PW01, ZS09, HT16, Pot23].

**Solutions explicites au problème inverse.** Le problème inverse du premier temps de passage n'a été traité qu'encore plus tard. A notre connaissance, il ne donne lieu à une frontière connue explicitement que dans dans des cas très restreints<sup>37</sup>; des méthodes numériques ont aussi été développées pour donner des solutions approchées [ZS09, Abu15].

Nous développons ici un cadre dans lequel on obtient une frontière explicite (constante, en fait) en utilisant la théorie des *lois quasi-stationnaires*, l'enjeu devenant alors de trouver des processus  $(X_t)_{t \geq 0}$  pour lesquels on peut exhiber une loi quasi-stationnaire explicite.

**Définition 1.** On dit qu'une mesure de probabilité  $\nu$  sur  $\mathbb{R}$  est une loi quasi-stationnaire<sup>38</sup> pour  $(X_t)_{t \geq 0}$  si pour tout borélien  $A$ , et  $t \geq 0$ , on a

$$\nu(A) = \mathbb{P}_\nu(X_t \in A | \tau_0 > t), \tag{11}$$

<sup>34</sup>On rappelle que la définition de frontière (voir Section 1.2) inclut la semi-continuité supérieure de la fonction  $b$ , et l'on rappelle que  $\tau_b$  est un temps d'arrêt pour le processus  $(X_t)_{t \geq 0}$ .

<sup>35</sup>Voir [Ler86, PS06] pour des références sur cette question, initialement étudiée par Schrödinger [Sch15].

<sup>36</sup>Voir [Pes02], Théorème 3.1.

<sup>37</sup>A part ceux où l'on choisit *a priori* une loi de temps d'arrêt solution du problème direct pour une frontière connue.

<sup>38</sup>On considère ici des lois quasi-stationnaires pour des processus tués en 0, d'où le rôle particulier joué par cet état; on fait ce choix car c'est dans ce cadre-là que la littérature mathématique est développée.

où  $\tau_0$  est le temps de passage en 0 (c.à.d.  $\tau_b$  dans le cas  $b \equiv 0$ ).

La propriété de Markov assure que dans ce cas, la loi de  $\tau_0$  sous  $\mathbb{P}_\nu$  est exponentielle<sup>39</sup> et, quitte à changer l'échelle temporelle, on peut choisir que ce soit la loi  $\mathcal{E}(1)$  : la loi  $\nu$  est donc non seulement une loi de départ pour laquelle l'hypothèse (H1) est vérifiée par la frontière  $b \equiv 0$ , mais également la loi, à tout instant positif, du processus conditionné à rester au-dessus de cette frontière jusqu'à cet instant, dans le cas où l'on choisit cette loi de départ. En particulier, notre **Théorème 1** assure que pour un processus vérifiant les hypothèses (H2) et (H3) ayant une loi quasi-stationnaire  $\nu$ , le  $N$ -BMP partant de  $\nu$  admet  $\nu$  comme limite hydrodynamique à tout instant.

*Loi quasi-stationnaire pour le mouvement brownien avec dérive.* L'étude des lois quasi-stationnaires pour le mouvement brownien tué (voir **MSM94**) permet d'obtenir une loi de départ explicite pour laquelle le problème admet une solution sous la forme d'une barrière linéaire. Nous rappelons ici comment obtenir cette solution en utilisant une méthode d'inversion de transformée de Laplace.

On considère donc un mouvement brownien  $(B_t^\alpha)_{t \geq 0}$  avec volatilité 1 et dérive  $-\alpha$ ,  $\alpha$  étant un réel donné. On cherche à déterminer quelles dérivées sont possibles pour qu'il existe une loi de départ  $\mu_0$  sur  $\mathbb{R}_+$  telle que le temps d'atteinte  $\tau_0^\alpha$  de 0 par ce mouvement brownien, défini par

$$\tau_0^\alpha = \inf\{t > 0, B_t^\alpha < 0\},$$

soit de loi exponentielle  $\mathcal{E}(1)$ . Pour une dérive  $\alpha \in \mathbb{R}$  donnée, le principe de réflexion (cas  $\alpha = 0$ ) puis le théorème de Girsanov permettent de calculer la transformée de Laplace du temps d'atteinte de 0 : on a pour  $x, \lambda > 0$

$$\mathbb{E}_x(e^{-\lambda\tau_0^\alpha}) = e^{x(\alpha - \sqrt{\alpha^2 + 2\lambda})}.$$

Ainsi, pour une loi de départ  $\mu_0$  à densité  $u_0$  à support sur  $\mathbb{R}_+$ , on a

$$\mathbb{E}_{\mu_0}(e^{-\lambda\tau_0^\alpha}) = \int_0^{+\infty} e^{x(\alpha - \sqrt{\alpha^2 + 2\lambda})} u_0(x) dx = \mathfrak{L}u_0(-\alpha + \sqrt{\alpha^2 + 2\lambda}),$$

où  $\mathfrak{L}u_0$  désigne la transformée de Laplace de la fonction  $\rho$ . Or on souhaite que  $\tau_0^\alpha$  soit de loi exponentielle, et l'on doit donc résoudre l'équation fonctionnelle

$$\forall \lambda > 0, \quad \mathfrak{L}u_0(-\alpha + \sqrt{\alpha^2 + 2\lambda}) = \frac{1}{1 + \lambda}.$$

On va donc inverser l'argument de la transformée de Laplace et appliquer une transformée de Laplace inverse : on a pour  $p \geq \alpha$

$$p = -\alpha + \sqrt{\alpha^2 + 2\lambda} \iff \lambda = \frac{p^2}{2} + \alpha p,$$

de sorte que l'on a finalement

$$\mathfrak{L}u_0(p) = \frac{1}{1 + \frac{p^2}{2} + \alpha p}.$$

Dans le cas favorable où le dénominateur admet deux racines réelles  $a_\pm = -\alpha \pm \sqrt{\alpha^2 - 2}$ , donc sous la condition  $|\alpha| \geq \sqrt{2}$ , cette équation admet une solution, donnée par

$$u_0^\alpha(x) = \frac{2}{a_+ - a_-} (e^{a_+x} - e^{a_-x}) = \begin{cases} \frac{2e^{-\alpha x}}{\sqrt{\alpha^2 - 2}} \operatorname{sh}(x\sqrt{\alpha^2 - 2}) & \text{si } |\alpha| > \sqrt{2}, \\ 2xe^{-x\sqrt{2}} & \text{si } |\alpha| = \sqrt{2}. \end{cases} \quad (12)$$

---

<sup>39</sup>Voir **MV12**.

Dans le cas  $\alpha \leq -\sqrt{2}$ , ces fonctions ne sont pas intégrables, mais dans le cas  $\alpha \geq \sqrt{2}$ , elles le sont et définissent bien des densités de probabilité à support dans  $\mathbb{R}_+$ . Finalement, pour  $\alpha \geq \sqrt{2}$ , le temps d'atteinte de 0 pour le mouvement brownien avec dérive  $-\alpha$ , issu d'une position suivant la densité  $u_0^\alpha$ , est de loi exponentielle  $\mathcal{E}(1)$ . Ainsi, pour un mouvement brownien standard dont la position initiale suit cette loi  $u_0^\alpha dx$ , le temps d'atteinte de la frontière linéaire  $t \mapsto \alpha t$  est encore de loi  $\mathcal{E}(1)$ , et l'hypothèse (H1) est donc vérifiée.

*Autres lois quasi-stationnaires explicites.* Plus généralement, l'étude des distributions quasi-stationnaires remonte aux travaux de Yaglom [Yag47] sur le comportement de processus de branchements sous-critiques conditionnés à survivre<sup>40</sup>. Divers critères<sup>41</sup> sont donnés pour l'existence de lois quasi-stationnaires dans le cas des processus de Lévy, de processus sur des espaces d'états discrets, de diffusions. Pour ce qui est des formules explicites, qui nous intéressent ici, Kyprianou et Palmowski [KP06] obtiennent à nouveau le cas du mouvement brownien avec dérive donné précédemment, mais aussi celui de certains processus de Lévy à sauts négatifs, puis Yamato [Yam23] obtient notamment une loi explicite dans le cas du processus de Poisson composé avec dérive négative et sauts de loi exponentielle. Certains cas pour des processus à espace d'état discret (comme certains processus de naissance et de mort [vD91], par exemple) ont des lois quasi-stationnaires explicites, donnant à nouveau un cadre dans lequel la frontière constante  $b \equiv 0$  vérifie l'hypothèse (H1).

**Existence d'une frontière.** Les premières avancées en lien avec le problème inverse du premier temps de passage sont dues à Dudley, Gutmann et Anulova [DG77, Anu80], mais dans un cadre où le temps d'arrêt ne se réalise pas exactement<sup>42</sup> comme un temps de passage sous une barrière pour le mouvement brownien. Au début des années 2000, un premier lien formel est fait par Avellaneda et Zhu [AZ01] entre le problème inverse du premier temps de passage et les problèmes à frontière libre : étant donné une fonction de survie  $R : (0, +\infty) \rightarrow [0, 1]$  caractérisant la loi d'un temps aléatoire, on considère le problème (13) suivant, dont les inconnues sont  $u : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$  de classe  $C^{1,2}$  sur  $(0, +\infty) \times \mathbb{R}$  et  $b : \mathbb{R}_+ \rightarrow \mathbb{R}$  :

$$\begin{cases} \partial_t u = A^* u, & \text{pour } x > b_t, \\ u(t, x) = 0 & \text{pour } x \leq b_t, \\ \int_{b_t}^{+\infty} u(t, x) dx = R(t), \end{cases} \quad (13)$$

$A^*$  étant l'adjoint du générateur d'une diffusion  $(X_t)_{t \geq 0}$  solution de l'é.d.s. à coefficients bornés et réguliers

$$dX_t = \beta(X_t) dt + \sigma(X_t) dB_t$$

défini<sup>43</sup> pour  $f$  de classe  $C^2$  par

$$A^* f(x) = \frac{1}{2}(\sigma^2 f)''(x) - (\beta f)'(x).$$

<sup>40</sup>Cela explique le rôle important de l'état 0 dans la littérature sur le sujet, puisque c'est l'état dans lequel les processus de branchement sont tués.

<sup>41</sup>Nous renvoyons à [MV12, CMSM13, vDP13] pour plus de détails sur la littérature autour des lois quasi-stationnaires.

<sup>42</sup>Anulova considère en fait une double barrière symétrique; d'ailleurs, elle montre donc que notre hypothèse est vérifiée dans le cas où  $X$  est un mouvement brownien standard sur  $(-\infty, 0)$  réfléchi en 0.

<sup>43</sup>Si  $u$  est une fonction des deux variables de classe  $C^2$  en espace, on peut donc appliquer l'opérateur  $A^*$  sur la deuxième variable : on définit donc pour  $(t, x) \in \mathbb{R}_+ \times \mathbb{R}$

$$A^* u(t, x) := A^*[u(t, \cdot)](x).$$

Les auteurs conjecturent que la sous-densité de la diffusion  $(X_t)_{t \geq 0}$  restant au-dessus de la frontière  $b$  donnée par

$$u(t, x) = \frac{d}{dx} \mathbb{P}(X_t \leq x, \tau_b > t),$$

est solution du problème [\[13\]](#) si la fonction de survie de  $\tau_b$  est  $R$  [\[AZ01\]](#). Cheng, Chen, Chadam et Saunders [\[CCCS06\]](#) montrent l'existence de solutions de viscosité<sup>44</sup> à [\[13\]](#), avant d'étendre l'étude au cas de diffusions inhomogènes en temps ( $\sigma$  et  $\beta$  dépendent de la variable temporelle) et de faire rigoureusement le lien entre le problème à frontière libre et le problème inverse du premier temps de passage [\[CCCS11\]](#) dans le cas d'un temps aléatoire à densité : ils démontrent l'existence d'une unique solution de viscosité au problème à frontière libre, le fait que celle-ci est l'unique solution du problème inverse du premier temps de passage correspondant, ainsi que le fait que si le problème admet une solution classique, cette dernière est aussi solution du problème inverse du premier temps de passage. Ekström et Janson [\[EJ16\]](#) ont ensuite montré l'existence et l'unicité dans le cas brownien, mais pour un temps aléatoire dont la loi est quelconque (pas nécessairement à densité, donc).

Enfin, la question a été étudiée dans un cadre général par Klump et Savov [\[KS23\]](#), qui établissent l'existence d'une unique solution au problème inverse du premier temps de passage pour une vaste classe de processus stochastiques (non nécessairement markoviens) : ils considèrent un processus stochastique  $(Y_t)_{t \geq 0}$  sur la droite réelle, adapté à une filtration complète et continue à droite, et vérifiant les jeux d'hypothèses suivantes ((E1-3) pour l'existence et (U1-3) pour l'unicité) :

(E1) pour tout  $t > 0$ , la loi de  $Y_t$  est diffuse,

(E2) la trajectoire de  $Y$  est presque sûrement continue à droite et *quasi-continue à gauche*<sup>45</sup>,

(E3) pour toute frontière  $b$ , on a presque sûrement

$$\tau_b = \tau'_b := \inf\{s > 0, Y_s \leq b_s\},$$

(U1)  $Y$  est un processus markovien sur un intervalle  $I$  de  $\mathbb{R}$  vérifiant  $\mathbb{Q}_x(Y_0 = x) = 1$  pour  $x \in I$ ,

(U2)  $Y$  est stochastiquement monotone,

(U3) pour tout temps aléatoire  $\xi > 0$ , il existe un intervalle  $J^\xi \subset (0, t^\xi)$ , où  $t^\xi$  est la borne supérieure du support de la loi de  $\xi$ , tel que pour toute frontière  $b$  à valeurs dans  $\bar{I}$  tel que  $\tau_b$  a la même loi que  $\xi$ , on a pour tout  $t$  dans  $J^\xi$

$$b_t = \inf(\text{supp}(\mathbb{Q}(X_t \in \cdot | \tau_b > t))).$$

**Théorème C** ([\[KS23\]](#), Théorèmes 2.3 et 2.8). *On a les résultats suivants.*

(i) *Si les hypothèses (E1-3) sont vérifiées, pour tout temps aléatoire  $\xi > 0$ , il existe une frontière  $b$  telle que  $\tau_b$  a la même loi que  $\xi$ .*

<sup>44</sup>Voir Section 2 de [\[CCCS06\]](#), ou bien [\[CIL92\]](#) pour plus d'informations sur ce type de solutions.

<sup>45</sup>On demande que pour toute suite croissante de temps d'arrêts  $(T_n)_{n \geq 0}$  convergeant vers un temps aléatoire  $T$ , on a presque sûrement sur  $\{T < +\infty\}$

$$X_T = \lim_{n \rightarrow +\infty} X_{T_n}.$$



(ii) Si, à  $\xi$  donné, pour une loi de départ  $\mu_0$  donnée sur  $I$ , les hypothèses (E1-3, U1-3) sont vérifiées pour la loi  $\mathbb{Q} = \mathbb{Q}_{\mu_0}$  et un certain intervalle  $J^\xi$ , toutes les frontières données par (i) coïncident sur l'intervalle  $J^\xi$ .

En particulier, Klump et Savov détaillent des conditions propres aux processus de Lévy et aux diffusions qui permettent de s'assurer que les hypothèses sont vérifiées. Celles-ci sont consignées dans les Théorèmes 2.13 et 2.16 de [KS23], auxquels nous renvoyons le lecteur. On obtient ainsi une classe de processus auxquels s'appliquent le Théorème 1, puisqu'ils vérifient donc l'hypothèse (H1) en plus des hypothèses (H2) et (H3) (voir Section 1.2).

Pour montrer (i), Klump et Savov construisent une suite de frontières  $(b^n)_{n \geq 1}$  satisfaisant

$$\mathbb{Q}_{\mu_0}(X_{t_1^n} \geq b_{t_1^n}^n, \dots, X_{t_{k_n}^n} \geq b_{t_{k_n}^n}^n) = \mathbb{Q}(\xi > t_{k_n}^n)$$

pour des grilles temporelles  $((t_i^n)_{1 \leq i \leq k_n})_{n \geq 1}$  de plus en plus fines, puis utilisent les propriétés de la  $\Gamma$ -convergence (voir [DM93]) pour construire une limite à cette suite de frontières et montrer qu'elle résout le problème inverse du premier temps de passage. La preuve de l'unicité (ii) repose sur des comparaisons de mesures obtenues en appliquant successivement la propriété de Markov aux instants  $t_{k_n}^n$ , d'où l'importance de l'hypothèse (U1).

On peut noter que pour un processus markovien, l'hypothèse (E2) est automatiquement vérifiée<sup>46</sup> dès que celui-ci possède la propriété de Feller (hypothèse (H2)).

L'hypothèse (U2) permet d'obtenir des principes de comparaison sur les solutions au problème inverse du premier temps de passage<sup>47</sup>. Nous utilisons cette même hypothèse (hypothèse (H3)) dans la preuve des Théorèmes 1 et 2 pour construire le couplage du  $N$ -BMP avec ses barrières stochastiques<sup>48</sup>. On peut noter que notre résultat redémontre l'unicité de la solution au problème inverse du premier temps de passage dans le cas où celle-ci existe, pour un processus qui vérifie les hypothèses (H1-3). En effet, considérons deux frontières  $b^1$  et  $b^2$  vérifiant (H1), pour un processus  $(X_t)_{t \geq 0}$  vérifiant par ailleurs (H2) et (H3). Alors pour  $t \geq 0$  et  $r$  réel, on a

$$\mathbb{Q}_{\mu_0}(X_t \leq r | \tau_{b^1} > t) = \mathbb{Q}_{\mu_0}(X_t \leq r | \tau_{b^2} > t),$$

puisque ces deux quantités sont d'après le Théorème 1 la limite (unique) de  $F^N(t, r)$ . Ainsi, ces deux lois conditionnelles ont le même support; avec l'hypothèse (U3), on a donc  $b_t^1 = b_t^2$  pour tout  $t$  dans  $J^\xi$ .

On peut noter que d'autres variantes du problème de temps de passage ont été étudiées dans la littérature mathématique ces dernières années, dans des versions qui permettent toujours de conjecturer un lien (au moins heuristique) avec la limite hydrodynamique de systèmes de particules :

- dans le problème du premier temps de passage randomisé [JKZ09], on fixe cette fois la frontière et l'on recherche une loi initiale permettant de réaliser une loi fixée pour le temps de passage à cette frontière,
- dans le problème inverse du premier temps de passage fixé à 0 [DP15], on autorise un changement de temps en plus du choix de la loi initiale, la frontière étant fixée à 0,
- une version du problème inverse du premier temps de passage avec *soft-killing* est étudiée dans [EEH14, EHW20, KK24] : cette fois les particules ne sont pas tuées instantanément au passage de la frontière, mais à taux constant (chaque durée  $dt$  passée sous la frontière donne une probabilité de l'ordre de  $\lambda dt$  d'être tué),

<sup>46</sup>Voir [RY99], Exercice III.2.33.

<sup>47</sup>Voir [KS23], Théorème 2.11.

<sup>48</sup>De fait, nous avons été amenés à utiliser cette hypothèse avant d'avoir eu connaissance des travaux de Klump et Savov [KS23].

- le problème dans le cas particulier du mouvement brownien réfléchi en 0 (*two-sided inverse first-passage time problem*) est aussi étudiée par Klump et Kolb [KK23], rejoignant la formulation initiale d'Anulova [Anu80]; cette formulation se réalise comme la limite hydrodynamique en dimension 1 du modèle symétrique des abeilles browniennes [BBNP22], mais également comme celle d'un modèle sans branchement mais avec sélection présenté par Klump [Klu23]; dans l'optique d'effectuer des simulations numériques pour déterminer la forme des frontières ainsi que la loi conditionnée au non-dépassement de ces deux frontières, ce modèle se révèle *a priori* moins coûteux que le  $N$ -BMP en temps de calcul.

**Régularité de la frontière.** La question de la régularité de la frontière obtenue comme solution du problème inverse du premier temps de passage est également importante. En effet, si la semi-continuité (obtenue dans [CCCS06] pour les diffusions et [KS23] dans le cas général) est suffisante pour démontrer avec notre méthode que le  $N$ -BMP admet une limite hydrodynamique, la continuité est requise dans notre preuve du fait que la position de la particule la plus basse converge vers cette frontière. Le premier cas où la continuité de la frontière est démontrée remonte à [CCCS11]<sup>49</sup>, dans le cas du mouvement brownien standard issu d'une masse de Dirac en 0, avec une loi pour le temps d'atteinte ayant une densité strictement positive sur  $(0, +\infty)$ ; un résultat similaire est obtenu par une autre méthode par Ekström et Janson [EJ16]<sup>50</sup>.

Une avancée importante sur cette question est due à Chen, Chadam et Saunders [CCS22] : ils étudient la régularité des solutions de viscosité exhibées dans [CCCS11], et parviennent au résultat suivant.

**Théorème D** ([CCCS06], Théorèmes 1&2, [CCCS11], Théorème 1 & [CCS22], Théorème 1). *Soit  $\xi > 0$  un temps aléatoire dont la fonction de survie  $R$  continue sur  $\mathbb{R}_+$ , avec  $R > 0$  sur  $\mathbb{R}_+$ , et  $(X_t)_{t \geq 0}$  la solution de l'é.d.s.*

$$dX_t = \beta(t, X_t) dt + \sigma(t, X_t) dB_t,$$

où  $\beta$  et  $\sigma$  sont régulières, bornées, avec  $\inf \sigma > 0$ , et  $(B_t)_{t \geq 0}$  est un mouvement brownien standard.

- (i) Le problème à frontière libre [13] associé à la diffusion  $X$  admet une unique solution de viscosité  $(u, b)$ .
- (ii) La frontière  $b$  est l'unique solution semi-continue supérieurement du problème inverse du premier temps de passage pour  $X$ .
- (iii) (a) Si  $R$  est de classe  $C^1$  sur  $\mathbb{R}_+$  avec  $R' < 0$ , et si  $R$  est deux fois dérivable, de dérivée seconde intégrable sur un intervalle  $(0, \delta)$ , avec  $\delta > 0$ , alors  $(u, b)$  est une solution classique<sup>51</sup> du problème à frontière libre, et  $b$  est  $\frac{1}{2}$ -Hölder sur  $(0, +\infty)$ .
- (b) Si de plus  $R$  est de classe  $C^{\alpha+\frac{1}{2}}$  sur  $(t_0, T)$  pour un réel  $\alpha > \frac{1}{2}$  non entier,  $b$  est de classe  $C^\alpha$  sur  $(t_0, T)$ .
- (c) Enfin, si pour un réel  $\alpha \geq \frac{1}{2}$  non entier,  $R$  est de classe  $C^{\alpha+\frac{1}{2}}$  sur  $\mathbb{R}_+$ , si  $u_0$  est nulle sur  $\mathbb{R}_-$  et de classe  $C^{2\alpha}$  sur  $[0, \delta]$ , et si  $u_0$  et  $R$  satisfont des conditions de compatibilité en 0 données par les dérivations successives en  $b_t$  de la condition de frontière libre, alors  $b_0 = 0$  et  $b$  est de classe  $C^\alpha$  sur  $\mathbb{R}_+$ .

<sup>49</sup>Voir [CCCS11] Proposition 6.

<sup>50</sup>Voir le paragraphe suivant sur le lien avec le problème d'arrêt optimal.

<sup>51</sup>Au sens où, comme pour le problème [2],  $u$  est de classe  $C^{1,2}$  et  $b$  est continue.

Ils démontrent donc que sous des hypothèses de régularité relativement faible, les solutions de viscosité du problème à frontière libre sont en fait des solutions classiques, et que plus la fonction de survie prescrite dans le problème inverse du premier temps de passage est régulière, plus la frontière l'est. En particulier, dans le cas qui nous intéresse, celui de la loi exponentielle, la fonction de survie est de classe  $C^\infty$  et admet une dérivée strictement négative<sup>52</sup>, elle satisfait donc toutes les hypothèses présentées dans le théorème précédent. On en déduit donc que la frontière est non seulement continue, mais aussi de classe  $C^\infty$ . Par la même occasion, on obtient aussi l'existence d'une solution classique au problème à frontière libre associé, ce qui généralise donc le résultat obtenu dans le cas brownien [BBP19].

**Lien avec la théorie de l'arrêt optimal.** Nous revenons pour finir sur les travaux d'Ekström et Janson [EJ16], puisqu'ils font le lien avec un autre problème étudié dans la littérature mathématique, celui de l'arrêt optimal. Plus précisément, ils démontrent une correspondance exacte, dans le cas du mouvement brownien standard, entre le problème inverse du premier temps de passage et le problème suivant : trouver la fonction  $v$  définie par

$$v(t, x) = \inf_{\tau} \mathbb{E} \left( R(t - \tau) \mathbf{1}_{\{\tau < t\}} + \mathbf{1}_{\{\tau = t, x + B_t \leq 0\}} \right) \quad (14)$$

la borne inférieure étant prise sur les temps d'arrêt du mouvement brownien  $(B_s)_{s \geq 0}$  prenant leurs valeurs dans l'intervalle  $[0, t]$ . Si l'on définit alors

$$b_t = \sup\{x \in \mathbb{R}, v(t, x) = R(t)\},$$

et

$$u(t, x) = \mathbb{P}(B_t \geq x, \tau_b > t),$$

alors  $b$  est solution du problème inverse du premier temps de passage (pour le brownien standard issu de 0) et  $u = v$ . Cette correspondance permet notamment aux auteurs d'étendre les résultats qui avaient été obtenus jusqu'alors aux cas de lois sans densité; ils retrouvent de plus la continuité de la frontière dans le cas de la loi exponentielle. La méthode utilisée pour s'attaquer au problème d'arrêt optimal est une méthode de discrétisation, idée que l'on retrouve notamment dans les preuves de Klump et Savov [KS23], ainsi que dans les encadrements utilisés dans le cadre du  $N$ -BBM et des abeilles browniennes [DMFPSL19a, DMFPSL19b, BBNP22]. Notons enfin qu'un problème d'arrêt optimal dans un cadre plus général est étudié dans [BEES18] : le but est trouver le temps d'arrêt  $\tau$  qui maximise  $\mathbb{E}(\psi(B_\tau, \tau))$  parmi tous les temps d'arrêt du mouvement brownien qui suivent une loi donnée pour le mouvement brownien,  $\psi$  étant une fonction fixée *a priori*. Les auteurs démontrent que, pour certains choix de la fonction  $\psi$ , la solution s'exprime comme le premier temps de passage le long d'une certaine frontière, donnant par là même une solution au problème inverse du premier temps de passage.

## 2 Flot comonotone pour des processus de Feller sur la droite réelle

Dans la Section 1.3.2 nous avons mentionné que la preuve que nous développons pour démontrer le Théorème 1 s'appuie sur un couplage du  $N$ -BMP avec deux  $b$ -BMP qui l'encadrent (pour l'ordre stochastique, voir [8]). Ce couplage requiert de pouvoir créer des paires de trajectoires conservant le même ordre au cours du temps, chaque trajectoire suivant la loi du processus sous-jacent  $(X_t)_{t \geq 0}$ ; le processus ainsi obtenu sur  $\mathbb{R}^2$  doit en prime être un processus de Markov

<sup>52</sup>Si la fonction de survie est constante sur un intervalle, cela signifie qu'on ne peut pas toucher la frontière sur cet intervalle; celle-ci doit donc y valoir  $-\infty$ , et l'on n'a donc pas de régularité possible.

possédant la propriété de Feller. Les résultats du [Chapitre III](#), qui sont présentés dans cette section, établissent l'existence d'un tel processus à partir d'hypothèses minimales.

## 2.1 Hypothèses

Rappelons que  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, (\mathbb{Q}_x)_{x \in \mathbb{R}}, (X_t)_{t \geq 0})$  désigne le processus markovien sous-jacent qui dirige la trajectoire des particules, avec  $\mathbb{Q}_x(X_0 = x) = 1$  pour tout réel  $x$ . On rappelle également que le semi-groupe  $(P_t)_{t \geq 0}$  est la famille d'opérateurs linéaires continus sur  $\mathcal{B}_b$  définis par

$$\forall f \in \mathcal{B}_b \quad \forall x \in \mathbb{R} \quad P_t f(x) = \mathbb{E}_x(f(X_t)).$$

Nous souhaitons construire un processus de Markov  $((X_t^1, X_t^2))_{t \geq 0}$  sur un espace probabilisé  $(\Omega', \mathcal{F}', (\mathcal{F}'_t)_{t \geq 0}, (\mathbb{Q}'_{x^1, x^2})_{x^1, x^2 \in \mathbb{R}})$  ayant les propriétés suivantes :

- (a) la propriété de Feller, en tant que processus de Markov sur  $\mathbb{R}^2$ , est vérifiée,
- (b) sous  $\mathbb{Q}'_{x^1, x^2}$ ,  $(X_t^1)_{t \geq 0}$  a la loi de  $(X_t)_{t \geq 0}$  sous  $\mathbb{Q}_{x^1}$  et  $(X_t^2)_{t \geq 0}$  a la loi de  $(X_t)_{t \geq 0}$  sous  $\mathbb{Q}_{x^2}$ ,
- (c) si  $x^1 \leq x^2$ ,  $\mathbb{Q}'_{x^1, x^2}(X_t^1 \leq X_t^2, \forall t \geq 0) = 1$ .

Comme nous l'expliquons ci-après, les hypothèses (H2) et (H3) sont nécessaires pour espérer obtenir les propriétés (a), (b) et (c); le [Théorème 3](#) vient montrer, en prenant  $n = 2$ , que ces hypothèses sont également suffisantes.

Les propriétés (b) et (c) entraînent l'hypothèse (H3) selon laquelle le processus  $(X_t)_{t \geq 0}$  est stochastiquement monotone. En effet, à  $t$  fixé et pour  $x^1 \leq x^2$ , la propriété (c) implique que  $\mathbb{Q}'_{x^1, x^2}(X_t^1 \leq X_t^2) = 1$ . On en déduit par la propriété (b) que la loi de  $X_t$  sous  $\mathbb{Q}_{x^1}$  est stochastiquement dominée par la loi de  $X_t$  sous  $\mathbb{Q}_{x^2}$ , c.à.d.  $P_t(x^1, \cdot) \leq_{\text{st}} P_t(x^2, \cdot)$  avec les notations de la [Section II.2](#).

Inversement, à  $t$  fixé, partant de l'hypothèse que  $P_t(x^1, \cdot) \leq_{\text{st}} P_t(x^2, \cdot)$ , un résultat classique de Strassen [\[Str65\]](#) entraîne l'existence d'un couplage  $(Y_t^1, Y_t^2)$  de ces deux lois, tel que  $Y_t^1 \leq Y_t^2$  presque sûrement. La difficulté est ici de réaliser un couplage entre trajectoires vérifiant la même propriété simultanément pour tous les instants  $t \geq 0$ , et possédant qui plus est les propriétés de Markov et de Feller.

D'autre part, la propriété (a) entraîne que le processus  $(X_t)_{t \geq 0}$  doit aussi avoir la propriété de Feller (hypothèse (H2)).

## 2.2 Résultats

Plutôt que de nous restreindre à des paires de particules, nous montrons en fait, pour tout entier  $n \geq 2$ , l'existence d'un  $n$ -uplet de trajectoires ayant la propriété de Feller, chaque trajectoire ayant la loi de  $(X_t)_{t \geq 0}$ , et l'ordre entre les trajectoires étant préservé presque sûrement au cours du temps. De plus, cette construction présente une propriété de cohérence, à savoir que pour  $k \leq n$ , la loi obtenue pour  $k$  trajectoires s'obtient en marginalisant la loi obtenue pour  $n$  trajectoires.

Etant donné un semi-groupe  $(P_t^{(n)})_{t \geq 0}$  sur  $\mathbb{R}^n$ , on note  $(P_t^{(n)}(\mathbf{x}, \cdot))_{\mathbf{x} \in \mathbb{R}^n}$  le noyau markovien associé à  $P_t^{(n)}$ . On dit alors que  $(P_t^{(n)})_{t \geq 0}$  préserve l'ordre entre les coordonnées si, pour tout  $\mathbf{x} \in \mathbb{R}^n$  et  $t \geq 0$ ,  $P_t^{(n)}(\mathbf{x}, \cdot)$  est à support dans l'ensemble  $\mathbb{R}_{\mathbf{x}}^n$  des éléments  $\mathbf{y} \in \mathbb{R}^n$  compatibles avec  $\mathbf{x}$ , c'est-à-dire vérifiant pour tous  $1 \leq i, j \leq n$

$$x^i \leq x^j \implies y^i \leq y^j.$$

De plus, si l'on dispose d'un semi-groupe  $(P_t^{(n)})_{t \geq 0}$  sur  $\mathbb{R}^n$  pour chaque entier  $n \geq 1$ , on dit que la famille de semi-groupes est *cohérente* si, pour tout  $k \leq n$ , tout  $\mathbf{x}$  dans  $\mathbb{R}^n$  et tout jeu d'indices  $1 \leq i_1 < \dots < i_k \leq n$ , la loi marginale de  $P_t^{(n)}(\mathbf{x}, \cdot)$  associée aux coordonnées d'indices  $(i_1, \dots, i_k)$  est  $P_t^{(k)}((x^{i_1}, \dots, x^{i_k}), \cdot)$ .

Nous avons alors démontré le résultat suivant, qui assure l'existence d'un couplage avec les propriétés recherchées, non pas d'une paire, mais de n'importe quel nombre fini de particules.

**Théorème 3.** *Si  $(P_t)_{t \geq 0}$  est un semi-groupe de Feller stochastiquement monotone sur  $\mathbb{R}$ , il existe une famille cohérente  $((P_t^{(n)})_{t \geq 0})_{n \geq 1}$  de semi-groupes de Feller préservant l'ordre entre les coordonnées telle que  $(P_t^{(1)})_{t \geq 0} = (P_t)_{t \geq 0}$ .*

Pour démontrer le [Théorème 3](#), nous construisons tout d'abord pour chaque entier  $n \geq 2$  un processus sur  $\mathbb{R}^n$  à des instants discrets régulièrement espacés, obtenu en utilisant de manière successive le *couplage comonotone*, qui consiste à appliquer à une même variable aléatoire de loi uniforme sur  $(0, 1)$  la pseudo-inverse des fonctions de répartition respectives des lois concernées. Ensuite, nous faisons tendre le pas de temps vers 0, et montrons l'existence d'un noyau markovien limite sur  $\mathbb{R}^n$  ayant les propriétés requises.

Nous avons également pu caractériser la famille de semi-groupes ainsi obtenue grâce à l'ordre super-modulaire sur les mesures de probabilité sur  $\mathbb{R}^n$ . Pour  $n \geq 2$ , on dit qu'une fonction  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  est *super-modulaire*<sup>53</sup> lorsque nous avons  $f(\mathbf{x}) + f(\mathbf{y}) \leq f(\mathbf{x} \vee \mathbf{y}) + f(\mathbf{x} \wedge \mathbf{y})$  pour tous  $\mathbf{x}, \mathbf{y}$  dans  $\mathbb{R}^n$ , en notant  $\mathbf{x} \vee \mathbf{y} = (\max(x^i, y^i))_{1 \leq i \leq n}$  et  $\mathbf{x} \wedge \mathbf{y} = (\min(x^i, y^i))_{1 \leq i \leq n}$ . On définit alors l'ordre super-modulaire sur l'ensemble des mesures de probabilité sur  $\mathbb{R}^n$  en notant  $\mu \leq_{\text{sm}} \nu$  si, pour toute fonction  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  super-modulaire mesurable bornée, on a

$$\int_{\mathbb{R}^n} f(\mathbf{x}) \mu(d\mathbf{x}) \leq \int_{\mathbb{R}^n} f(\mathbf{x}) \nu(d\mathbf{x}).$$

Pour une famille  $(\mu_i)_{1 \leq i \leq n}$  de lois sur  $\mathbb{R}$  donnée, le couplage comonotone entre ces lois maximise l'ordre super-modulaire parmi toutes les lois sur  $\mathbb{R}^n$  ayant les lois  $\mu_i$  comme marginales<sup>54</sup>.

En exploitant cette propriété, nous montrons que le processus limite construit dans la preuve du [Théorème 3](#) maximise l'ordre super-modulaire parmi tous les processus de Feller sur  $\mathbb{R}^n$  dont les lois marginales unidimensionnelles suivent la loi de  $(X_t)_{t \geq 0}$ . Nous avons alors choisi d'appeler *flot comonotone* associé au processus  $(X_t)_{t \geq 0}$  la<sup>55</sup> famille  $((\check{P}_t^{(n)})_{t \geq 0})_{n \geq 1}$  de semi-groupes construite dans la preuve du [Théorème 3](#) (même si nous n'avons pas construit explicitement de flot stochastique mais seulement une famille cohérente de semi-groupes; nous renvoyons à la [Section 2.3](#) pour une discussion sur la notion de flot stochastique).

Plus précisément, si l'on note  $\mathfrak{M}_n$  l'ensemble des semi-groupes de Feller  $(Q_t^{(n)})_{t \geq 0}$  sur  $\mathbb{R}^n$  tels que les lois marginales 1-dimensionnelles associées à chaque opérateur  $Q_t^{(n)}$  sont données par le semi-groupe  $(P_t)_{t \geq 0}$  du processus sous-jacent, nous avons démontré le résultat suivant.

**Théorème 4.** *Soit  $((\check{P}_t^{(n)})_{t \geq 0})_{n \geq 1}$  le flot comonotone associé à  $(X_t)_{t \geq 0}$ . Alors pour tous  $n \geq 1$ ,  $t \geq 0$  et  $\mathbf{x} \in \mathbb{R}^n$ , la loi  $\check{P}_t^{(n)}(\mathbf{x}, \cdot)$  est maximale pour l'ordre super-modulaire, parmi l'ensemble  $\{M_t^{(n)}(\mathbf{x}, \cdot), (M_s^{(n)})_{s \geq 0} \in \mathfrak{M}_n\}$ .*

<sup>53</sup>Voir [MS02](#).

<sup>54</sup>Voir [MS02](#).

<sup>55</sup>Nous ne démontrons pas l'unicité de la famille de processus obtenue par passage à la limite dans la preuve du [Théorème 3](#), mais celle-ci découle de la propriété de maximalité pour l'ordre super-modulaire ([Théorème 4](#)).

## 2.3 Discussion

Pour remplacer le [Théorème 3](#) dans un contexte plus général, on peut noter qu'il répond dans le cadre des processus de Feller au problème présenté dans [FM01](#) sur l'équivalence, pour une famille de variables aléatoires donnée, entre la propriété de monotonie stochastique, et la réalisabilité de celle-ci *via* un couplage. Cette équivalence a été démontrée dans de nombreux cas incluant les chaînes de Markov en temps discret [KKO77](#) et les processus de saut [Zha00](#) sur un espace polonais muni d'un ordre partiel, les chaînes de Markov en temps continu sur un espace dénombrable [LMS00](#) et certaines classes de systèmes de particules en interaction [Lig05](#).

Mentionnons deux cas particuliers du [Théorème 3](#) pour lesquels on peut construire un couplage plus simplement que par la méthode utilisée dans notre démonstration :

- les processus de Lévy : l'homogénéité en espace de ces processus permet de définir le couplage parallèle (voir [Figure II.1a](#)), c'est-à-dire de définir, pour  $1 \leq i \leq n$  et  $t \geq 0$ ,  $X_t^i = x^i + X_t$ , où  $(X_t)_{t \geq 0}$  est une version du processus vérifiant  $X_0 = 0$ ,
- les processus de Feller à trajectoires continues : on peut cette fois faire appel au couplage de Doeblin (voir [Figure II.1b](#)), dans lequel les trajectoires sont indépendantes les unes des autres jusqu'au premier instant où elles se rencontrent; à partir de cet instant, on leur fait suivre la même trajectoire.

Revenons pour finir sur la notion de *flot stochastique* évoquée à la [Section 2.2](#), qui est étudiée en détail par Le Jan et Raimond [LJR04](#). Cet objet est l'analogue du flot d'une équation différentielle pour les processus markoviens : c'est une famille de fonctions aléatoires <sup>56</sup> $(\varphi_{s,t})_{0 \leq s \leq t}$  de  $\mathbb{R}$  dans  $\mathbb{R}$  telles que pour tout  $x$  dans  $\mathbb{R}$ ,  $(\varphi_{0,t}(x))_{t \geq 0}$  a la même loi que  $(X_t)_{t \geq 0}$  sous  $\mathbb{Q}_x$ . Le flot comonotone que nous avons construit n'est pas un flot stochastique à proprement parler, mais seulement une famille cohérente de semi-groupes; le Théorème 1 de [LJR04](#) permet d'en déduire l'existence d'un flot stochastique associé à cette famille de semi-groupes.

## 3 Problème de transport optimal en champ moyen

Dans le [Chapitre III](#), nous présentons des travaux effectués dans le cadre du CEMRACS 2022 qui ne possèdent pas de lien direct avec les problèmes discutés jusqu'ici, et sont consacrés à la résolution numérique d'un problème de transport optimal en champ moyen. Ce problème consiste à contrôler la dérive d'un processus de type McKean-Vlasov et avec une loi fixée à l'instant initial 0 et à l'instant final  $T > 0$ , de façon à minimiser en moyenne un certain coût fonction de la trajectoire du processus. Nous avons mis en œuvre des méthodes d'apprentissage automatique, dans lesquelles des réseaux de neurones sont utilisés pour modéliser le problème sous plusieurs formes, puis entraînés pour en donner une solution approchée.

### 3.1 Contrôle et transport optimal en champ moyen

Les jeux à champ moyen <sup>57</sup> sont une classe de problèmes dans lesquels un grand nombre de joueurs sont opposés, chacun cherchant à minimiser son coût de déplacement le long d'une trajectoire. La trajectoire et le coût de déplacement dépendent des interactions entre joueurs, et sont simplifiées par l'intervention du *champ moyen* : plutôt que de considérer des interactions individuelles, on se place du point de vue de l'un des joueurs qui interagit avec la distribution

---

<sup>56</sup>Nous restons volontairement vague sur la définition précise d'un flot stochastique et renvoyons aux travaux de Le Jan et Raimond [LJR04](#) pour une définition rigoureuse.

<sup>57</sup>Voir [LL06a](#), [LL06b](#), [LL07](#), [HMC06](#), [HCM07](#)

totale de joueurs. Le problème de contrôle en champ moyen<sup>58</sup> est une variante des jeux à champ moyen dans laquelle on cherche cette fois à minimiser un coût moyen pour tous les joueurs (c.à.d. que cette fois, les joueurs coopèrent), avec le même type d'interactions.

On considère des agents se déplaçant dans l'espace  $\mathcal{Q} = \mathbb{R}^d$ , avec un horizon de temps fixé  $T > 0$ . On note  $\mathcal{Q}_T = [0, T] \times \mathcal{Q}$ . Pour modéliser le déplacement des agents au cours du temps en tenant compte des interactions avec le champ moyen, on suppose donc que ceux-ci suivent une dynamique de type McKean-Vlasov<sup>59</sup>, c.à.d. que leur trajectoire est solution de l'é.d.s.

$$dX_t^v = \beta(X_t^v, \mu_t^v, v(t, X_t^v)) dt + \sigma dB_t, \quad (15)$$

faisant intervenir les objets suivants :

- $X_t^v$  désigne la position de l'agent à l'instant  $t \in [0, T]$ , et  $\mu_t^v$  est un élément de l'espace  $\mathcal{P}_2(\mathcal{Q})$  des lois sur  $\mathcal{Q}$  admettant un moment d'ordre 2, qui désigne la loi de  $X_t^v$ ,
- $v : \mathcal{Q}_T \rightarrow \mathbb{R}^k$  est le terme de *contrôle*,
- $\beta : \mathcal{Q} \times \mathcal{P}_2(\mathcal{Q}) \times \mathbb{R}^k \rightarrow \mathcal{Q}$  et la constante  $\sigma > 0$  sont respectivement la dérive et la volatilité de l'é.d.s.,
- $(B_t)_{t \geq 0}$  est un mouvement brownien standard en dimension  $d$ .

On considère alors un coût de déplacement infinitésimal  $f : \mathcal{Q} \times \mathcal{P}_2(\mathcal{Q}) \times \mathbb{R}^k \rightarrow \mathbb{R}$  et un coût terminal  $g : \mathcal{Q} \times \mathcal{P}_2(\mathcal{Q}) \rightarrow \mathbb{R}$ . Le *problème de contrôle en champ moyen* consiste alors à trouver un contrôle  $v^* : \mathcal{Q}_T \rightarrow \mathbb{R}^k$  minimisant, en moyenne, le coût de déplacement et le coût terminal, c.à.d. minimisant la fonctionnelle

$$J^{MFC} : v \mapsto \mathbb{E} \left( \int_0^T f(X_t^v, \mu_t^v, v(t, X_t^v)) dt + g(X_T^v, \mu_T^v) \right), \quad (16)$$

sous la contrainte donnée par une loi de départ  $\rho_0 \in \mathcal{P}_2(\mathcal{Q})$  pour l'é.d.s..

Le problème considéré dans le [Chapitre III](#) est une variante de ce problème appelé *problème de transport optimal en champ moyen*. Cette fois, nous imposons, en plus de la loi initiale, une répartition terminale pour les agents, c.à.d. que l'on impose sur les solutions de l'é.d.s. [\(15\)](#) une loi de départ  $\rho_0$  et une loi finale  $\rho_T$ , qui sont deux éléments de  $\mathcal{P}_2(\mathcal{Q})$  fixés. Cette deuxième contrainte remplace le coût terminal, et le problème est donc de trouver un contrôle  $v^*$  minimisant la fonctionnelle

$$J^{MFO} : v \mapsto \mathbb{E} \left( \int_0^T f(X_t^v, \mu_t^v, v(t, X_t^v)) dt \right), \quad (17)$$

sous la contrainte

$$\begin{cases} dX_t^v = \beta(X_t^v, \mu_t^v, v(t, X_t^v)) dt + \sigma dB_t, \\ X_0^v \sim \rho_0, X_T^v \sim \rho_T. \end{cases} \quad (18)$$

Notons que, contrairement au cas du problème de contrôle en champ moyen, la condition de loi terminale impose une restriction *a priori* sur les contrôles admissibles, puisque l'on est contraint de choisir parmi les contrôles qui permettent de transporter *via* l'é.d.s. [\(15\)](#) la loi  $\rho_0$  sur la loi  $\rho_T$ .

L'étude théorique du problème de transport optimal en champ moyen est, à notre connaissance, toujours un problème ouvert<sup>60</sup> que nous n'avons pas cherché à étudier plus avant. Les

<sup>58</sup>Voir [BFY13](#) pour une introduction détaillée à ce sujet.

<sup>59</sup>Voir [McK66](#), [Szn91](#).

<sup>60</sup>Nous renvoyons à [BFY13](#), [AL15](#) pour une étude sur le problème de contrôle en champ moyen. Une solution est donnée dans [HT23](#), mais les hypothèses faites sont restrictives et ne permettent pas de prendre en compte le modèle de congestion que nous avons étudié, par exemple.

travaux présentés dans le [Chapitre III](#) sont exclusivement consacrés à la mise en œuvre de méthodes numériques et à la comparaison des résultats qu’elles produisent sur quelques exemples concrets.

## 3.2 Description des méthodes numériques

Dans cette section, nous présentons brièvement les outils relatifs aux méthodes d’apprentissage profond que nous avons utilisés, ainsi que les trois méthodes numériques que nous avons proposées. Nous renvoyons au [Chapitre III](#) pour de plus amples détails.

Les trois méthodes que nous proposons sont basées sur l’utilisation des réseaux de neurones, qui sont des fonctions dépendant d’un paramètre  $\theta$  vivant dans un espace de grande dimension. L’intérêt des réseaux de neurones, obtenus comme compositions successives d’applications linéaires<sup>61</sup> et de fonctions d’*activation*, est de ne pas avoir à considérer une grille spatio-temporelle fixée sur laquelle on chercherait à résoudre le problème, et dont la taille augmenterait exponentiellement vite avec la dimension. A la place, l’introduction du paramètre  $\theta$  permet, à chaque étape de l’entraînement, d’évaluer le réseau de neurone en n’importe quel point de l’espace, et l’on cherche alors à lui faire vérifier les conditions du problème sur un certain nombre de points simulés, plutôt que sur tous les points de la grille.

Pour chacune des méthodes, nous définissons donc une fonctionnelle  $\mathcal{L}(\theta)$  dépendant de ce paramètre  $\theta$ , et c’est cette fonctionnelle, ou une approximation numérique de celle-ci, que nous cherchons à minimiser. En pratique, nous utilisons pour cela une méthode de *descente de gradient stochastique*, en réalisant  $\mathcal{L}(\theta)$  comme l’espérance de  $\mathcal{L}(\theta; S)$ , où  $S$  est une variable choisie convenablement, décrivant les points sur lesquels nous cherchons à optimiser le réseau de neurones. Nous approchons alors numériquement  $\mathbb{E}(\mathcal{L}(\theta; S))$  par la méthode de Monte-Carlo, c.à.d. que l’on considère une famille  $(S_n)_{1 \leq n \leq N}$  de réalisations de  $S$  et que nous calculons

$$\mathcal{L}(\theta; S_1, \dots, S_N) = \frac{1}{N} \sum_{n=1}^N \mathcal{L}(\theta, S_n).$$

Grâce à la structure offerte par le réseau de neurones, nous pouvons appliquer un algorithme de dérivation automatique<sup>62</sup> pour calculer, à  $S_1, \dots, S_N$  fixés, le gradient de  $\mathcal{L}(\theta; S_1, \dots, S_N)$  en fonction de  $\theta$ . Comme pour l’algorithme classique de descente de gradient, on actualise alors la valeur du paramètre  $\theta$  en le déplaçant dans le sens opposé au gradient, puis l’on répète cette opération. On observe alors au fur et à mesure de l’entraînement une baisse de la valeur de  $\mathcal{L}(\theta; S_1, \dots, S_N)$ .

### 3.2.1 Méthode directe

La première méthode que nous avons implémentée consiste à paramétrer directement le contrôle, c.à.d. que l’on remplace la fonction  $v$  par une fonction paramétrée  $v_\theta$  dans [\(17\)](#) et [\(18\)](#). On réalise alors différentes approximations pour calculer la quantité  $J^{MFOt}(v_\theta)$ , que l’on cherche à optimiser en fonction de  $\theta$ . Nous simulons une approximation de l’é.d.s. [\(15\)](#) à l’aide de la méthode d’Euler-Maruyama, mais nous n’avons pas de moyen d’assurer la condition terminale : pour cela, nous ajoutons un terme de pénalisation, revenant donc à une formulation en termes de problème de contrôle en champ moyen, avec la fonction  $g$  donnée par

$$\forall x \in \mathcal{Q} \quad \forall \mu \in \mathcal{P}_2(\mathcal{Q}) \quad g(x, \mu) = C\mathcal{W}_2(\mu, \rho_T),$$

<sup>61</sup>Le paramètre  $\theta$  est en fait l’ensemble des coefficients des matrices formées par ces applications linéaires.

<sup>62</sup>Comme le réseau de neurones est obtenu comme composition d’applications dont la dérivation ne pose pas de problème, nous pouvons appliquer la règle de dérivation en chaîne.



où  $W_2$  désigne la distance de Wasserstein sur  $\mathcal{P}_2(\mathcal{Q})$  et  $C$  une constante positive. Cette pénalisation ne permet pas d'imposer la condition terminale, mais permet de chercher un compromis entre l'optimisation du coût de déplacement et la bonne approximation de la distribution à l'instant  $T$ .

Les autres approximations sont d'ordre numérique : nous approchons l'intégrale et l'espérance dans (16) par la méthode de Monte-Carlo, ce qui permet d'appliquer l'algorithme de descente de gradient stochastique pour améliorer la valeur de  $\theta$  et donc celle du contrôle  $v_\theta$ .

### 3.2.2 Résolution d'une é.d.p. couplée

Notre deuxième algorithme consiste à résoudre par la méthode "Deep Galerkin" [SS18] un couple d'é.d.p. équivalent au problème de minimisation de  $J^{MFO T}$ , dont le principe est le suivant : supposons que l'on souhaite approcher la solution de l'équation de la chaleur sur un domaine  $\mathcal{U}$  borné à bord régulier, avec condition au bord donnée par une fonction  $h : \partial\mathcal{U} \rightarrow \mathbb{R}$  et condition initiale  $f_0 : \mathcal{U} \rightarrow \mathbb{R}$ . Alors on peut introduire la fonctionnelle

$$J(f) = \left\| \partial_t f - \frac{1}{2} \Delta f \right\|_{L^2([0,T] \times \mathcal{U})}^2 + \|f - h\|_{L^2([0,T] \times \partial\mathcal{U})}^2 + \|f - f_0\|_{L^2(\mathcal{U})}^2,$$

de sorte que  $f$  est solution du problème si et seulement si  $J(f)$  est égale à 0. On remplace alors la fonction  $f$  par un réseau de neurones  $f_\theta$ , et l'on approche numériquement les normes sur les espaces  $L^2$  par la méthode de Monte-Carlo, ce qui permet à nouveau d'appliquer l'algorithme de descente de gradient stochastique pour minimiser une approximation de  $J(f_\theta)$ , et donc donner une solution approchée de l'équation de la chaleur.

Notre méthode consiste donc à déterminer un couple d'é.d.p. équivalent au problème de transport optimal en champ moyen, et d'appliquer cet algorithme pour le résoudre numériquement (voir la Section III.3.2 pour les détails concernant ce système d'é.d.p., que nous choisissons d'omettre ici). Au moins formellement, étant donné un contrôle  $v$ , la loi  $\mu_t^v$  admet une densité solution de l'équation de Kolmogorov-Fokker-Planck. Notre problème devient alors un problème de contrôle optimal, et la théorie du contrôle permet alors de déterminer une seconde équation (dite de Hamilton-Jacobi-Bellman) couplée à la première comme condition d'optimalité pour le contrôle. Nous obtenons finalement le système d'équations (III.11), dont nous approchons numériquement la solution par l'algorithme Deep Galerkin.

### 3.2.3 Méthode de lagrangien augmenté pour une formulation duale

Pour cette dernière méthode, nous travaillons sur une troisième formulation du problème de transport optimal en champ moyen, qui est une formulation duale à la formulation en termes d'é.d.p. résolue par la deuxième méthode. Cette dualité permet d'obtenir un problème équivalent<sup>63</sup> au couple d'é.d.p., en termes de minimisation d'une somme de deux fonctionnelles convexes, ce qui permet d'appliquer des algorithmes classiques d'optimisation convexe. En l'occurrence, nous reprenons l'algorithme ADMM (Alternating Direction Method of Multipliers, voir [BPC<sup>+</sup>11]) qui consiste à minimiser alternativement les deux fonctionnelles, que l'on couple au principe de la méthode Deep Galerkin, c.à.d. que l'on remplace les arguments des deux fonctionnelles ainsi que les multiplicateurs de Lagrange du problème par des réseaux de neurones, et que l'on applique un algorithme de descente de gradient stochastique pour entraîner alternativement les paramètres de chaque réseau de neurones. Cette méthode est décrite en détail dans la Section III.3.3.

<sup>63</sup>Ces deux problèmes sont en dualité, et c'est le Théorème de Fenchel-Rockafellar (voir [Roc70], Théorème 31.1) qui nous incite à penser que les deux problèmes sont équivalents.

## 4 Questions en suspens et perspectives

### 4.1 Limite hydrodynamique de systèmes de particules

**Liens entre les différents problèmes.** Les problèmes présentés dans la [Section 1](#) de cette introduction tournent majoritairement autour de trois questions, dépendant du processus de Feller  $(X_t)_{t \geq 0}$  de générateur  $A$ , et d'une loi de départ à densité  $u_0$ , que nous rappelons ici :

- (i) **limite hydrodynamique du  $N$ -BMP** : existence d'une fonction  $F : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$  telle que la fonction de répartition empirique du  $N$ -BMP

$$F^N(t, r) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{X_t^i \leq r\}}$$

converge vers  $F$ ,

- (ii) **problème à frontière libre** : existence d'une fonction  $u : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$  et une frontière  $b : (0, +\infty) \rightarrow \mathbb{R}$  telle que

$$\begin{cases} \partial_t u = A^* u + u, & \text{pour } x > b_t, \\ u(t, x) = 0, & \text{pour } x \leq b_t, \\ \int_{b_t}^{+\infty} u(t, x) dx = 1, \\ u(0, x) = u_0(x), \end{cases}$$

- (iii) **problème inverse du premier temps de passage** : existence d'une frontière  $b$  telle que le temps d'arrêt

$$\tau_b = \inf\{s > 0, X_s < b_s\}$$

soit de loi exponentielle  $\mathcal{E}(1)$ .

Les résultats les plus significatifs évoqués jusque-là concernant les liens entre ces trois problèmes, ainsi que le [Théorème 1](#) démontré dans le [Chapitre 1](#) basé sur [BF23a](#), sont résumés dans la [Figure 3](#) : on retrouve en gris clair les résultats concernant le cas du mouvement brownien; en gris foncé, celui des diffusions régulières; et en noir, les cas plus généraux dont les hypothèses ont été discutées dans la [Section 1.4.2](#). En particulier, le Théorème 1 de [CCCS11](#), le Théorème 1 de [CCS22](#) puis le Théorème 1 de [BF23a](#) ([Théorème 1](#)) assurent ensemble l'existence de solutions aux trois problèmes dans le cas où  $(X_t)_{t \geq 0}$  est une diffusion à coefficients réguliers et bornés.

Une piste de recherche naturelle serait de poursuivre l'étude des équivalences entre ces problèmes dans un cadre général. En particulier, le problème à frontière libre (ii) n'est étudié, à notre connaissance, que dans le cas des diffusions à coefficients réguliers; or on sait que le problème (iii) admet des solutions pour une classe générale de processus, et que cela fournit une limite hydrodynamique (problème (i)). On aimerait donc savoir si, comme dans les cas concernés par [CCCS11](#), la frontière qui résout le problème (iii) donne également une solution du problème à frontière libre (ii) correspondant dans un cadre plus général. De même, on pourrait chercher à étudier ces liens pour des modèles faisant intervenir d'autres règles de sélection, comme ceux mentionnés dans la [Section 1.4.1](#) ([CDMGPI6](#), [Ata23](#) par exemple). L'écueil principal est celui de la régularité des objets considérés : il semble par exemple difficile, partant simplement de l'existence d'une limite hydrodynamique, de prouver directement que celle-ci permet de définir une frontière régulière, qui permettrait d'en déduire l'existence de solutions fortes au problème (ii).

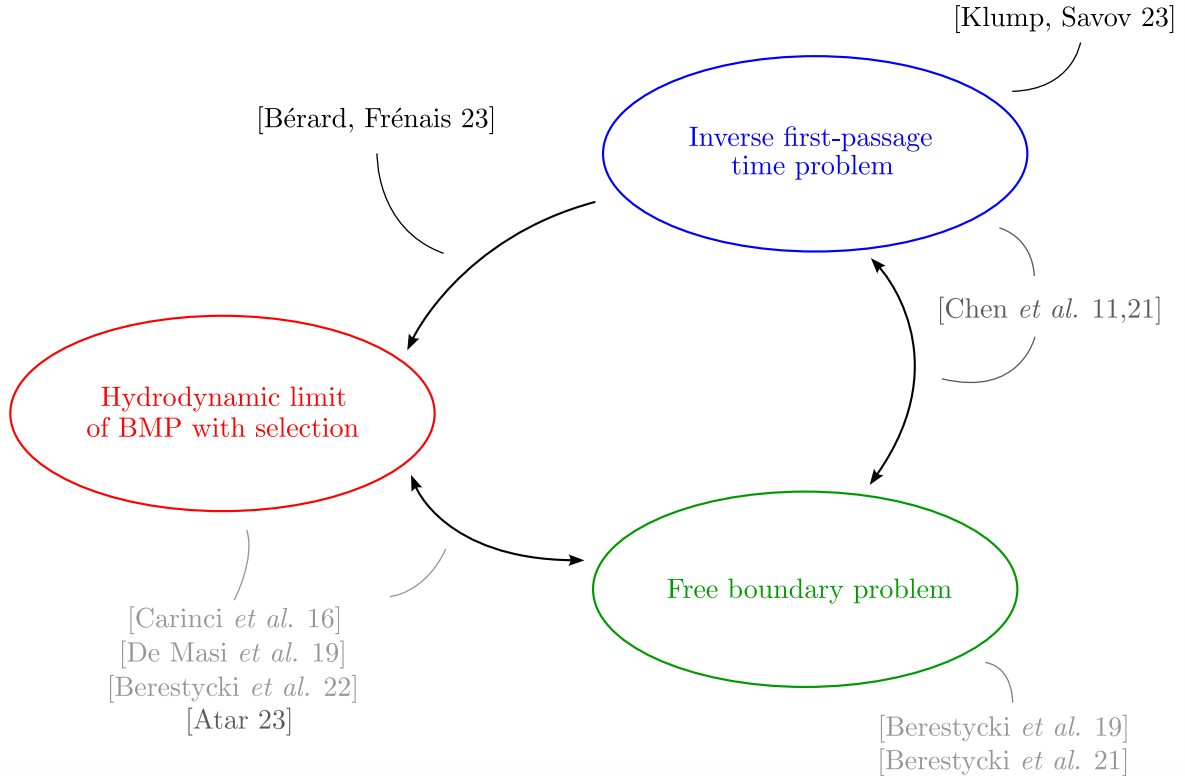


Figure 3: Limites hydrodynamiques, problèmes à frontière libre et problème inverse du premier temps de passage

Une autre question est justement celle de la régularité de la frontière en fonction de celle des trajectoires de  $(X_t)_{t \geq 0}$ . Pour démontrer le [Théorème 2](#) nous faisons par exemple deux hypothèses supplémentaires par rapport au [Théorème 1](#), à savoir la continuité des trajectoires de  $(X_t)_{t \geq 0}$  (hypothèse (H4)) et celle de la frontière  $b$  (hypothèse (H5)). Pour toute une classe de processus à trajectoires continues, à savoir les diffusions concernées par le [Théorème D](#), la frontière est effectivement continue; ce résultat se généralise-t-il à tous les processus à trajectoires continues pour lesquels le problème inverse du premier temps de passage admet une solution ?

Enfin, nous pourrions également prendre en compte le quatrième problème, celui d'arrêt optimal [\(14\)](#) introduit par Ekström et Janson [\[EJ16\]](#) : sa correspondance n'est établie qu'avec le problème (iii), et dans le cadre particulier du mouvement brownien standard; cette correspondance pourrait s'étendre à des processus plus généraux.

**Comportement en temps long du  $N$ -BMP.** Une autre étude intéressante serait celle du comportement en temps long du  $N$ -BMP et de sa limite hydrodynamique. Commençons par les questions autour du comportement en temps long de la limite hydrodynamique du  $N$ -BMP, c.à.d. que l'on passe d'abord à la limite  $N \rightarrow +\infty$ , puis l'on regarde le comportement de la limite  $F(\cdot, t)$  quand  $t$  tend vers l'infini. Rappelons que, dans le cadre du [Théorème 1](#), cette limite hydrodynamique est la loi du processus conditionné à rester au-dessus de la frontière  $b$ . On peut donc se demander quels sont les comportements limites, quand  $t$  tend vers l'infini, de  $F(\cdot, t)$  et de la frontière  $b_t$ .

*Lois quasi-stationnaires et quasi-limites.* Nous avons déjà vu que, si  $\nu$  est une loi quasi-stationnaire pour le processus  $(X_t)_{t \geq 0}$  tué en 0, le  $N$ -BMP partant de la loi initiale  $\nu^{\otimes N}$  admet pour limite hydrodynamique la loi  $\nu$ , le comportement asymptotique étant bien compris car la frontière et la loi ne dépendent pas du temps. Si le processus  $(X_t)_{t \geq 0}$  est de plus homogène en espace, on peut considérer la limite hydrodynamique pour le processus  $(X'_t)_{t \geq 0} = (X_t + \alpha t)_{t \geq 0}$ ; en partant de cette même loi quasi-stationnaire pour  $(X_t)_{t \geq 0}$ , nous obtenons une frontière qui est une droite de pente  $\alpha$  et une limite hydrodynamique qui est la translation de la loi quasi-stationnaire le long de cette droite. Dans le cas du  $N$ -BBM par exemple, le problème à frontière libre (2) qui caractérise la limite hydrodynamique admet des solutions sous la forme de propagation d'onde de vitesse  $\alpha$ , pour tout  $\alpha \geq \sqrt{2}$ . Pour les obtenir, on étudie les lois quasi-stationnaires du mouvement brownien avec dérive  $-\alpha$ , qui sont les lois à densité  $u_0^\alpha$  calculées plus haut (12). Ainsi, le problème à frontière libre (2) avec donnée initiale  $u_0^\alpha$  a pour solution la frontière linéaire de pente  $\alpha$  et la fonction  $u^\alpha$  définie par

$$\forall t \geq 0 \forall x \in \mathbb{R} \quad u^\alpha(t, x) = u_0^\alpha(x - \alpha t).$$

Cependant, ce comportement est restreint aux lois de départ qui sont quasi-stationnaires, alors que nous étudions la limite hydrodynamique partant d'une loi de départ quelconque. Nous aimerions savoir si ces solutions sont attractives, au sens où, partant d'une loi quelconque, la frontière  $b$  aurait une pente limite, et la limite hydrodynamique se rapprocherait, quand  $t$  tend vers l'infini, de la loi obtenue par translation d'une loi quasi-stationnaire associée à cette pente. En supposant que la solution du problème (2) converge effectivement vers un front d'onde, Berestycki, Brunet et Derrida [BBD18] obtiennent pour le  $N$ -BBM un développement asymptotique de la position de la frontière  $b_t$ . Mais à ce jour, la question de la convergence vers un front d'onde reste ouverte, même dans le cas brownien : Berestycki et Trough ont récemment conjecturé<sup>64</sup>, par analogie avec les solutions de l'équation de F-KPP<sup>65</sup>, que la solution  $(u, b)$  du problème (2) converge, après translation, vers la solution stationnaire  $u_0^{\sqrt{2}}$  de dérive minimale, si et seulement si la queue de distribution de la loi initiale présente une décroissance suffisamment rapide.

Notons enfin que dans le cas des abeilles browniennes [BBNP21, BBNP22]<sup>66</sup>, cette question est déjà résolue. En effet, le problème à frontière libre (5) qui caractérise la limite hydrodynamique admet une unique solution stationnaire  $(F_\infty, \mathbf{b}_\infty)$ , qui est l'unique fonction propre principale d'intégrale 1 du Laplacien sur un disque centré en 0, avec condition de Dirichlet homogène; le rayon du disque (qui caractérise ici la frontière libre) est choisi de manière à ce que la valeur propre principale du Laplacien sur ce disque soit égale à 1. Berestycki, Brunet, Nolen et Penington démontrent<sup>67</sup> que, quelle que soit la loi de départ, la limite hydrodynamique des abeilles browniennes converge uniformément vers cette solution stationnaire quand  $t$  tend vers l'infini.

*Limite hydrodynamique d'une loi stationnaire pour le  $N$ -BMP.* L'autre point de vue que l'on pourrait adopter serait de d'abord passer à la limite  $t \rightarrow +\infty$ , à  $N$  fixé, et d'étudier ensuite l'objet obtenu. Cette question n'est étudiée, à notre connaissance, que dans deux contextes.

Le premier est à nouveau celui des abeilles browniennes. Dans [BBNP22], les auteurs démontrent en effet l'existence, pour  $N$  suffisamment grand, d'une unique loi stationnaire  $\psi^N$  pour

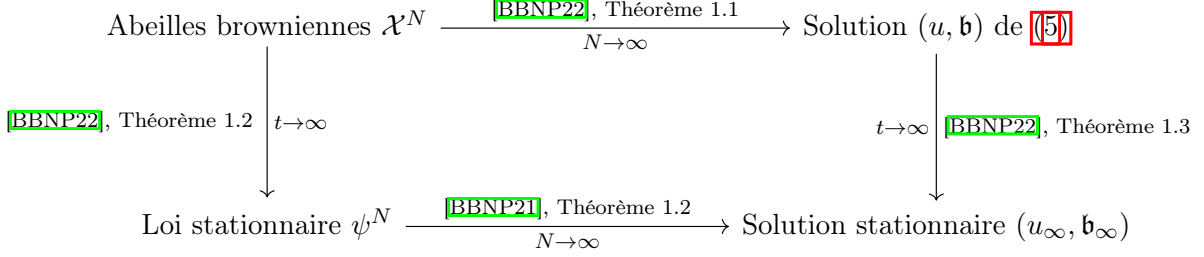
<sup>64</sup>Voir [BT24], Conjecture 1.6.

<sup>65</sup>En effet, le problème (2) appartient à la même classe d'universalité que l'équation de F-KPP, et le comportement asymptotique de la frontière obtenus par [BBD18] sont les mêmes que ceux obtenus pour l'équation de F-KPP par Bramson [Bra78, Bra83].

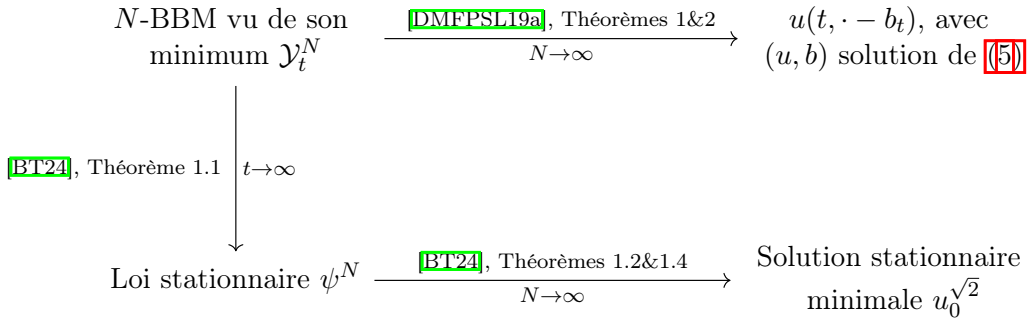
<sup>66</sup>Comme nous l'avons expliqué dans la Section 1.2, cela nous donne des indications sur le cas du  $N$ -processus de Bessel branchant.

<sup>67</sup>Voir [BBNP21], Théorème 1.2.

le modèle des abeilles browniennes à  $N$  particules, puis que cette suite de lois stationnaires converge, quand  $N$  tend vers l'infini, vers la solution stationnaire du problème (5); ces différentes convergences sont résumées dans le diagramme ci-dessous.



Plus récemment, Berestycki et Tough [BT24] ont établi des résultats analogues pour le  $N$ -BBM. Nous avons déjà vu que cette fois, le système de particules n'a pas de comportement stationnaire, puisqu'au contraire il est possible d'observer, au moins en choisissant convenablement la condition initiale, un déplacement de front d'onde. En revanche, vu de sa particule la plus basse, le  $N$ -BBM possède une loi stationnaire : rappelons que les positions des particules sont notées  $X_t^i$ , et celle de la particule la plus basse  $m_t^N$ ; Berestycki et Tough ont alors montré que le processus  $(\mathcal{Y}_t^N)_{t \geq 0}$  défini par  $Y_t^{N,i} = X_t^i - m_t^N$  admet, pour tout entier  $N \geq 1$ , une loi stationnaire  $\psi^N$ . De plus, ils ont démontré un *principe de sélection* par le  $N$ -BBM, au sens suivant : le problème à frontière libre (2) qui caractérise la limite hydrodynamique du  $N$ -BBM admet plusieurs solutions de type front d'onde, données par les lois quasi-stationnaires du mouvement brownien avec dérive; chacune de ces lois quasi-stationnaire est une candidate pour la loi limite des  $\psi^N$ , mais c'est en fait toujours celle de célérité minimale  $\sqrt{2}$  qui est effectivement obtenue, quelle que soit la loi de départ. Le diagramme obtenu est donc cette fois le suivant.



Notons que la Conjecture 1.6 de Berestycki et Tough [BT24] pourrait en partie combler la partie manquante de ce diagramme (pour les lois initiales dont la queue de distribution est suffisamment petite). En revanche, il existe dans le cas brownien des lois initiales pour lesquelles le diagramme ne commute pas (par exemple si l'on part d'une loi quasi-stationnaire  $u_0^\alpha$ , pour  $\alpha > \sqrt{2}$ ).

Nous aimerions donc étudier l'existence de lois stationnaires pour le  $N$ -BMP : par exemple, dans le cadre de processus de Lévy<sup>68</sup>, nous pouvons espérer que sous certaines hypothèses, le  $N$ -BMP vu depuis sa position la plus basse admette une loi stationnaire, de même que le  $N$ -BBM. Nous pourrions alors également étudier la convergence, quand  $N$  tend vers l'infini, de ces lois stationnaires, et déterminer si l'on observe également un principe de sélection, comme démontré par Berestycki et Tough dans le cas brownien.

<sup>68</sup>Si nous ne supposons pas que le processus est homogène en espace, le processus  $(\mathcal{Y}_t^N)_{t \geq 0}$  n'est plus markovien.

*Lien avec un modèle particulière de type Fleming-Viot.* Un comportement comparable à celui décrit dans les paragraphes précédents est déjà observé pour une instance particulière du modèle de Fleming-Viot [BHM96], qui se décrit comme suit : on considère  $N$  particules se déplaçant dans un certain domaine ouvert  $\mathcal{U} \subset \mathbb{R}^d$  fixé, indépendamment les unes des autres et selon la trajectoire d'un processus donné (par exemple un mouvement brownien); dès que l'une des particules sort de  $\mathcal{U}$ , elle saute à la position de l'une des  $N - 1$  autres particules, choisie aléatoirement de manière uniforme.

Le comportement asymptotique de ce modèle (en temps et en taille de population) a été étudié dans plusieurs contextes; il est analogue à celui du  $N$ -BBM dans le cas brownien, ce qui laisse imaginer que cette analogie pourrait s'étendre aux cas plus généraux présentés plus bas.

- Le même type de diagramme que ceux obtenus plus haut dans le cas des abeilles browniennes a été obtenu pour le modèle de Fleming-Viot pour un mouvement brownien sur un domaine borné régulier de  $\mathbb{R}^d$  (on peut donc choisir un disque centré en 0) : celui-ci admet d'une part<sup>69</sup> comme limite hydrodynamique la loi du mouvement brownien conditionné à ne pas sortir de ce domaine, et d'autre part<sup>70</sup>, pour tout  $N \geq 1$ , une unique loi stationnaire qui converge, quand  $N$  tend vers l'infini, vers la solution stationnaire  $(u_\infty, \mathbf{b}_\infty)$  au problème (5).
- De même, le modèle de Fleming-Viot pour le mouvement brownien avec dérive et tué en 0 se comporte comme le  $N$ -BBM. On peut appliquer le résultat de Villemonais [Vil14], qui établit dans un contexte général que la limite hydrodynamique de ce modèle de Fleming-Viot est, à  $t$  fixé, la loi du processus conditionné à ne pas être tué avant l'instant  $t \geq 0$ ; et ce modèle admet<sup>71</sup> également une loi stationnaire, à  $N$  fixé, qui converge, quand  $N$  tend vers l'infini, vers la *limite de Yaglom*<sup>72</sup> du mouvement brownien avec dérive tué en 0, qui est la loi de densité  $u_0^{\nu^2}$ . Comme dans le cas du  $N$ -BBM, le diagramme de convergence ne commute pas : le modèle de Fleming-Viot sélectionne la loi quasi-stationnaire minimale, alors que la convergence en temps long de la limite hydrodynamique dépend *a priori* de la loi initiale.
- Plus généralement, le fait que la loi du processus sous-jacent, conditionnée à la non-extinction, est la limite hydrodynamique du modèle de Fleming-Viot associé a été démontré dans un contexte général [Vil14]. Il a aussi été démontré pour plusieurs types de processus  $(X_t)_{t \geq 0}$  ayant une unique loi quasi-stationnaire  $\nu$  que le modèle de Fleming-Viot associé admet, à  $N$  fixé, une unique loi stationnaire  $\nu^N$ , et que celle-ci converge vers  $\nu$  quand  $N$  tend vers l'infini [FM07, AFG11, TN22]. De plus, le même phénomène de sélection que dans le cas du mouvement brownien a été mis en évidence dans le cas où  $(X_t)_{t \geq 0}$  a plus d'une loi quasi-stationnaire. Il a par exemple été démontré dans les cas du processus de Galton-Watson, de processus de naissance et de mort, et de diffusions [AFGJ16, Vil15, CV21], que le modèle de Fleming-Viot converge encore<sup>73</sup> vers la limite de Yaglom du processus  $(X_t)_{t \geq 0}$ .

<sup>69</sup>Voir [BHM00, GK04].

<sup>70</sup>Voir [BHM00].

<sup>71</sup>Voir [Tou23].

<sup>72</sup>Par définition (voir par exemple [MV12]), la *limite de Yaglom* d'un processus  $(X_t)_{t \geq 0}$  est à la fois la mesure  $\nu$  quasi-limite pour toutes les masses de Dirac, c.à.d. vérifiant pour tout état  $x$

$$\nu = \lim_{t \rightarrow +\infty} \mathbb{P}_x(X_t \in \cdot | \tau_0 > t),$$

et celle, parmi toutes les lois quasi-stationnaires pour  $(X_t)_{t \geq 0}$ , qui donne la durée de vie moyenne  $\mathbb{E}_\nu(\tau_0)$  minimale.

<sup>73</sup>Toujours au sens où il admet à  $N$  fixé une loi stationnaire qui converge vers  $\nu$  quand  $N$  tend vers l'infini.

- Nous renvoyons aux travaux de Groisman et Jonckheere pour une discussion détaillée sur les liens entre ce modèle particulière de Fleming-Viot et le  $N$ -BMP, ainsi que le principe de sélection que nous avons décrit, qu'ils présentent en détail dans le cadre des processus de Lévy [GJ19]. Ils démontrent également la correspondance exacte entre les solutions de type front d'onde pour le problème [2] généralisé (c.à.d. en remplaçant le laplacien par l'adjoint du générateur du processus de Lévy) et les lois quasi-stationnaires du processus associé [GJ18].

Nous pourrions donc explorer à quel point l'analogie des comportements entre le modèle de Fleming-Viot et le  $N$ -BMP, constatée dans le cas brownien, s'étend à des contextes plus généraux.

**Autres questions.** Nous mentionnons enfin des questions qui sont apparues au cours de discussions ou de lectures, qui ont retenu notre attention sans que nous ayons le temps de les étudier avant l'écriture de ce manuscrit.

- En étudiant le comportement asymptotique du  $N$ -BBM, Berestycki et Togh [BT24] ont pu affaiblir l'hypothèse de positions de départ i.i.d.<sup>74</sup> Il serait intéressant de vérifier si leur démonstration utilisant la théorie du transport optimal pour construire un couplage entre des  $N$ -BBM avec condition initiale i.i.d. ou non s'adapte au cas général du  $N$ -BMP.
- Nous nous demandons ce qu'il advient à l'échelle attendue pour les fluctuations : rappelons le résultat [4] que nous obtenons au Théorème 1,

$$\mathbb{P} \left( \sup_{r \in \mathbb{R}} |F^N(t, r) - F(t, r)| > N^{-\beta} \right) \leq c_2 e^{-c_1 N^{\frac{1-2\beta}{3}}},$$

qui est valable pour  $\beta \in (0, \frac{1}{2})$ . Dans le cas  $\beta = \frac{1}{2}$ , nous nous attendrions *a priori* à observer une convergence en loi pour la différence  $\sqrt{N}(F^N - F)$ .

- La question de la généralisation en dimension supérieure est naturelle, partant notamment du modèle des abeilles browniennes [BBNP22]. Notons par exemple que la notion de monotonie stochastique, nécessaire à notre démonstration, demande une adaptation dans cas-là puisque l'ordre total disparaît. Cependant, l'approche utilisée dans [BBNP22], où l'on se ramène à la dimension 1 en passant par la norme, pourrait éventuellement permettre d'adapter certaines parties de notre preuve.
- L'hypothèse de monotonie stochastique faite sur le processus  $(X_t)_{t \geq 0}$  est un argument important dans notre preuve, ainsi que dans la partie de celle de Klump et Savov [KS23] portant sur l'unicité de la frontière solution du problème inverse du premier temps de passage. Il est alors légitime de se demander si le  $N$ -BMP peut admettre une limite hydrodynamique dans les cas où il existerait plusieurs solutions au problème inverse du premier temps de passage, et si l'on observe alors un phénomène de sélection d'une frontière particulière parmi les solutions du problème inverse du premier temps de passage.

## 4.2 Flot comonotone d'un processus markovien sur la droite réelle

Nous avons mentionné à la Section 2.3 le fait que, sous les hypothèses du Théorème 3 (monotonie stochastique et propriété de Feller), les résultats du Chapitre II combinés au Théorème 1 de [LJR04] assurent l'existence d'un flot stochastique associé à la famille de semi-groupes

---

<sup>74</sup>Voir [BT24], Théorème A.1.

$((\check{P}_t^{(n)})_{t \geq 0})_{n \geq 1}$ , pour lequel il serait intéressant d’obtenir des résultats de régularité supplémentaires (comme par exemple la monotonie en espace et le caractère càdlàg en espace et en temps).

D’autre part, étant donné un processus  $(X_t)_{t \geq 0}$  solution d’une é.d.s.

$$dX_t = \beta(X_t) dt + \sigma(X_t) dB_t,$$

avec des hypothèses appropriées sur les coefficients  $\beta$  et  $\sigma$ , on peut construire un flot de solutions fortes à cette é.d.s. à partir des trajectoires d’un unique mouvement brownien standard (voir [Kun97, Kun19]). Nous supposons que, dans ce contexte, le flot comonotone coïncide avec le flot de solutions de l’é.d.s., mais nous n’avons pas cherché à approfondir cette question et la gardons en tête pour de futurs travaux.

Enfin, une question naturelle serait d’étudier la réalisabilité de la propriété de monotonie stochastique pour un processus de Markov général sans supposer la propriété de Feller<sup>75</sup>, dans l’esprit des résultats obtenus par Kellerer<sup>76</sup>.

### 4.3 Transport optimal en champ moyen

Les trois méthodes numériques que nous avons proposées fournissent de bons résultats dans les cas où nous les avons mises en œuvre : en particulier, elles approchent correctement la solution explicite connue dans le cas où la dérive dépend linéairement du contrôle, et le coût de manière quadratique<sup>77</sup>; et elles permettent d’observer que les agents se répartissent dans l’espace pour coopérer lorsqu’ils sont soumis à un effet de congestion rendant le déplacement plus coûteux dans les zones denses<sup>78</sup>.

Une piste de recherche dans la continuité du travail mené serait d’explorer plus avant le choix des hyperparamètres pour améliorer la performance des méthodes proposées, notamment en grande dimension.

Enfin, la question de la résolution théorique du problème de transport optimal en champ moyen est une piste de recherche naturelle, qui ne donne lieu à ce jour à des résultats que dans des cas particuliers relativement restreints. Il serait intéressant d’étudier ce problème, notamment en revenant à une version pénalisée sous la forme du problème de contrôle en champ moyen : ce dernier problème est étudié sur le plan théorique<sup>79</sup>, et l’on pourrait espérer étendre la méthode proposée par Hernández et Tangpi [HT23] pour approcher le problème de transport optimal en champ moyen par une version pénalisée, et en déduire l’existence de solutions.

---

<sup>75</sup>Nous expliquons à la Section C.1 pourquoi nous pensons qu’il manque un argument dans la démonstration proposée dans [KKO77] (Théorème 5).

<sup>76</sup>Voir [Kel72], Théorème 1. Pour une version en anglais, nous renvoyons au Théorème 2.15 de [BJ22], dont les auteurs démontrent également, entre autres choses, l’existence, pour une famille donnée  $(\mu_t)_{t \geq 0}$  de lois sur  $\mathbb{R}$  croissante pour l’ordre stochastique, d’un processus markovien  $(X_t)_{t \geq 0}$  à trajectoires croissantes presque sûrement et tel que la loi de  $X_t$  est  $\mu_t$ , pour tout  $t \geq 0$ .

<sup>77</sup>Voir Section III.4.1

<sup>78</sup>Voir Section III.4.2

<sup>79</sup>Voir [BFY13, AL15].



# Chapter I

## Hydrodynamic limit of the $N$ -BMP

This chapter contains the main part of [BF23a], that can be found at <https://hal.science/hal-04296610>. Appendices A and B contain some additional details to this work.

### I.1 Introduction

We consider a branching-selection stochastic process that consists in a finite family of  $N \geq 2$  particles moving on the real line in continuous time. At rate 1, each particle gives birth to a new particle at its current location (branching), and, in order to preserve a constant number of particles, the particle currently located at the lowest position<sup>1</sup> is removed instantly (selection). Between such branching-selection events, particles move independently according to a continuous-time Markov process  $X$  on  $\mathbb{R}$ .

This process belongs to a family of models introduced by Brunet and Derrida [BD97], which then gave rise to a series of related works (see Section I.1.2 for a bibliographical discussion). Among these models, the so-called  $N$ -branching Brownian motion ( $N$ -BBM), where the underlying Markov process is the standard Brownian motion on  $\mathbb{R}$ , fits into our setting. In [DMFPSL19a, DMFPSL19b], De Masi, Ferrari, Presutti and Soprano-Loto studied the hydrodynamic limit of the  $N$ -BBM (fixed time  $t$ , with  $N \rightarrow +\infty$ ) and showed that it can be described as the solution of a free boundary problem. Our main goal is to extend this result to the case of a general Markov process on  $\mathbb{R}$ . Accordingly, we call this process the  $N$ -branching Markov process ( $N$ -BMP).

#### I.1.1 Statement of the main results

Let  $X = (X_t)_{t \geq 0}$  be a time-homogeneous Markov process on  $\mathbb{R}$  (in the sense of [RY99], Chapter III) with transition kernel  $(P_t)_{t \geq 0}$ , defined on a filtered measurable space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0})$  equipped with a family of probability measures  $(\mathbb{Q}_x)_{x \in \mathbb{R}}$ . We consider a probability measure  $\mu_0$  on  $\mathbb{R}$  and denote by  $\mathbb{Q}_{\mu_0}$  the probability measure on  $(\Omega, \mathcal{F})$  in the case where the distribution of  $X_0$  is  $\mu_0$ . As usual, we also denote by  $(P_t)_{t \geq 0}$  the semigroup of  $X$ , which is defined on the space  $\mathcal{B}_b$  of bounded real-valued Borel functions on  $\mathbb{R}$  by

$$\forall f \in \mathcal{B}_b \quad \forall t \geq 0 \quad \forall x \in \mathbb{R} \quad P_t f(x) = \mathbb{E}_x(f(X_t)) = \int_{\mathbb{R}} f(y) P_t(x, dy).$$

We will also write  $\mathcal{C}_b$  (resp.  $\mathcal{C}_0$ ) for the set of continuous bounded (resp. vanishing at  $\pm\infty$ ) real-valued functions on  $\mathbb{R}$ .

Let  $b : (0, +\infty) \rightarrow \mathbb{R}$  be an upper semi-continuous curve, let also<sup>2</sup>  $b_0 = \limsup_{t \rightarrow 0^+} b_t \in$

<sup>1</sup>If several particles are simultaneously located at the lowest position, ties are broken arbitrarily.

<sup>2</sup>This way,  $b$  is upper semi-continuous on  $[0, +\infty)$ .

$[-\infty, +\infty)$ , and define<sup>3</sup>

$$\tau_b = \inf\{s \geq 0, X_s < b_s\}.$$

Our main results relate the hydrodynamic limit of the  $N$ -BMP to the existence of a curve  $b$  (which depends on  $\mu_0$ ) satisfying the following assumption:

- (i) for all  $t > 0$ ,  $\mathbb{Q}_{\mu_0}(\tau_b > t) = e^{-t}$ .

In other words, (i) means that, under  $\mathbb{Q}_{\mu_0}$ ,  $\tau_b$  follows the exponential distribution  $\mathcal{E}(1)$ . We write for  $r \in \mathbb{R}$

$$U(t, r) = \mathbb{Q}_{\mu_0}(X_t \leq r, \tau_b > t),$$

so that under Assumption (i), the function  $e^t U(t, \cdot)$  is the c.d.f. of  $X$  conditioned on staying above the curve  $b$  up until  $t$ , that is

$$e^t U(t, r) = \mathbb{Q}_{\mu_0}(X_t \leq r | \tau_b > t).$$

Moreover, we rely on the following two additional assumptions on  $X$ :

- (ii)  $X$  is a Feller process<sup>4</sup>,  
 (iii)  $X$  is stochastically non-decreasing, *i.e.* for every non-decreasing bounded function  $f$ ,  $x \leq y$  and  $t \geq 0$ , we have  $P_t f(x) \leq P_t f(y)$ .

We make the usual assumption that the filtration  $(\mathcal{F}_t)_{t \geq 0}$  is complete and right-continuous and that the process  $X$  takes values in the set  $\mathcal{D}(\mathbb{R}_+, \mathbb{R})$  of real-valued càdlàg functions on  $\mathbb{R}_+$  (see for instance Theorem 19.15 in [Ka02]). In this context,  $\tau_b$  is a stopping time<sup>5</sup> such that, on  $\{\tau_b < +\infty\}$ , we have  $X_{\tau_b} \leq b_{\tau_b}$ .

Given an integer  $N \geq 2$ , we denote by  $X_t^1, \dots, X_t^N$  the positions<sup>6</sup> of the particles in the  $N$ -BMP at time  $t \geq 0$ , and by  $F^N(t, \cdot)$  the empirical c.d.f. of the  $N$ -BMP:

$$F^N(t, r) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{X_t^i \leq r\}}.$$

We will always consider the case where the initial positions of the particles are i.i.d. with common distribution  $\mu_0$ , and we denote by  $\mathbb{P}$  the corresponding probability measure governing the  $N$ -BMP.

---

<sup>3</sup>Notice that with our assumptions, we have

$$\tau_b = \tau'_b := \inf\{s > 0, X_s < b_s\},$$

which is more commonly used in the literature concerning the inverse first-passage time problem. Let us prove this by contradiction: assume that  $\tau_b < \tau'_b$ . Then we necessarily have  $X_0 < b_0$  and  $\tau'_b > 0$ . Since  $X$  has càdlàg paths, we then have  $X_t < b_0$  for all  $t$  small enough. Recalling that  $b_0 = \limsup_{t \rightarrow 0^+} b_t$ , there exists a non-increasing sequence  $(t_k)_k$  s.t.  $t_k \rightarrow 0$  and  $b_{t_k} \rightarrow b_0$ . Now by right-continuity, we have  $X_{t_k} \rightarrow X_0$ , so that  $X_0 < b_0$  implies that  $X_{t_k} < b_{t_k}$  for  $k$  large enough. This proves  $\tau'_b = 0$ , which is a contradiction.

<sup>4</sup>By Feller process, we mean that its semigroup is Feller: it maps  $\mathcal{C}_0$  into  $\mathcal{C}_0$  and is strongly continuous (see §III.2 in [RY99]). We will also use that these semigroups map  $\mathcal{C}_b$  into  $\mathcal{C}_b$  (which we can deduce from the former assumption by writing any bounded continuous function as a limit of bounded functions that vanish at infinity).

<sup>5</sup>See Section A.2.1.2 for a proof of this fact.

<sup>6</sup>The details on how we set indices for the particles are discussed in Section A.1.2

**Theorem I.1.** Consider a time horizon  $T > 0$ , and assume (i), (ii) and (iii) are fulfilled. Then for every  $t$  in  $[0, T]$ ,  $\|F^N(t, \cdot) - e^t U(t, \cdot)\|_\infty$  goes a.s. to 0 as  $N$  goes to infinity, where  $\|\cdot\|_\infty$  is the supremum norm on  $\mathcal{B}_b$ . Furthermore, for every  $\beta \in (0, \frac{1}{2})$ , there exist positive constants  $c_1, c_2 > 0$  depending only on  $\beta$  and  $T$  such that for all  $N$ , we have

$$\mathbb{P}\left(\|F^N(t, \cdot) - e^t U(t, \cdot)\|_\infty > N^{-\beta}\right) \leq c_2 e^{-c_1 N^{\frac{1-2\beta}{3}}}. \quad (\text{I.1})$$

A natural question is then whether, as  $N$  goes to infinity, the trajectory of the lowest particle position in the  $N$ -BMP approaches  $b$ . We investigate this question in [Theorem I.2](#). To state our result, we need an additional assumption on  $X$ , known in the literature as *Ray's estimate*: denote by  $\tau_y$  the (downwards) crossing time of  $y$  by  $X$ , i.e.

$$\tau_y = \inf\{s \geq 0, X_s < y\},$$

and assume that:

(iv) for all real numbers  $x > y$  and  $\kappa > 0$ , we have

$$\mathbb{Q}_x(\tau_y \leq t) = o(t^\kappa), \quad (\text{I.2})$$

(v) the curve  $b$  is continuous on  $(0, +\infty)$ .

Let  $m_t^N = \min_{1 \leq i \leq N} X_t^i$  be the lowest particle position at time  $t$ .

**Theorem I.2.** Consider  $0 < t_0 < T$  and assume (i) to (v) are fulfilled. Then, almost surely, the trajectory  $t \mapsto m_t^N$  converges to the curve  $t \mapsto b_t$ , uniformly on  $[t_0, T]$  as  $N$  goes to infinity. Furthermore, for any  $\varepsilon > 0$  and  $\kappa > 0$  there exists a constant  $C > 0$  (depending on  $\varepsilon, \kappa, T$  and also on  $X$ ) such that for all  $N$ , we have

$$\mathbb{P}\left(\sup_{t \in [t_0, T]} |m_t^N - b_t| > \varepsilon\right) \leq CN^{-\kappa}. \quad (\text{I.3})$$

In addition, these results still hold with  $t_0 = 0$  provided that  $b$  is continuous at 0.

## I.1.2 Discussion

### I.1.2.1 Assumptions

Our initial understanding of Assumption (i) was as a probabilistic reformulation of an existence result for a free boundary problem, as discussed in the next paragraph. After (almost) completing the present manuscript, we realized that Assumption (i) had already been studied in the context of the so-called inverse first-passage time problem, for which a substantial body of literature exists; our main reference is Klump and Savov [\[KS23\]](#), whose introduction contains an overview of the subject and relevant works. In [\[KS23\]](#), it is proved that Assumption (i) is satisfied for every Feller process such that  $\mathbb{Q}_{\mu_0}(X_t \in \cdot)$  is diffuse for all  $t$  and such that for all upper semi-continuous curve  $b' : (0, +\infty) \rightarrow \mathbb{R}$ ,  $\mathbb{Q}_{\mu_0}(\tau_{b'} = \tilde{\tau}_{b'}) = 1$ , where  $\tilde{\tau}_{b'} = \inf\{s \geq 0, X_s \leq b'_s\}$ ; it is also proved that these assumptions are satisfied by a wide class of Lévy processes and diffusion processes.

One can also understand Assumption (i) as equivalent (at least at a heuristic level, we refrain from giving a precise mathematical statement here) to the existence of a solution to the following problem:

$$\begin{cases} \partial_t u(t, x) = A^* u(t, x) + u(t, x), & x > b_t, \\ u(t, x) = 0, & x \leq b_t, \end{cases} \quad (\text{I.4})$$

where  $u(t, \cdot)$  is a probability density function,  $A^*$  is the adjoint of the infinitesimal generator of  $X$ , and  $b$  is part of the problem as a free boundary. In the case where  $X$  is a standard Brownian motion, so that  $A^* = A = \frac{1}{2}\partial_x^2$ , this problem has been solved by Berestycki, Brunet and Penington in [BBP19] and by Lee<sup>7</sup> in [Lee18].

Another situation in which we know that Assumption (i) is verified is the case of Markov processes starting from a quasi-stationary distribution on  $[0, +\infty)$ , in which case  $b$  is constant: it is well-known that in this case, the first crossing time of the origin is exponentially distributed<sup>8</sup>. Various authors have given some conditions under which such distribution exist, starting with the case of the Brownian motion with negative drift [MSM94], followed by more general one-dimensional diffusion (see [CMSM13] and references therein). Several conditions have also been given for Lévy processes [KP06, GJ18, Yam23].

Assumption (ii) (Feller property) is quite classical for Markov processes and we do not discuss it any further.

Assumption (iii) (stochastic monotonicity) is automatically verified for Feller processes with continuous paths as well as Lévy processes (see [BF23b] for a more detailed discussion about this assumption).

Assumption (iv) (Ray's estimate) is verified for any Feller process with continuous paths (see Section 3.8 in [IM12], with a method from Ray [Ray56]).

Assumption (v) is, to our knowledge, established for standard Brownian motion (see [BBP19, CCCS11]) and, more generally, diffusions with smooth coefficients [CCS22].

### I.1.2.2 Related works

The hydrodynamic limit shown in Theorem I.1 has already been established by De Masi, Ferrari, Presutti and Soprano-Loto for the  $N$ -BBM in [DMFPSL19a, DMFPSL19b] (allowing a more general non-local branching mechanism in [DMFPSL19b]), with a control upon the probabilities of deviations for the supremum norm given by a negative power of  $N$ . The multi-dimensional case (the so-called Brownian bees) has also been studied by Berestycki, Brunet, Nolen and Penington in [BBNP22], where this time the killed particle is the furthest away from the origin. Taking Euclidean norms of particles' positions, this model leads to an  $N$ -BMP driven by the opposite of a Bessel process. The authors of [BBNP22] not only obtain the hydrodynamic limit, but also a convergence result analogous to our Theorem I.2, though only at a fixed time  $t$  instead of the uniform convergence over  $[t_0, T]$ .

As in [DMFPSL19a, DMFPSL19b, BBNP22], our proof relies on a comparison between the  $N$ -BMP and auxiliary stochastic processes. However, our version of these processes is based on a different definition where the curve  $b$  plays a central role. This approach allows us to get improved bounds (stretched exponential vs. negative power) for the hydrodynamic limit, thus allowing us to also prove the uniform convergence of the lowest position, extending the strategy used in [BBNP22]<sup>9</sup>. It is worth mentioning that the result obtained in Theorem I.1 is universal: the constants that control the deviations of the distribution of the  $N$ -BMP do not depend on the underlying Markov process  $X$  or the boundary  $b$  (as soon as the assumptions are verified). However, in Theorem I.2, the constants depend on  $X$  both through Ray's estimate and the moduli of continuity of  $b$ .

<sup>7</sup>To be precise, the former prove the existence of global solution for (I.4) with  $A = \frac{1}{2}\partial_x^2$ , while the latter proves in [Lee18] the existence of local solutions to a more general version of the free boundary problem corresponding to the version of the  $N$ -BBM with non-local branching that appears in [DMFPSL19b].

<sup>8</sup>See for instance [CMSM13] Theorem 2.2. Notice that we might need to rescale the process in time to obtain the exponential distribution with parameter 1.

<sup>9</sup>For this last problem, the negative power bound we obtain comes from the use of Ray's estimate, which holds for very general Markov processes; in the Brownian case, our strategy would lead to a stretched exponential bound by using the stronger control on (I.2) available in this specific case.

In addition to these works, the hydrodynamic limit has also been investigated for a variety of branching-selection systems with a different selection and/or branching mechanism, e.g. [DR11, GJM20, GSL21, Ata23], often in connection with free boundary problems.

### I.1.3 Outline of the proof

To establish the hydrodynamic limit in [DMFPSL19a, DMFPSL19b, BBNP22], a key idea is to compare the particle process with stochastic barriers in which selection is either delayed or anticipated so that it occurs at fixed times rather than at the branching times. Here we use a different approach that takes advantage of the assumed existence of a curve  $b$  satisfying condition (i), and compare the  $N$ -BMP to another process that we call the  $b$ -BMP. To be specific, starting from  $n$  i.i.d. initial positions with distribution  $\mu_0$ , the  $b$ -BMP is obtained by running a (dyadic) branching Markov process (BMP) based on the dynamics of  $X$ , while removing (or "killing") particles as soon as they cross the moving boundary  $b$ . When  $b$  satisfies condition (i), the many-to-one formula (see Lemma I.A below) implies that the expected population size of the  $b$ -BMP at any given time  $t$  is equal to  $n$ , since the exponential growth due to branching is, on average, exactly offset by the exponential decay due to killing below  $b$  while starting with distribution  $\mu_0$ . This suggests that the  $b$ -BMP starting with  $n = N$  particles should in some sense be close to the  $N$ -BMP, where selection keeps the population size exactly equal to  $N$ .

The  $b$ -BMP turns out to be more tractable than the  $N$ -BMP, since the way particles are removed in the  $b$ -BMP does not involve the complicated interaction implied by the  $N$ -BMP selection mechanism<sup>10</sup>. When the initial size  $n$  is large, we are able to prove in Proposition I.1 that, for the  $b$ -BMP at time  $t$ , with high probability, the population size is indeed of order  $n$ , and the empirical distribution of particle positions is close to  $\mathbb{Q}_{\mu_0}(X_t \in \cdot | \tau_b > t)$  (we actually prove explicit, non-asymptotic bounds).

Our idea is then to bound the  $N$ -BMP  $\mathcal{X}$  between two copies  $\mathcal{X}^+$ ,  $\mathcal{X}^-$  of the  $b$ -BMP starting respectively with  $n = \lceil N(1+\delta) \rceil$  and  $n = \lfloor N(1-\delta) \rfloor$  particles. More precisely, in Proposition I.4, we build a coupling between  $\mathcal{X}$  and  $\mathcal{X}^+$ ,  $\mathcal{X}^-$  in such a way that, provided that  $\#\mathcal{X}_s^+ \geq N$  for all  $s \leq t$ , one has  $\mathcal{X}_t \preceq \mathcal{X}_t^+$ , and provided that  $\#\mathcal{X}_s^- \leq N$  for all  $s \leq t$ , one has  $\mathcal{X}_t^- \preceq \mathcal{X}_t$ , where, given two real-valued tuples  $x, y$  (possibly with different sizes),  $x \preceq y$  means that, for all  $r$  in  $\mathbb{R}$ , we have  $\#\{i, x^i \geq r\} \leq \#\{i, y^i \geq r\}$ . The stochastic monotonicity (Assumption (iii)) is key to the construction of the coupling.

Then, thanks to our control on the population size of the  $b$ -BMP, we are able to prove in Proposition I.3 that, given a finite time horizon  $T$ , for large  $N$ , with high probability, the population size of  $\mathcal{X}^+$  (resp.  $\mathcal{X}^-$ ) remains above (resp. below)  $N$  over the time-interval  $[0, T]$ , for suitably small values of  $\delta$ . In turn, this allows us to compare the empirical distribution of particles within the  $N$ -BBM at time  $t$  to  $\mathbb{Q}_{\mu_0}(X_t \in \cdot | \tau_b > t)$ , leading to Theorem I.1.

The proof of Theorem I.2 then combines the result of Theorem I.1 together with Ray's estimate to control the difference between  $b_t$  and  $m_t^N$ . Our proof is similar in spirit to that of Berestycki, Brunet, Nolen and Penington in [BBNP22] (namely Section 4.2 about the proof of their Proposition 1.6). Compared to [BBNP22], we make the most of our improved bounds to obtain uniform convergence, even though we are using a weaker control on the trajectories of  $X$  (given by Ray's estimate).

### I.1.4 Organization of the paper

The rest of the paper is organized as follows. We start by studying the  $b$ -BMP and the stochastic barriers in Section I.2, where we prove a hydrodynamic limit result for the  $b$ -BMP (Proposi-

<sup>10</sup>Such an observation is not new: with  $b$  a suitably chosen linear boundary, it plays a substantial role in the study the long-time behaviour of the  $N$ -BBM [Mai16] or the  $N$ -BRW (branching random walk) [BG10].

tion I.1) and then deduce Proposition I.2. In this section we also study the size of the  $b$ -BMP in order to prove Proposition I.3. In Section I.3, we prove Proposition I.4 (up to some technical results that we postpone to the appendices) and deduce Theorem I.1. Finally, in Section I.4 we prove Theorem I.2 thanks to the additional regularity Assumptions (iv) and (v).

Appendix A is devoted to technical arguments used to prove the existence of a coupling between the  $N$ -BMP and its stochastic barriers. More precisely in Section A.1, we build on the bases given by Ikeda, Nagasawa and Watanabe in [INW68b] to show some results about the pasting of coupled Markov processes, which we use in Section A.2 to formally construct our coupling and show that it has the desired properties. Appendix B contains a technical lemma on the geometric distribution used in the main proofs of the paper.

## I.2 Study of the $b$ -BMP

### I.2.1 Hydrodynamic limit of the $b$ -BMP

Recall the definition of the (dyadic) branching Markov process (BMP), a special case of the class of processes studied in [INW68a]: we have a finite number of particles on the real line following independent trajectories of the Markov process  $X$ ; each particle branches at rate 1 independently of the others, giving birth to another particle at the same location, which immediately starts to follow yet another independent trajectory of the Markov process  $X$ , and so on. The corresponding  $b$ -BMP is obtained by removing particles as soon as they cross the moving boundary  $b$ .

To be more precise, let us denote by  $\mathfrak{X}$  a BMP starting at time 0 with a single particle whose location follows the distribution  $\mu_0$ , and let  $\mathcal{N}_t^{\mathfrak{X}}$  denote the population size of  $\mathfrak{X}$  at time  $t$ . By identifying the trajectory of a particle before its birth to that of its parent, we can consider, for each  $j = 1, \dots, \mathcal{N}_t^{\mathfrak{X}}$ , the trajectory  $(X_s^j)_{s \geq 0}$  of the  $j$ -th particle<sup>11</sup> at time  $t$ , and the corresponding crossing time  $\tau_b^j$  of the boundary. The number of particles in the corresponding  $b$ -BMP whose locations are above  $r$  at time  $t$ , is then given by

$$\chi_t(\mathfrak{X}, r) = \sum_{j=1}^{\mathcal{N}_t^{\mathfrak{X}}} \mathbf{1}_{\{X_t^j > r, \tau_b^j > t\}}.$$

We now recall the many-to-one lemma, which we state here in the version that will be useful throughout the paper (see Lemma 1 and Section 4.1 in [HR17]).

**Lemma I.A** (Many-to-one lemma). *Suppose that  $\mathfrak{X} = (X_t^i)_{t \geq 0, 1 \leq i \leq \mathcal{N}_t^{\mathfrak{X}}}$  is a BMP starting from one particle with distribution  $\mu_0$ , and let  $f : \mathcal{D}([0, t], \mathbb{R}) \rightarrow \mathbb{R}$  be a bounded measurable function on the set of càdlàg functions from  $[0, t]$  to  $\mathbb{R}$ , for some positive time  $t$ . Then we have*

$$\mathbb{E} \left( \sum_{i=1}^{\mathcal{N}_t^{\mathfrak{X}}} f(X_{[0,t]}^i) \right) = e^t \mathbb{E}_{\mu_0}(f(X_{[0,t]})). \quad (\text{I.5})$$

Defining  $G(t, r)$  for  $r$  in  $\mathbb{R}$  and  $t \geq 0$  by

$$G(t, r) = 1 - e^t U(t, r) = e^t \mathbb{Q}_{\mu_0}(X_t > r, \tau_b > t) = \mathbb{Q}_{\mu_0}(X_t > r | \tau_b > t), \quad (\text{I.6})$$

we deduce from the many-to-one lemma (Lemma I.A) that, for  $r$  in  $\mathbb{R}$  and  $t \geq 0$ ,

$$\mathbb{E}(\chi_t(\mathfrak{X}, r)) = G(t, r). \quad (\text{I.7})$$

<sup>11</sup>The particle present at time 0 is given number 1; successively born particles are then numbered by birth order. The quantities we consider are in fact independent of which numbering scheme we adopt.

Now let  $\mathfrak{X}^1, \mathfrak{X}^2, \dots$  denote an i.i.d. sequence of copies of  $\mathfrak{X}$ , so that, for all  $n \geq 1$ ,  $\mathfrak{X}^1, \dots, \mathfrak{X}^n$  joined together form a BMP starting with  $n$  particles. Accordingly, define the population sizes  $\mathcal{N}_t^i$ , the positions  $(X_t^{ij})_{1 \leq j \leq \mathcal{N}_t^i}$  and the crossing times  $(\tau_b^{ij})_{1 \leq j \leq \mathcal{N}_t^i}$  corresponding to  $\mathfrak{X}^i$ , for  $i \geq 1$ , and set

$$G^n(t, r) = \frac{1}{n} \sum_{i=1}^n \chi_t(\mathfrak{X}^i, r),$$

so that  $G^n$  is the complement of the empirical c.d.f. of a  $b$ -BMP starting from  $n$  particles.

Given [\(I.7\)](#) and the fact that the random variables  $\chi_t(\mathfrak{X}^1, r), \chi_t(\mathfrak{X}^2, r), \dots$  are i.i.d., the law of large numbers shows that  $G(t, r)$  is the limit of  $G^n(t, r)$  as  $n$  goes to infinity, so that the hydrodynamic limit of the  $b$ -BMP can be identified as the underlying Markov process  $X$  conditioned on not crossing the boundary. The following result provides a much more precise control on the difference between  $G^n$  and  $G$ , with quantitative non-asymptotic bounds on the sup-norm  $\|G^n(t, \cdot) - G(t, \cdot)\|_\infty$ .

**Proposition I.1.** *Consider a time horizon  $T > 0$ , and let  $c_3 = -\ln(1 - e^{-T}) > 0$ ; for all  $t \in [0, T]$ , let  $C(t) = 2e^{2t} - e^t$ . Then for any  $\eta > 0$ ,  $t \in [0, T]$ ,  $n \geq \mathcal{N}_0 = \mathcal{N}_0(\eta, t) = \max\left(\frac{8C(t)}{\eta^2}, 2\right)$  and  $\alpha \in (0, \frac{1}{2})$ , we have the following bound:*

$$\mathbb{P}(\|G^n(t, \cdot) - G(t, \cdot)\|_\infty > \eta) \leq 16n^2 e^{-\frac{n^{1-2\alpha}\eta^2}{32}} + 32e^T n^2 e^{-c_3 n^\alpha}. \quad (\text{I.8})$$

Invoking the Borel-Cantelli lemma, an immediate corollary of [Proposition I.1](#) is that, for all  $t \geq 0$ , almost surely,  $\lim_{n \rightarrow +\infty} \|G^n(t, \cdot) - G(t, \cdot)\|_\infty = 0$ . As another corollary, we have the following bound on the total population size of the  $b$ -BMP.

**Corollary I.1.** *Let  $\mathcal{X} = (X_t^{ij}, 1 \leq i \leq n, 1 \leq j \leq \mathcal{N}_t^i, 0 \leq t \leq T)$  be a BMP starting from  $n$  particles independently distributed according to  $\mu_0$ , and let  $N_t$  be the population size at time  $t$  of the  $b$ -BMP made from  $\mathcal{X}$ . Then [12](#) for any  $\eta > 0$ ,  $t \in [0, T]$ ,  $n \geq \mathcal{N}_0(\eta, t)$  and  $\alpha \in (0, \frac{1}{2})$ , we have*

$$\mathbb{P}(|N_t - n| > n\eta) \leq 16n^2 e^{-\frac{n^{1-2\alpha}\eta^2}{32}} + 32e^T n^2 e^{-c_3 n^\alpha}. \quad (\text{I.9})$$

**Remark I.1.** *Note that:*

- *The exponential decay is driven by universal constants, in the sense that they depend on  $T$  but not on the initial distribution  $\mu_0$  nor on the choice of the Markov process that drives the BMP.*
- *The fact that a  $b$ -BMP starting from  $n$  particles is the joining of  $n$  independent copies of  $b$ -BMPs starting with one particle is key to the proof of the proposition, and this is why we study this process rather than studying the  $N$ -BMP directly.*
- *Our proof is an adaptation of the proof of Theorem 12.4 in [\[DGL96\]](#), which is a quantitative version of the Glivenko-Cantelli theorem, using symmetrization ideas introduced by Dudley [\[Dud78\]](#) and Pollard [\[Pol84\]](#). Here, we adapted this proof to the context of branching Markov processes.*

*Proof of [Proposition I.1](#)*

*Step 1. Symmetrization by a second load of BMPs.*

Fix an integer  $n$  and  $\eta > 0$ , let  $\tilde{\mathfrak{X}}^1, \dots, \tilde{\mathfrak{X}}^n$  be  $n$  other independent (between themselves and from the  $\mathfrak{X}^i$ ) copies of  $\mathfrak{X}$ , and use the tildes in the notation for all random objects linked

<sup>12</sup>The definition of  $\mathcal{N}_0$  is given in [Proposition I.1](#)

with them. Also write  $\mathcal{X} = (\mathfrak{X}^1, \dots, \mathfrak{X}^n)$  and  $\tilde{\mathcal{X}} = (\tilde{\mathfrak{X}}^1, \dots, \tilde{\mathfrak{X}}^n)$ . Let us first prove that for  $n \geq \frac{8C(t)}{\eta^2}$ , we have

$$\mathbb{P}(\|G^n(t, \cdot) - G(t, \cdot)\|_\infty > \eta) \leq 2\mathbb{P}\left(\|G^n(t, \cdot) - \tilde{G}^n(t, \cdot)\|_\infty > \frac{\eta}{2}\right). \quad (\text{I.10})$$

To this end, define the real-valued random variable  $r^*$  by

$$r^* = \begin{cases} \inf\{r \in \mathbb{R}, |G^n(t, r) - G(t, r)| > \eta\} & \text{if the set is not empty,} \\ 0 & \text{otherwise.} \end{cases}$$

Note that  $r^*$  is  $\mathcal{X}$ -measurable. First we have

$$\begin{aligned} \mathbb{P}\left(\|G^n(t, \cdot) - \tilde{G}^n(t, \cdot)\|_\infty > \frac{\eta}{2}\right) &\geq \mathbb{P}\left(|G^n(t, r^*) - \tilde{G}^n(t, r^*)| > \frac{\eta}{2}\right) \\ &\geq \mathbb{P}\left(|G^n(t, r^*) - G(t, r^*)| > \eta, \right. \\ &\quad \left. |\tilde{G}^n(t, r^*) - G(t, r^*)| < \frac{\eta}{2}\right). \end{aligned}$$

Then we condition on  $\mathcal{X}$  and use that both  $G^n(t, r^*)$  and  $G(t, r^*)$  are  $\mathcal{X}$ -measurable to get

$$\begin{aligned} \mathbb{P}\left(\|G^n(t, \cdot) - \tilde{G}^n(t, \cdot)\|_\infty > \frac{\eta}{2}\right) &\geq \mathbb{E}\left(\mathbf{1}_{\{|G^n(t, r^*) - G(t, r^*)| > \eta\}} \mathbb{P}\left(|\tilde{G}^n(t, r^*) - G(t, r^*)| < \frac{\eta}{2} \mid \mathcal{X}\right)\right). \end{aligned} \quad (\text{I.11})$$

Assume that

$$\mathbb{P}\left(|\tilde{G}^n(t, r^*) - G(t, r^*)| < \frac{\eta}{2} \mid \mathcal{X}\right) \geq \frac{1}{2} \text{ a.s.} \quad (\text{I.12})$$

From [\(I.11\)](#) and [\(I.12\)](#), one has

$$\mathbb{P}\left(\|G^n(t, \cdot) - \tilde{G}^n(t, \cdot)\|_\infty > \frac{\eta}{2}\right) \geq \frac{1}{2}\mathbb{P}\left(|G^n(t, r^*) - G(t, r^*)| > \eta\right),$$

and, noting that  $|G^n(t, r) - G(t, r)|$  is greater than  $\eta$  for some  $r$  in  $\mathbb{R}$  if and only if it is for  $r = r^*$ , [\(I.10\)](#) follows. We now prove [\(I.12\)](#). Since  $r^*$  is  $\mathcal{X}$ -measurable, while  $\tilde{\mathfrak{X}}^1, \dots, \tilde{\mathfrak{X}}^n$  are i.i.d. copies of  $\mathfrak{X}$  independent from  $\mathcal{X}$ , we have

$$\mathbb{E}(\tilde{G}^n(t, r^*) \mid \mathcal{X}) = G(t, r^*)$$

and

$$\mathbb{V}(\tilde{G}^n(t, r^*) \mid \mathcal{X}) = \frac{1}{n} \mathbb{V}\left(\sum_{j=1}^{\mathcal{N}_t^{\tilde{\mathfrak{X}}}} \mathbf{1}_{\{X_t^j > r^*, \tau_b^j > t\}} \mid \mathcal{X}\right),$$

where  $\mathfrak{X}$  is assumed to be independent from  $\mathcal{X}$ . We can then bound the sum by  $\mathcal{N}_t^{\tilde{\mathfrak{X}}}$ , which follows the geometric distribution  $\mathcal{G}(e^{-t})$ , and thus we have

$$\mathbb{V}(\tilde{G}^n(t, r^*) \mid \mathcal{X}) \leq \frac{1}{n} \mathbb{E}\left((\mathcal{N}_t^{\tilde{\mathfrak{X}}})^2\right) = \frac{C(t)}{n},$$

where  $C(t) = 2e^{2t} - e^t$ . By Chebyshev's inequality, we obtain

$$\mathbb{P}\left(|\tilde{G}^n(t, r^*) - G(t, r^*)| < \frac{\eta}{2} \mid \mathcal{X}\right) \geq 1 - \frac{4C(t)}{n\eta^2},$$



and under the condition  $n \geq \frac{8C(t)}{\eta^2}$ , the conditional probability is indeed greater than  $\frac{1}{2}$ .

*Step 2. Second symmetrization with random signs.*

Let  $\varepsilon_1, \dots, \varepsilon_n$  be i.i.d. Rademacher random variables independent from  $\mathcal{X}$  and  $\tilde{\mathcal{X}}$ . Since  $\chi_t(\mathfrak{X}^i, r)$  and  $\chi_t(\tilde{\mathfrak{X}}^i, r)$  have the same distribution and are independent for every real number  $r$ ,  $\chi_t(\mathfrak{X}^i, r) - \chi_t(\tilde{\mathfrak{X}}^i, r)$  and  $\varepsilon_i (\chi_t(\mathfrak{X}^i, r) - \chi_t(\tilde{\mathfrak{X}}^i, r))$  also have the same distribution, hence so do

$\|G^n(t, \cdot) - \tilde{G}^n(t, \cdot)\|_\infty$  and  $\left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i (\chi_t(\mathfrak{X}^i, \cdot) - \chi_t(\tilde{\mathfrak{X}}^i, \cdot)) \right\|_\infty$ . Then

$$\mathbb{P} \left( \|G^n(t, \cdot) - \tilde{G}^n(t, \cdot)\|_\infty > \frac{\eta}{2} \right) = \mathbb{P} \left( \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i (\chi_t(\mathfrak{X}^i, \cdot) - \chi_t(\tilde{\mathfrak{X}}^i, \cdot)) \right\|_\infty > \frac{\eta}{2} \right),$$

and we can now use the triangle inequality and the union bound to obtain

$$\begin{aligned} & \mathbb{P} \left( \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i (\chi_t(\mathfrak{X}^i, \cdot) - \chi_t(\tilde{\mathfrak{X}}^i, \cdot)) \right\|_\infty > \frac{\eta}{2} \right) \\ & \leq \mathbb{P} \left( \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i \chi_t(\mathfrak{X}^i, \cdot) \right\|_\infty > \frac{\eta}{4} \right) + \mathbb{P} \left( \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i \chi_t(\tilde{\mathfrak{X}}^i, \cdot) \right\|_\infty > \frac{\eta}{4} \right) \\ & = 2\mathbb{P} \left( \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i \chi_t(\mathfrak{X}^i, \cdot) \right\|_\infty > \frac{\eta}{4} \right) \end{aligned}$$

and thus

$$\mathbb{P} \left( \|G^n(t, \cdot) - \tilde{G}^n(t, \cdot)\|_\infty > \frac{\eta}{2} \right) \leq 2\mathbb{P} \left( \left\| \frac{1}{n} \sum_{i=1}^n \varepsilon_i \chi_t(\mathfrak{X}^i, \cdot) \right\|_\infty > \frac{\eta}{4} \right). \quad (\text{I.13})$$

*Step 3. Conditioning on  $\mathcal{X}$  and using Hoeffding's inequality.*

We have to bound the r.h.s. in [\(I.13\)](#). The main issue is that we have to estimate a supremum over  $\mathbb{R}$ . However, conditionally on  $\mathcal{X}$ , we actually take the supremum over a finite number of random variables (only the  $\varepsilon_i$  remain random), which enables us to use a union bound.

To see this, let us consider  $n$  realizations  $x^1, \dots, x^n$  of the BMPs  $\mathfrak{X}^1, \dots, \mathfrak{X}^n$ , with respective population sizes  $n_t^i$  at time  $t$ , and  $n_t = \sum_{i=1}^n n_t^i$  the total size of the population, so that particles present at time  $t$  are indexed by pairs  $(i, j)$  with  $1 \leq i \leq n$  and  $1 \leq j \leq n_t^i$ . Ordering particles according to their position at time  $t$  leads to an ordering of the pairs of the form  $(i_k, j_k)$ , for  $1 \leq k \leq n_t$ , such that the sequence  $(x_t^{i_k j_k})_{1 \leq k \leq n_t}$  is non-decreasing. Then, as  $r$  runs over the real line,  $\sum_{i=1}^n \varepsilon_i \chi_t(x^i, r) = \sum_{k=1}^{n_t} \varepsilon_{i_k} \mathbf{1}_{\{x_t^{i_k j_k} > r, \tau_b^{i_k j_k} > t\}}$  takes at most  $n_t + 1$  values, because each indicator function can change its value at most once (and it may never do, if the crossing time of the boundary is less than  $t$ ). Thus taking the supremum over  $\mathbb{R}$  of the values of  $\sum_{i=1}^n \varepsilon_i \chi_t(x^i, r)$  is in fact taking the maximum over  $n_t + 1$  random variables. We obtain that the event on the r.h.s. of [\(I.13\)](#), conditionally on  $\mathcal{X}$ , is a finite union, and we dominate its conditional probability by the number of events in the union times the maximum probability of one of these events, which can again be written as a supremum over  $\mathbb{R}$ :

$$\mathbb{P} \left( \sup_{r \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i \chi_t(X^i, r) \right| > \frac{\eta}{4} \mid \mathcal{X} \right) \leq (n_t + 1) \sup_{r \in \mathbb{R}} \mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i \chi_t(X^i, r) \right| > \frac{\eta}{4} \mid \mathcal{X} \right), \quad (\text{I.14})$$

where  $\mathcal{N}_t = \sum_{i=1}^n \mathcal{N}_t^i$ . Now we use the Hoeffding inequality for sums of independent bounded random variables (see for instance Theorem 8.1 in [DGL96]) to bound the probability in the r.h.s. of (I.14), uniformly with respect to  $r$ : conditionally on  $\mathcal{X}$ ,  $\varepsilon_i \chi_t(X^i, r)$  is bounded by  $\pm \mathcal{N}_t^i$  and centered, hence, setting  $A_t := \exp\left(-\frac{n^2 \eta^2}{32 \sum_{i=1}^n (\mathcal{N}_t^i)^2}\right)$ , we have that

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n \varepsilon_i \chi_t(X^i, r)\right| > \frac{\eta}{4} \mid \mathcal{X}\right) \leq 2A_t.$$

Now since the bound is uniform with respect to  $r$ , (I.14) leads to

$$\mathbb{P}\left(\sup_{r \in \mathbb{R}} \left|\frac{1}{n} \sum_{i=1}^n \varepsilon_i \chi_t(X^i, r)\right| > \frac{\eta}{4} \mid \mathcal{X}\right) \leq 2(\mathcal{N}_t + 1)A_t. \quad (\text{I.15})$$

*Step 4. Final domination.*

Thanks to steps 1 to 3, we obtain by concatenating (I.10), (I.13) and (I.15)

$$\mathbb{P}(\|G^n(t, \cdot) - G(t, \cdot)\|_\infty > \eta) \leq 8\mathbb{E}((\mathcal{N}_t + 1)A_t). \quad (\text{I.16})$$

Now we fix  $K > 0$  and split the r.h.s. depending on whether all  $\mathcal{N}_t^i$  are smaller than  $K$  or not.

- On  $\{\mathcal{N}_t^i \leq K, \forall i\}$ , then  $A_t$  is smaller than  $e^{-\frac{n\eta^2}{32K^2}}$  and  $\mathcal{N}_t$  smaller than  $nK$ , so that

$$\mathbb{E}\left((\mathcal{N}_t + 1)A_t \mathbf{1}_{\{\mathcal{N}_t^i \leq K, \forall i\}}\right) \leq (nK + 1)e^{-\frac{n\eta^2}{32K^2}}. \quad (\text{I.17})$$

- For the complement, we use the union bound and get

$$\mathbb{E}\left((\mathcal{N}_t + 1)A_t \mathbf{1}_{\{\mathcal{N}_t^i \leq K, \forall i\}^c}\right) \leq \mathbb{E}\left(\left(\sum_{i=1}^n \mathcal{N}_t^i + 1\right) A_t \sum_{i=1}^n \mathbf{1}_{\{\mathcal{N}_t^i > K\}}\right),$$

then expand the sums and use independence: the  $\mathcal{N}_t^i$  are i.i.d. random variables with geometric distribution and  $A_t$  is smaller than 1, so we obtain

$$\begin{aligned} & \mathbb{E}\left(\left(\sum_{i=1}^n \mathcal{N}_t^i + 1\right) A_t \sum_{i=1}^n \mathbf{1}_{\{\mathcal{N}_t^i > K\}}\right) \\ & \leq n\mathbb{P}(\Gamma > K) + n\mathbb{E}(\Gamma \mathbf{1}_{\{\Gamma > K\}}) + n(n-1)\mathbb{E}(\Gamma)\mathbb{P}(\Gamma > K), \end{aligned} \quad (\text{I.18})$$

where  $\Gamma$  has geometric distribution with parameter  $e^{-t}$ . Finally we use Lemma B.1 to get

$$\mathbb{E}\left((\mathcal{N}_t + 1)A_t \mathbf{1}_{\{\mathcal{N}_t^i \leq K, \forall i\}^c}\right) \leq n e^{-c(t)K} + n(K+2)e^t e^{-c(t)K} + n(n-1)e^t e^{-c(t)K}, \quad (\text{I.19})$$

with  $c(t) = -\ln(1 - e^{-t}) > 0$ .

We now take  $K = n^\alpha$  with  $\alpha \in (0, \frac{1}{2})$ , and use that  $c(t) \geq c_3 := c(T)$  and  $e^t \leq e^T$  to get the following bound (which is uniform with respect to  $t$ ): then (I.19) gives for all  $n$

$$\mathbb{E}\left((\mathcal{N}_t + 1)A_t \mathbf{1}_{\{\mathcal{N}_t^i \leq K, \forall i\}^c}\right) \leq (n + n(n^\alpha + 2)e^T + n(n-1)e^T) e^{-c_3 n^\alpha}. \quad (\text{I.20})$$

Summing up step 4, we take [\(I.16\)](#), [\(I.17\)](#) and [\(I.20\)](#) together to obtain

$$\mathbb{P}(\|G^m(t, \cdot) - G(t, \cdot)\|_\infty > \eta) \leq 8(n^{1+\alpha} + 1) e^{-\frac{n^{1-2\alpha}\eta^2}{32}} + 8(n + n(n^\alpha + 2)) e^T + n(n-1) e^T e^{-c_3 n^\alpha}.$$

Now, as soon as  $n \geq 2$  (which is implied by the condition on  $n$  obtained in step 1 when  $\eta$  is small anyway), we can simplify this bound using  $n^{1+\alpha} + 1 \leq 2n^2$  and  $n + n(n^\alpha + 2) e^T + n(n-1) e^T \leq 4n^2 e^T$ , and we obtain [\(I.8\)](#).  $\square$

*Proof of [Corollary I.1](#).* Notice that almost surely

$$\lim_{r \rightarrow -\infty} G^m(t, r) = \frac{N_t}{n}$$

while  $G(t, r)$  goes to 1 as  $r$  goes to  $-\infty$ , so that

$$\mathbb{P}(|N_t - n| > n\eta) \leq \mathbb{P}(\|G^m(t, \cdot) - G(t, \cdot)\|_\infty > \eta),$$

and we now deduce [\(I.9\)](#) from [\(I.8\)](#).  $\square$

## I.2.2 Application to $\mathcal{X}^+$ and $\mathcal{X}^-$

Fix  $\delta \in (0, \frac{1}{2})$  and consider two  $b$ -BMPs  $\mathcal{X}^{\pm, \delta, N}$  respectively starting from  $N_\delta^+ := \lceil N(1 + \delta) \rceil$  and  $N_\delta^- := \lfloor N(1 - \delta) \rfloor$  i.i.d. positions with distribution  $\mu_0$ . We will drop the dependence of the processes in  $\delta$  and  $N$  in the notation from now, and simply write  $\mathcal{X}^+$  and  $\mathcal{X}^-$ . We will refer to these processes as *upper* and *lower stochastic barrier*: we use the same terminology as in the related works even though the construction is different because these processes play a similar role in the proof.

In [Proposition I.2](#) below, we specialize the result of [Proposition I.1](#) to these stochastic barriers. Let us define  $G^{\pm, \delta, N}(t, r) = \frac{1}{N} \sum_{x \in \mathcal{X}_t^\pm} \mathbf{1}_{\{x > r\}}$  (note that we normalize by  $N$  instead of  $N^\pm$ ).

**Proposition I.2.** *For all  $\delta, \alpha$  in  $(0, \frac{1}{2})$  and  $\eta$  in  $(0, 1)$  there exist an explicit integer  $\mathcal{N}_1 = \mathcal{N}_1(\eta, \delta, T)$  such that for all  $t$  in  $[0, T]$  and  $N \geq \mathcal{N}_1$ , we have*

$$\mathbb{P}\left(\left\|G^{\pm, \delta, N}(t, \cdot) - (1 \pm \delta)G(t, \cdot)\right\|_\infty > \eta\right) \leq 64N^2 e^{-\frac{\eta^2(\frac{1}{2}-\delta)^{1-2\alpha}N^{1-2\alpha}}{512}} + 128N^2 e^{-c_3(\frac{1}{2}-\delta)^\alpha N^\alpha}. \quad (\text{I.21})$$

*Proof of [Proposition I.2](#).* By definition, we have

$$\mathbb{P}\left(\left\|G^{\pm, \delta, N}(t, \cdot) - (1 \pm \delta)G(t, \cdot)\right\|_\infty > \eta\right) = \mathbb{P}\left(\left\|G^{N_\delta^\pm}(t, \cdot) - \frac{N(1 \pm \delta)}{N_\delta^\pm}G(t, \cdot)\right\|_\infty > \frac{N}{N_\delta^\pm}\eta\right),$$

where  $G^{N_\delta^\pm} = \frac{N}{N_\delta^\pm} G^{\pm, \delta, N}$  is the complement of the empirical c.d.f. of a  $b$ -BMP starting with  $n = N_\delta^\pm$  particles, as in [Proposition I.1](#). Now we have for  $N \geq \frac{2}{\eta}$

$$\left|\frac{N(1 \pm \delta)}{N_\delta^\pm} - 1\right| \leq \frac{N}{2N_\delta^\pm}\eta,$$

and since  $G$  is bounded by 1, we obtain

$$\left\|\frac{N(1 \pm \delta)}{N_\delta^\pm}G(t, \cdot) - G(t, \cdot)\right\|_\infty \leq \frac{N}{2N_\delta^\pm}\eta,$$

and hence

$$\mathbb{P} \left( \left\| G^{\pm, \delta, N}(t, \cdot) - (1 \pm \delta)G(t, \cdot) \right\|_{\infty} > \eta \right) \leq \mathbb{P} \left( \left\| G^{N_{\delta}^{\pm}}(t, \cdot) - G(t, \cdot) \right\|_{\infty} > \frac{N}{2N_{\delta}^{\pm}} \eta \right).$$

We will now use [Proposition I.1](#): let  $\eta' = \frac{N}{2N_{\delta}^{\pm}} \eta$ , so that the condition on  $N_{\delta}^{\pm}$  is that  $N_{\delta}^{\pm} \geq \max \left( \frac{8C(T)}{\eta'^2}, \frac{2}{\eta}, 2 \right)$ . We can choose  $\eta < 1$ , so that the second condition implies the third one. Now, since  $N_{\delta}^{\pm}$  is always between  $N(1 - \delta) - 1$  and  $N(1 + \delta) + 1$ , the second condition is achieved as soon as  $N \geq \frac{2}{1 - \delta}$ . Furthermore, for  $N \geq 2$ ,  $\frac{N_{\delta}^{\pm}}{N}$  is between  $\frac{1}{2} - \delta$  and  $\frac{3}{2} + \delta$ , so the first condition is

$$N_{\delta}^{\pm} \geq \frac{32C(T)}{\eta^2} \left( \frac{3}{2} + \delta \right) \quad i.e. \quad N \geq \frac{32C(T) \left( \frac{3}{2} + \delta \right) + 1}{1 - \delta}.$$

We can now write

$$\mathcal{N}_1 = \mathcal{N}_1(\delta, \eta, T) = \max \left( \frac{\frac{2}{\eta} + 1}{1 - \delta}, \frac{32C(T) \left( \frac{3}{2} + \delta \right) + 1}{1 - \delta} \right),$$

so that for  $N \geq \mathcal{N}_1$ , [\(I.8\)](#) in [Proposition I.1](#) gives for  $\alpha$  in  $(0, \frac{1}{2})$

$$\mathbb{P} \left( \left\| G^{\pm, \delta, N}(t, \cdot) - (1 \pm \delta)G(t, \cdot) \right\|_{\infty} > \eta \right) \leq 16(N_{\delta}^{\pm})^2 e^{-\frac{\eta^2(N_{\delta}^{\pm})^{1-2\alpha}}{32}} + 32(N_{\delta}^{\pm})^2 e^{-c_3(N_{\delta}^{\pm})^{\alpha}},$$

which can be rewritten [\(I.21\)](#)

$$\mathbb{P} \left( \left\| G^{\pm, \delta, N}(t, \cdot) - (1 \pm \delta)G(t, \cdot) \right\|_{\infty} > \eta \right) \leq 64N^2 e^{-\frac{\eta^2 \left( \frac{1}{2} - \delta \right)^{1-2\alpha} N^{1-2\alpha}}{512}} + 128N^2 e^{-c_3 \left( \frac{1}{2} - \delta \right)^{\alpha} N^{\alpha}}$$

as desired using again the inequality  $\frac{1}{2} - \delta \leq \frac{N_{\delta}^{\pm}}{N} \leq \frac{3}{2} + \delta \leq 2$ .  $\square$

As explained in the introduction, we expect that, for large  $N$ ,  $\mathcal{X}^+$  (resp.  $\mathcal{X}^-$ ) will have more (resp. less) than  $N$  particles with high probability on any compact time interval. We thus define the favorable events

$$\begin{cases} \Gamma_t^+ = \Gamma_t^{+, \delta, N} = \{ \# \mathcal{X}_s^+ \geq N, \forall s \leq t \}, \\ \Gamma_t^- = \Gamma_t^{-, \delta, N} = \{ \# \mathcal{X}_s^- \leq N, \forall s \leq t \}. \end{cases}$$

The following proposition provides the control we need on the probability of these events for large  $N$ , when  $\delta$  goes to zero as a negative power of  $N$  (this is the case we consider in subsequent proofs [I.3](#)).

**Proposition I.3.** *Take any  $\beta \in (0, \frac{1}{2})$  and set  $\delta = \frac{N^{-\beta}}{2}$ . Then there exist positive constants  $c_4, c_5$  depending only on  $T$  and  $\beta$  such that for all  $N \geq 2$ , we have*

$$\mathbb{P}((\Gamma_T^{\pm})^c) \leq c_5 e^{-c_4 N^{\frac{1-2\beta}{3}}}. \quad (I.22)$$

<sup>13</sup>We chose to display the bounds as in [\(I.22\)](#) by taking  $\delta = \frac{N^{-\beta}}{2}$  for the sake of clarity, and because this is the form that we need to prove the other results. In fact, the proof leads to the following more explicit although more complicated bound, for any fixed  $\delta \in (0, \frac{1}{2})$ . For instance we obtain

$$\mathbb{P}((\Gamma_T^+)^c) \leq 3N e^{-c_3 N^{\alpha}} + 3N^{1+\alpha} \left( e^{-c_3 \left( \frac{N\delta}{2} + 1 \right)} + 36N^2 e^{-\frac{\delta^2 N^{1-2\alpha}}{128(1+\delta)^2}} + 72e^T N^2 e^{-c_3 N^{\alpha}} \right)$$

for any  $N$  greater than a certain  $\mathcal{N}_{2,+}(\delta, T)$  explicitly given in the proof of [Proposition I.3](#)

We now prove [Proposition I.3](#). Thanks to [Corollary I.1](#), we control the size of the stochastic barriers at any fixed time, but the favorable events  $\Gamma_T^\pm$  depend on the whole dynamic of the barriers between 0 and  $T$ , and not only on the position at one given time. The proof of [Proposition I.3](#) relies on the use of [Corollary I.1](#) on a finite (though random) set of times. Indeed, to check if  $\Gamma_T^\pm$  is realized, we only need to check if the cardinalities are big or small enough at the instants when they may change, meaning when a particle appears or disappears. Even better, for  $\Gamma_T^+$ , we only need to check death events: if  $N_t \geq N$  for some time  $t$ , then this will at least stay true until the next death event, because the number of particles will not decrease until then. Similarly, for  $\Gamma_T^-$ , we only check branching events, because  $\Gamma_T^-$  can only be invalidated at an instant when the number of particles increases. In the following, we make detailed proof for the bound on the probability of  $\Gamma_T^+$  and quickly explain what happens for  $\Gamma_T^-$ .

*Proof of [Proposition I.3](#).* We assume that our  $b$ -BMP is obtained by joining  $N_\delta^+$  i.i.d.  $b$ -BMPs obtained from  $N_\delta^+$  BMPs  $\mathfrak{X}^1, \dots, \mathfrak{X}^{N_\delta^+}$  (which we call families) from which particles are removed as soon as they cross the barrier  $b$ . We denote by  $N_t^i$  the population size at time  $t$  of the  $i$ -th  $b$ -BMP, and let  $N_t = \sum_{i=1}^{N_\delta^+} N_t^i$ , so that  $\#\mathcal{X}_t^+ = N_t$ . Then, for  $1 \leq i \leq N_\delta^+$  and  $\ell \geq 1$ , define  $\sigma_{i\ell}$  as the  $\ell$ -th instant of crossing of the  $i$ -th family.

Our goal is then to study  $\mathbb{P}(N_{\sigma_{i\ell}} < N, \sigma_{i\ell} \leq T)$  for all  $i, \ell$ . Let  $A_{i\ell} = \{\sigma_{i\ell} \leq T\}$ . Here we use the fact that the branching families are independent, so we can separate the  $i$ -th family from the others: by the union bound, we have

$$\begin{aligned} \mathbb{P}(N_{\sigma_{i\ell}} < N, A_{i\ell}) &\leq \mathbb{P}(|N_{\sigma_{i\ell}} - N_\delta^+| > N\delta, A_{i\ell}) \\ &\leq \mathbb{P}\left(\left|\sum_{k \neq i} N_{\sigma_{i\ell}}^k - (N_\delta^+ - 1)\right| > \frac{N\delta}{2}, A_{i\ell}\right) + \mathbb{P}(|N_{\sigma_{i\ell}}^i - 1| > \frac{N\delta}{2}, A_{i\ell}). \end{aligned} \quad (\text{I.23})$$

For the second term in the r.h.s. of [\(I.23\)](#), we note that, on  $A_{i\ell}$ , one has that  $N_{\sigma_{i\ell}}^i \leq \mathcal{N}_T^i$ , and that  $\mathcal{N}_T^i$  follows a geometric distribution with parameter  $e^{-T}$ , so that we have for  $N \geq \frac{2}{\delta}$

$$\begin{aligned} \mathbb{P}(|N_{\sigma_{i\ell}}^i - 1| > \frac{N\delta}{2}, A_{i\ell}) &= \mathbb{P}(N_{\sigma_{i\ell}}^i > \frac{N\delta}{2} + 1, A_{i\ell}) \\ &\leq \mathbb{P}(\mathcal{N}_T^i > \frac{N\delta}{2} + 1) \\ &\leq e^{-c_3(\frac{N\delta}{2} + 1)}, \end{aligned} \quad (\text{I.24})$$

where  $c_3 = -\ln(1 - e^{-T})$ . For the first term, take  $n = N_\delta^+ - 1$ ,  $\eta = \frac{N\delta}{2} (N_\delta^+ - 1)^{-1}$  and  $\eta' = \frac{\delta}{2(1+\delta)}$ , so that

$$n\eta = \frac{N\delta}{2} \quad \text{and} \quad \eta \geq \eta'.$$

We can thus rewrite

$$\mathbb{P}\left(\left|\sum_{k \neq i} N_{\sigma_{i\ell}}^k - (N_\delta^+ - 1)\right| > \frac{N\delta}{2}, A_{i\ell}\right) = \mathbb{P}\left(\left|\sum_{k \neq i} N_{\sigma_{i\ell}}^k - n\right| > n\eta, A_{i\ell}\right).$$

Then, since for  $k \neq i$  the  $\sigma_{i\ell}$  are independent from the  $k$ -th family's dynamic, we can disintegrate with respect to the value of  $\sigma_{i\ell}$ :

$$\mathbb{P}\left(\left|\sum_{k \neq i} N_{\sigma_{i\ell}}^k - n\right| > n\eta, A_{i\ell}\right) = \int_0^T \mathbb{P}\left(\left|\sum_{k \neq i} N_t^k - n\right| > n\eta\right) h_{\sigma_{i\ell}}(dt), \quad (\text{I.25})$$

where  $h_{\sigma_{i\ell}}$  is the distribution of  $\sigma_{i\ell}$ . Now since  $\eta \geq \eta'$ , we have for all  $t$

$$\mathbb{P}\left(\left|\sum_{k \neq i} N_t^k - n\right| > n\eta\right) \leq \mathbb{P}\left(\left|\sum_{k \neq i} N_t^k - n\right| > n\eta'\right), \quad (\text{I.26})$$

and this new integrand is exactly the quantity that appears in [Corollary I.1](#), which we can apply provided that  $N_\delta^+ - 1$  is greater than  $\mathcal{N}_0(\eta', t)$ , for all  $t$  in  $[0, T]$ . Since  $\mathcal{N}_0(\eta', t)$  increases with  $t$ , we only need to have  $N_\delta^+ \geq 3$  and  $N_\delta^+ \geq \frac{8C(T)}{\eta'^2} + 1$ . Thus, if we take  $N \geq \mathcal{N}_{2,+} = \mathcal{N}_{2,+}(\delta, T)$ , where  $\mathcal{N}_{2,+}(\delta, T) := \max\left(\frac{32(1+\delta)^2 C(T)}{\delta^2} + 1, \frac{3}{1+\delta}\right)$ , we can apply [Corollary I.1](#) and we have

$$\mathbb{P}\left(\left|\sum_{k \neq i} N_t^k - n\right| > n\eta'\right) \leq 16(N_\delta^+ - 1)^2 e^{-\frac{(N_\delta^+ - 1)^{1-2\alpha} \delta^2}{128(1+\delta)^2}} + 32 e^T (N_\delta^+ - 1)^2 e^{-c_3(N_\delta^+ - 1)^\alpha}.$$

Now if  $N \geq \mathcal{N}_{2,+}$ , we have  $N \leq N_\delta^+ - 1 \leq N(1 + \delta)$ , so we obtain

$$\mathbb{P}\left(\left|\sum_{k \neq i} N_t^k - n\right| > n\eta'\right) \leq 16(1 + \delta)^2 N^2 e^{-\frac{\delta^2 N^{1-2\alpha}}{128(1+\delta)^2}} + 32 e^T (1 + \delta)^2 N^2 e^{-c_3 N^\alpha}.$$

This bound is uniform with respect to  $t \in [0, T]$ , so we can also use it to bound the l.h.s. of [\(I.25\)](#). In view of [\(I.23\)](#) and [\(I.24\)](#), we have shown for  $N \geq \mathcal{N}_{2,+}(\delta, T)$ , that

$$\mathbb{P}(N_{\sigma_{i\ell}} < N, A_{i\ell}) \leq e^{-c_3(\frac{N\delta}{2} + 1)} + 16(1 + \delta)^2 N^2 e^{-\frac{\delta^2 N^{1-2\alpha}}{128(1+\delta)^2}} + 32 e^T (1 + \delta)^2 N^2 e^{-c_3 N^\alpha}.$$

This can be simplified for  $N$  large enough as

$$\mathbb{P}(N_{\sigma_{i\ell}} < N, A_{i\ell}) \leq c_5 e^{-c_4 N^{\frac{1-2\beta}{3}}} \quad (\text{I.27})$$

with  $c_4, c_5 > 0$  depending only on  $\beta$  and  $T$  if we recall  $\delta = \frac{N^{-\beta}}{2}$ , absorb the polynomials into the decreasing exponentials and take in them the best compromise  $\alpha = \frac{1-2\beta}{3}$  for the powers of  $N$ .

Now since  $N_t$  can decrease only when a particle is removed from the  $b$ -BMP, and that the time of such a removal must be one of the  $\sigma_{i\ell}$  for some  $1 \leq i \leq N_\delta^+$  and  $\ell \geq 1$ , we have the inclusion

$$(\Gamma_T^+)^c \subset \bigcup_{\substack{1 \leq i \leq N_\delta^+ \\ \ell \geq 1}} \{N_{\sigma_{i\ell}} < N, A_{i\ell}\}.$$

Noting that the number of crossings in the  $i$ -th family up to time  $t$  is bounded above by  $\mathcal{N}_T^i$ , we have, for any integer  $K \geq 1$ , that  $\{\mathcal{N}_T^i \leq K\} \cap \bigcup_{\ell \geq 1} \{N_{\sigma_{i\ell}} < N, A_{i\ell}\} \subset \bigcup_{\ell=1}^K \{N_{\sigma_{i\ell}} < N, A_{i\ell}\}$ . By the union bound and [\(I.27\)](#) we thus have that

$$\mathbb{P}((\Gamma_T^+)^c) \leq \sum_{i=1}^{N_\delta^+} \left( \mathbb{P}(\mathcal{N}_T^i > K) + \sum_{\ell=1}^K \mathbb{P}(N_{\sigma_{i\ell}} < N, A_{i\ell}) \right) \leq N_\delta^+ \left( e^{-c_3 K} + K c_5 e^{-c_4 N^{\frac{1-2\beta}{3}}} \right).$$

Finally we take  $K = \lceil N^{\frac{1-2\beta}{3}} \rceil$ , use  $N_\delta^+ \leq N(1 + \delta) + 1 \leq 2(1 + \delta)N \leq 3N$  to obtain for all  $N \geq 2$  (up to changing the constants  $c_4, c_5 > 0$ )

$$\mathbb{P}((\Gamma_T^+)^c) \leq c_5 e^{-c_4 N^{\frac{1-2\beta}{3}}}. \quad (\text{I.22})$$

As we mentioned earlier, everything works the same way for  $\Gamma_T^-$ , except that we work with branching times instead of death times (which does not modify the argument because we bounded the number of death times by the number of branching times anyway). We let the reader check that the precise bound we could obtain is

$$\begin{aligned} \mathbb{P}((\Gamma_T^-)^c) &\leq N e^{-c_3 N^\alpha} + N^{1+\alpha} \left( e^{-c_3 \left(\frac{N\delta}{2} + 1\right)} \right. \\ &\quad \left. + 16N^2 e^{-\frac{(1-\delta)^{-1-2\alpha} \delta^2 N^{1-2\alpha}}{128 \cdot 2^{1-2\alpha}}} + 32 e^T N^2 e^{-c_3 \frac{(1-\delta)^\alpha N^\alpha}{2\alpha}} \right), \end{aligned} \quad (\text{I.28})$$

true as soon as  $N \geq \mathcal{N}_{2,-}(\delta, T) = \max\left(\frac{32(1-\delta)^2 C(T)}{\delta^2} + 1, \frac{4}{1-\delta}\right)$ , which again, up to modifying the constants, leads for all  $N \geq 2$  to

$$\mathbb{P}((\Gamma_T^-)^c) \leq c_5 e^{-c_4 N^{\frac{1-2\beta}{3}}}. \quad (\text{I.22})$$

□

### I.3 Hydrodynamic limit of the $N$ -BMP

The key ingredient to prove [Theorem I.1](#) is the definition of a suitable coupling between the  $N$ -BMP and the  $b$ -BMPs  $\mathcal{X}^+$  and  $\mathcal{X}^-$  studied in [Section I.2.2](#). Recall the definitions of the favorable events:

$$\begin{cases} \Gamma_T^+ = \{\#\mathcal{X}_s^+ \geq N, \forall s \leq T\} \\ \Gamma_T^- = \{\#\mathcal{X}_s^- \leq N, \forall s \leq T\} \end{cases}.$$

and of the ordering  $x \preceq y$  between two real-valued tuples  $x, y$  (possibly with different sizes):

$$\forall r \in \mathbb{R} \quad \#\{i, x^i \geq r\} \leq \#\{i, y^i \geq r\}.$$

**Proposition I.4.** *There exists a coupling of an  $N$ -BMP  $\mathcal{X}$  and its upper stochastic barrier  $\mathcal{X}^+$  (resp. lower stochastic barrier  $\mathcal{X}^-$ ) such that on  $\Gamma_T^+$  (resp.  $\Gamma_T^-$ ), we have  $\mathcal{X}_t \preceq \mathcal{X}_t^+$  (resp.  $\mathcal{X}_t^- \preceq \mathcal{X}_t$ ) for all  $t$  in  $[0, T]$ .*

The proof below contains the essential arguments needed to prove [Proposition I.4](#), but some of the more technical aspects related to the pasting of Markov processes in a general setting are deferred to [Appendix A](#), with a fully detailed technical construction of the coupling given in [Section A.2](#).

*Proof of [Proposition I.4](#).* A key ingredient in the proof is Theorem 1 in [\[BF23b\]](#), which states that, provided that the Markov process  $X$  satisfies properties (ii) (Feller semigroup) and (iii) (stochastic monotonicity), we can build a (Feller) Markov process  $(X^1, X^2)$  whose coordinates are versions of the process  $X$ , such that, starting from  $(X_0^1, X_0^2) = (x^1, x^2)$ , with  $x^1 \leq x^2$ , one has  $X_t^1 \leq X_t^2$  for every  $t \geq 0$ . We call this process the monotone pair coupling.

We now explain the construction of a coupling between  $\mathcal{X}$  and  $\mathcal{X}^+$ . As long as  $\#\mathcal{X}_t^+ \geq N$ , each particle in  $\mathcal{X}$  will be paired with a particle in  $\mathcal{X}^+$  in such a way that the former always lies below the latter, so that  $\mathcal{X}_t \preceq \mathcal{X}_t^+$ .

- Initial positions: Recall that  $\mathcal{X}$  (resp.  $\mathcal{X}^+$ ) starts with  $N$  (resp.  $N_\delta^+ = \lceil N(1 + \delta) \rceil$ ) particles whose initial positions form an i.i.d. family with common distribution  $\mu_0$ . Since  $N_\delta^+ \geq N$ , we can always couple the initial configurations in such a way that every particle in  $\mathcal{X}$  is paired with a particle in  $\mathcal{X}^+$  having the same initial position.

- Between branching/killing events: As long as no branching/killing event affecting  $\mathcal{X}$  or  $\mathcal{X}^+$  occurs, paired particles evolve according to the monotone pair coupling, so their ordering is preserved, while the remaining (unpaired) particles evolve independently.
- Branching events: Branching events affecting paired particles are synchronized, so that paired particles simultaneously branch at rate 1. Such an event leads to two newborn particles, which are then paired together: this is consistent with the ordering constraint since the positions of the paired parent particles are ordered at the time of birth. Note that the  $\mathcal{X}^+$  particle that was paired with the lowest particle in  $\mathcal{X}$  is now unpaired, because the latter has just been killed as a consequence of the branching event, but it is still true that each particle in  $\mathcal{X}$  is paired with a particle in  $\mathcal{X}^+$ . Finally, unpaired particles branch independently.
- Crossing events: Remember that particles in  $\mathcal{X}^+$  are killed when they cross the moving boundary  $b$ . As a consequence, when a paired particle in  $\mathcal{X}^+$  crosses the boundary, the particle in  $\mathcal{X}$  it was paired to must be given another partner. This is possible as long as  $\mathcal{X}^+$  contains at least  $N$  particles at the time of the crossing: by definition, each of these particles lies above the barrier, among which one at least is not yet paired, while the particle in  $\mathcal{X}$  that was paired to the particle that has just crossed, has to lie below the barrier; as a consequence, we can indeed find a partner in  $\mathcal{X}^+$  such that the ordering is preserved. In the case where  $\mathcal{X}^+$  contains strictly less than  $N$  particles at the time of the crossing, we cease to consider pairs of particles and let the processes  $\mathcal{X}$  and  $\mathcal{X}^+$  evolve independently.

We can make a similar construction of a coupling between  $\mathcal{X}$  and  $\mathcal{X}^-$ . This time, we ensure that, as long as  $\#\mathcal{X}_t^- \leq N$ , each particle in  $\mathcal{X}^-$  is paired with a particle in  $\mathcal{X}$  in such a way that the former always lie below the latter.

- Initial positions: We start the  $N$ -BMP from  $N$  initial positions independently distributed according to  $\mu_0$ ,  $N_\delta^-$  of which are shared between  $\mathcal{X}$  and  $\mathcal{X}^-$ .
- Between branching/killing events: Paired particles subsequently evolve using the monotone pair coupling, while unpaired particles evolve independently.
- Crossing events: When a paired particle in  $\mathcal{X}^-$  crosses the boundary, the particle in  $\mathcal{X}$  it was paired to stops being paired and starts evolving independently, and it is still true that each particle in  $\mathcal{X}^-$  is paired to a particle in  $\mathcal{X}$ .
- Branching events: Branching events affecting paired particles are synchronized, as in the previous coupling between  $\mathcal{X}$  and  $\mathcal{X}^+$ , and the two newborn particles are paired accordingly. In the case where a particle in  $\mathcal{X}^-$  was paired with the lowest particle in  $\mathcal{X}$ , we have to find it a new partner, since this particle is removed as a consequence of the branching event. This is possible as long as  $\mathcal{X}^-$  contains at most  $N$  particles at the time of the branching: in this case, at most  $N - 1$  particles in  $\mathcal{X}^-$  are paired with particles in  $\mathcal{X}$  just before the branching, so at least one particle in  $\mathcal{X}$  is not yet paired. This particle lies above the position of the (now defunct) lowest particle, while the particle in  $\mathcal{X}^-$  that was paired to the lowest particle lies below, so we can pair these particles together while preserving the order. In the case where  $\mathcal{X}^-$  contains strictly more than  $N$  particles, we cease to consider pairs of particles and let the processes  $\mathcal{X}$  and  $\mathcal{X}^-$  evolve independently.

Note that, since  $\tau_b$  has the exponential  $\mathcal{E}(1)$  distribution with respect to  $\mathbb{Q}_{\mu_0}$ ,  $\mathbb{P}$ -a.s. branching and crossing events always occur at distinct times, so we do not have to consider situations where these events occur simultaneously in the definition of the coupling.  $\square$



Now we can finally prove [Theorem I.1](#) by putting the different pieces together: on  $\Gamma_T^\pm$  – which has a very high probability – the coupling ensures an ordering of the empirical c.d.f., so that  $F^N(t, \cdot)$  is squeezed between two functions that converge to small perturbations of  $e^t U(t, \cdot)$ , again up to a small probability. Then taking the perturbation small enough, we obtain the limit for  $F^N(t, \cdot)$ .

*Proof of [Theorem I.1](#).* Fix  $\beta > 0$ , let  $\eta = N^{-\beta}$  and focus on

$$\mathbb{P}(\|F^N(t, \cdot) - e^t U(t, \cdot)\|_\infty > \eta) = \mathbb{P}(\|1 - F^N(t, \cdot) - G(t, \cdot)\|_\infty > \eta).$$

We first separate this event in such a way that we will be able to use the couplings: we have

$$\begin{aligned} \mathbb{P}(\|F^N(t, \cdot) - e^t U(t, \cdot)\|_\infty > \eta) &\leq \mathbb{P}(\exists r \in \mathbb{R}, 1 - F^N(t, r) - G(t, r) > \eta) \\ &\quad + \mathbb{P}(\exists r \in \mathbb{R}, 1 - F^N(t, r) - G(t, r) < -\eta). \end{aligned}$$

Now for every  $r \in \mathbb{R}, \delta > 0$  and  $\varepsilon > 0$ , the stochastic ordering  $\mathcal{X}_t \preceq \mathcal{X}_t^+$  implies that, on  $\Gamma_T^{+, \delta, N}$ ,

$$1 - F^N(t, r) \leq G^{+, \delta, N}(t, r),$$

so that if  $\|G^{+, \delta, N}(t, \cdot) - (1 + \delta)G(t, \cdot)\|_\infty \leq \varepsilon$ ,  $1 - F^N(t, r) \leq G(t, r)(1 + \delta) + \varepsilon$ , which we can rewrite as  $1 - F^N(t, r) - G(t, r) \leq \delta G(t, r) + \varepsilon \leq \delta + \varepsilon$ . Hence if we take  $\delta = \varepsilon = \frac{\eta}{2} = \frac{N^{-\beta}}{2}$ ,

$$\begin{aligned} \mathbb{P}(\exists r \in \mathbb{R}, 1 - F^N(t, r) - G(t, r) > \eta) &\leq \mathbb{P}\left((\Gamma_T^{+, \delta, N})^c\right) + \mathbb{P}\left(\{\exists r \in \mathbb{R}, 1 - F^N(t, r) - G(t, r) > \eta\} \cap \Gamma_T^{+, \delta, N}\right) \\ &\leq \mathbb{P}\left((\Gamma_T^{+, \delta, N})^c\right) + \mathbb{P}\left(\|G^{+, \delta, N}(t, \cdot) - (1 + \delta)G(t, \cdot)\|_\infty > \frac{\eta}{2}\right), \end{aligned}$$

and we can now use the results of [Propositions I.2](#) and [I.3](#) to bound this quantity: we have for all  $\alpha, \beta$  in  $(0, \frac{1}{2})$  and  $N$  large enough

$$\begin{aligned} \mathbb{P}(\exists r \in \mathbb{R}, 1 - F^N(t, r) - G(t, r) > \eta) &\leq c_5 e^{-c_4 N^{\frac{1-2\beta}{3}}} + 64N^2 e^{-\frac{\eta^2 (\frac{1-\eta}{2})^{1-2\alpha} N^{1-2\alpha}}{2048}} + 128N^2 e^{-c_3 (\frac{1-\eta}{2})^\alpha N^\alpha}, \end{aligned}$$

and we can again optimize the value of  $\alpha$  to obtain for some constants  $c_1, c_2 > 0$  and all  $N$

$$\mathbb{P}\left(\exists r \in \mathbb{R}, 1 - F^N(t, r) - G(t, r) > N^{-\beta}\right) \leq \frac{c_2}{2} e^{-c_1 N^{\frac{1-2\beta}{3}}}, \quad (\text{I.29})$$

The same reasoning with the other term, using the coupling with  $\mathcal{X}^-$ , gives

$$\begin{aligned} \mathbb{P}(\exists r \in \mathbb{R}, 1 - F^N(t, r) - G(t, r) < -\eta) &\leq \mathbb{P}\left((\Gamma_T^{-, \delta, N})^c\right) + \mathbb{P}\left(\|G^{-, \delta, N}(t, \cdot) - (1 - \delta)G(t, \cdot)\|_\infty > \frac{\eta}{2}\right), \end{aligned}$$

which is again bounded by exponentially decaying quantities thanks to [Propositions I.2](#) and [I.3](#). Putting these equations together we deduce that

$$\mathbb{P}\left(\|F^N(t, \cdot) - e^t U(t, \cdot)\|_\infty > N^{-\beta}\right) \leq c_2 e^{-c_1 N^{\frac{1-2\beta}{3}}},$$

which is exactly [\(I.1\)](#). Thanks to the Borel-Cantelli lemma, the almost sure convergence of  $F^N(t, \cdot)$  to  $e^t U(t, \cdot)$  in the supremum norm is a straightforward consequence of [\(I.1\)](#).  $\square$

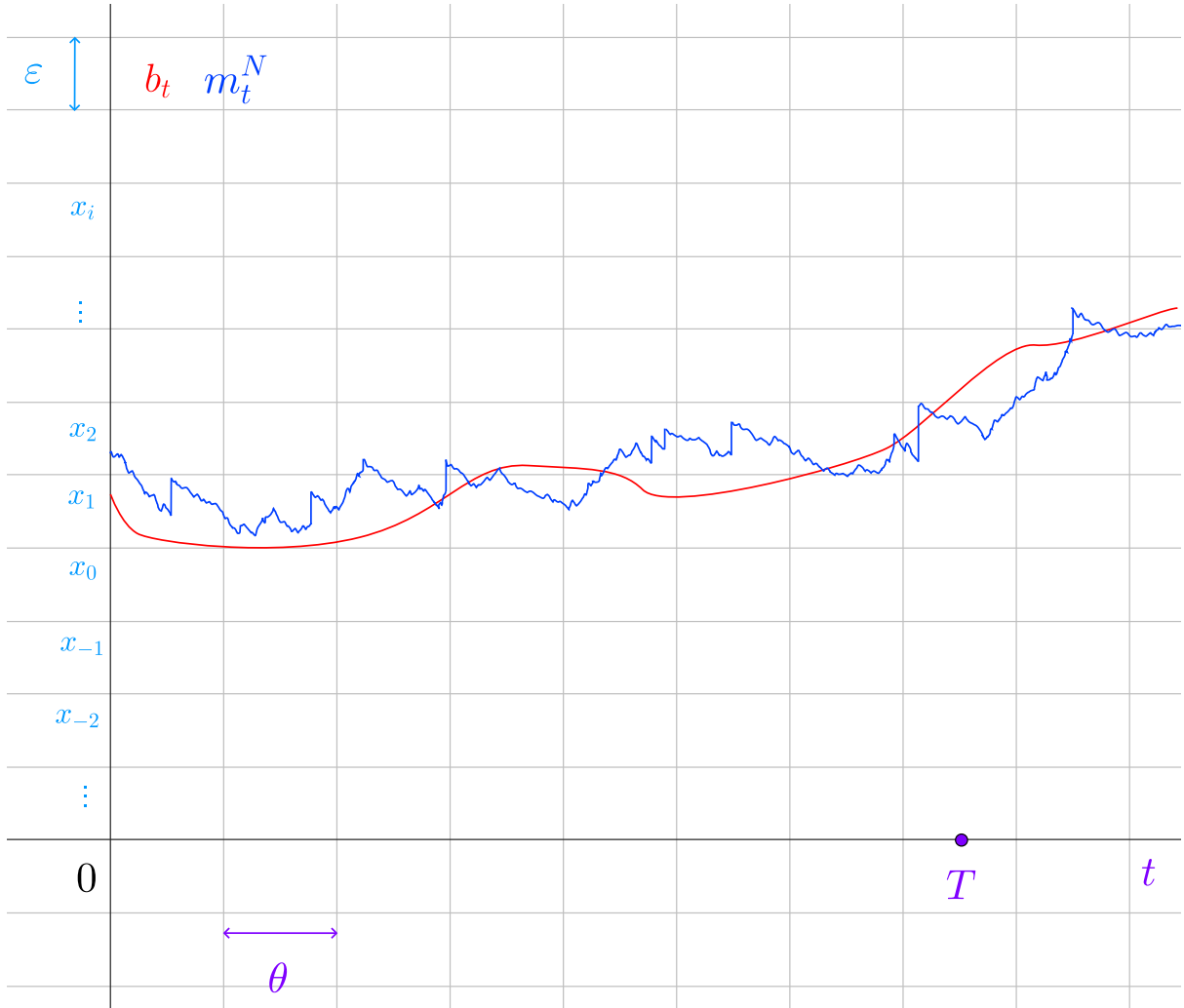


Figure I.1: A realization of  $m_t^N$  on the space-time grid described in the proof of [Theorem I.2](#)

#### I.4 Convergence of the minimum of the $N$ -BMP

We now turn to the proof of [Theorem I.2](#) about the convergence of  $m_t^N$ , which is the lowest particle position of the  $N$ -BMP at time  $t$ , towards  $b_t$ , when  $N$  goes to infinity. As mentioned in the introduction, our proof is similar in spirit to that of [BBNP22](#) (namely Section 4.2 about the proof of their Proposition 1.6). The idea is to combine the quantitative bound [\(I.1\)](#) of [Theorem I.1](#) together with Ray's estimate, in order to control the difference between  $b_t$  and  $m_t^N$  on a space-time grid (see [Figure I.1](#)). Berestycki, Brunet, Nolen and Penington combine a precise control on Brownian trajectories together with their hydrodynamic limit estimate (involving negative powers of  $N$ ) to prove convergence at any fixed time, while we exploit our improved hydrodynamic limit bound combined with a weaker but more generally valid control on the trajectories (provided by Ray's estimate) to obtain uniform convergence on a time interval. *Proof of [Theorem I.2](#).* We first give the proof assuming that  $t_0 > 0$ , and then explain the (minor) additional steps needed to deal with the case  $t_0 = 0$ .

*Step 1. Building a suitable space-time grid.*

Fix  $\varepsilon > 0$  and discretize space with step  $\varepsilon$ , starting from  $\min_{[t_0, T]} b$ , i.e. set for  $i \in \mathbb{Z}$

$$x_i = \min_{[t_0, T]} b + i\varepsilon.$$

Now discretize time with step  $\theta = 2N^{-\beta}$ , for a given  $\beta \in (0, \frac{1}{2})$ , starting from 0. Let  $s_0 \in (0, t_0]$ . Since  $b$  is uniformly continuous on  $[s_0, T]$ , we have, for all large enough  $N$  (depending on  $\varepsilon, b, s_0, T, \beta$ ) that  $|b_t - b_s| < \varepsilon$  for all  $s, t \in [s_0, T]$  such that  $|t - s| \leq \theta$ , and, from now on, we assume that this assumption holds. Also, the boundedness of  $b$  on  $[t_0, T]$  allows us to denote by  $K > 0$  the smallest integer such that  $x_K \geq \max_{[t_0, T]} b$ .

Thanks to Ray's estimate (Assumption (iv)), for every  $\kappa > 0$ , there is a constant  $C_\kappa$  (depending also on  $\varepsilon, t_0, T, b$ ) such that<sup>14</sup> for all  $-6 \leq i \leq K + 5$ , and all  $N \geq 1$ ,

$$\mathbb{Q}_{x_{i+1}}(\tau_{x_i} \leq \theta) \leq C_\kappa \theta^\kappa. \quad (\text{I.30})$$

A useful observation is that, due to stochastic monotonicity (Assumption (iii)), this also gives a bound on the probability for one particle to cross the interval  $[x_i, x_{i+1}]$  downwards in a short instant:

$$\forall x \geq x_{i+1}, \mathbb{Q}_x(\tau_{x_i} \leq \theta) \leq C_\kappa \theta^\kappa. \quad (\text{I.31})$$

To deduce (I.31) from (I.30), we invoke the monotone pair coupling already used in the proof of Proposition I.4, to show that  $\mathbb{Q}_x(\tau_{x_i} \leq \theta) \leq \mathbb{Q}_{x_{i+1}}(\tau_{x_i} \leq \theta)$  when  $x \geq x_{i+1}$ , and (I.31) ensues. Now, using the strong Markov property of  $X$ , it turns out that (I.31) can be strengthened into

$$\forall x, \mathbb{Q}_x(\exists s \in [0, \theta] \text{ such that } X_s \geq x_{i+1}, \tau_{x_i} \leq \theta) \leq C_\kappa \theta^\kappa. \quad (\text{I.32})$$

*Step 2. Lower bound at a fixed time.*

Our goal is to establish (I.44), showing that  $b_t$  is an approximate lower bound for  $m_t^N$  at every  $t = k\theta$  in  $(t_0 - \theta, T]$ , where  $k$  is an integer. Crucially, the various bounds we derive must hold uniformly with respect to  $k$ , so we systematically emphasize the dependency of the various constants with respect to the model and grid parameters in the sequel.

We assume that  $N$  is large enough (depending on  $s_0, t_0, \beta$ ) so that  $t_0 - 2\theta \geq s_0$ , and consider an integer  $k$  such that  $t := k\theta \in (t_0 - \theta, T]$ . For  $i \in \mathbb{Z}$ , introduce the numbers

$$r_i = \max\{x_j \text{ such that } x_j < b_t\} - i\varepsilon,$$

and note that  $r_0, \dots, r_5$  are elements of the space grid subject to (I.30). By definition we have  $b_t > r_0$ , and, since  $\theta$  is an  $\varepsilon$ -modulus of continuity for  $b$  on  $[s_0, T]$  and  $s_0 \leq t - \theta \leq t \leq T$ , we have that  $b_{t-\theta} > r_0 - \varepsilon = r_1$ .

Now let  $\eta = \frac{\theta}{2} = N^{-\beta}$ , and, before delving into the details, consider the following informal description of the behaviour of the  $N$ -BMP over the time-interval  $[t - \theta, t]$  (see Figure I.2).

- Since  $b_{t-\theta} > r_1$ , the hydrodynamic limit of the  $N$ -BMP (I.1) implies that with a high probability, fewer than  $N\eta$  particles lie below  $r_1$  at time  $t - \theta$ .
- Comparison with a BMP (without killing) then shows that, with high probability, between time  $t - \theta$  and  $t$ , the total number of particles in the  $N$ -BMP that are, or descend from, particles lying below  $r_1$  at time  $t - \theta$ , does not exceed  $\sim N\eta e^\theta$ .

<sup>14</sup>It is crucial to note that we invoke Ray's estimate for  $\mathbb{Q}_{x_{i+1}}(\tau_{x_i} \leq \theta)$  simultaneously over a **finite** set of indices  $i$ .

- With high probability, between  $t - \theta$  and  $t$  we have  $\sim N(e^\theta - 1)$  newborn particles in the  $N$ -BMP, so we kill  $\sim N(e^\theta - 1)$  particles.
- Thanks to Ray's estimate, with high probability, particles that did not lie below  $r_1$  at time  $t - \theta$ , do not lie below  $r_2$  at any time between  $t - \theta$  and  $t$ .
- As long as there are particles lying below  $r_2$ , the particles that are killed are among them (because we kill the lowest one).
- Since  $N(e^\theta - 1) \sim 2N\eta e^\theta$ , we deduce from the previous observations that, with high probability, there has to be a time within the interval  $[t - \theta, t]$  at which no particle is lying below  $r_2$ .
- Hence, thanks to Ray's estimate, with high probability there are no particles lying below  $r_3$  at time  $t$ .

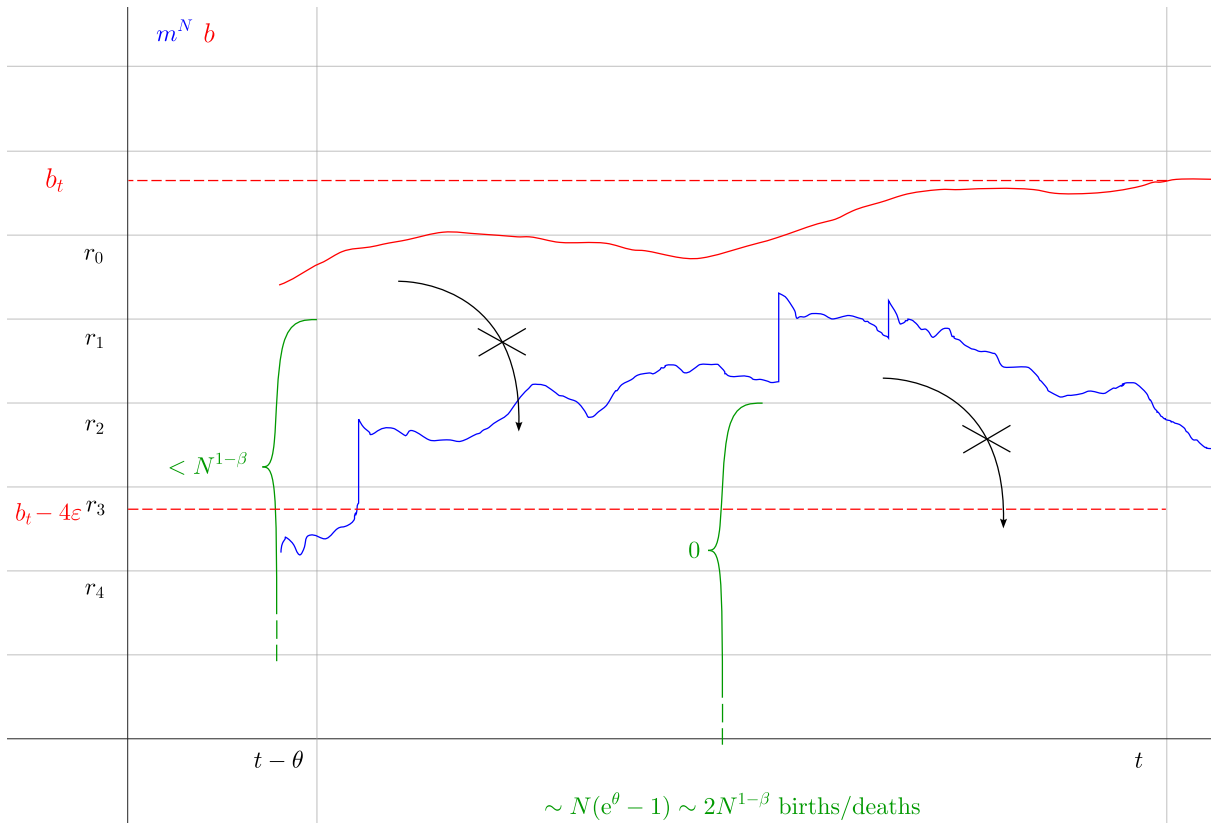


Figure I.2: Sketch of the second step of the proof of Theorem I.2

We now turn the previous informal description into precise arguments. Since  $b_{t-\theta} > r_1$ , we have that  $U(t - \theta, r_1) = 0$ , so that, by (I.1),

$$\mathbb{P}\left(|F^N(t - \theta, r_1)| > N^{-\beta}\right) \leq c_2 e^{-c_1 N^{\frac{1-2\beta}{3}}}.$$

Denoting by  $K_-$  the number of particles lying below  $r_1$  at time  $t - \theta$ , the above bound can be rewritten as

$$\mathbb{P}(K_- > N^{1-\beta}) \leq c_2 e^{-c_1 N^{\frac{1-2\beta}{3}}}. \quad (\text{I.33})$$

Now assume that, starting from time  $t - \theta$ , our  $N$ -BMP is built from a BMP (without killing), which is itself obtained by joining  $N$  independent BMPs (or "families"), each starting from a single particle. The  $N$ -BMP is obtained by removing particles (together with their descendants) from the BMP, according to the selection mechanism of the  $N$ -BMP.

The population of the BMP at time  $t$ , hereafter denoted  $\mathcal{N}_t$ , is the sum of  $N$  independent random variables  $G_i$ , each with geometric distribution  $\mathcal{G}(e^{-\theta})$ . For  $s$  in  $[t - \theta, t]$ , denote by  $X_s^{ij}$  the position of the  $j$ -th particle from the  $i$ -th family at time  $s$ , where  $1 \leq j \leq G_i$ . For  $x > y$ , let  $\Xi_{x,y}$  denote the event that at least one particle crosses downwards the interval  $[y, x]$  on the time interval  $[t - \theta, t]$ , i.e.

$$\Xi_{x,y} = \{\exists i \in \llbracket 1, N \rrbracket \exists j \in \llbracket 1, G_i \rrbracket, (X_{t-\theta}^{ij} \geq x \text{ and } \exists s \in [t - \theta, t] X_s^{ij} < y)\}.$$

Denote by  $\mathcal{N}'$  the population size at time  $t$  of the set of particles that lied below  $r_1$  at time  $t - \theta$  together with their descendants in the BMP, so that  $\mathcal{N}'$  is the sum of the  $K_-$  corresponding random variables  $G_i$ . Letting  $B = \{\forall s \in [t - \theta, t], m_s^N < r_2\}$ , we now claim that

$$B \cap \Xi_{r_1, r_2}^c \subset \{\mathcal{N}_t - N < \mathcal{N}'\}. \quad (\text{I.34})$$

Indeed, from the condition that the minimum particle position in the  $N$ -BMP is always less than  $r_2$  during the time interval  $[t - \theta, t]$ , we deduce that any particle of the  $N$ -BMP that is killed during this time interval has to lie below  $r_2$  at the time of its killing. From the condition that no BMP particle lying above  $r_1$  at time  $t - \theta$  or one of its descendants, ever goes below  $r_2$  during the time interval  $[t - \theta, t]$ , we conclude that every particle that is killed is taken among particles that were below  $r_1$  at time  $t - \theta$  and their descendants in the BMP. Now, the population of particles in the BMP at time  $t$  is divided into the  $N$  particles that survived the selection mechanism, and the  $\mathcal{N}_t - N$  particles (including their descendants) that have been killed. We have just seen that this last subset is contained into the subset of  $\mathcal{N}'$  particles (including their descendants in the BMP) that start below  $r_1$  at time  $t - \theta$ . Moreover, the inclusion has to be strict since there is at least one particle below  $r_2$  in the  $N$ -BMP at time  $t$ , which must be, or descend from, a particle that started below  $r_1$  at time  $t - \theta$ , and has survived selection. As a consequence, [\(I.34\)](#) is proved.

Now, using [\(I.34\)](#), we have that

$$\mathbb{P}(B) = \mathbb{P}(B \cap \Xi_{r_1, r_2}) + \mathbb{P}(B \cap \Xi_{r_1, r_2}^c) \leq \mathbb{P}(\Xi_{r_1, r_2}) + \mathbb{P}(\mathcal{N}_t - N < \mathcal{N}').$$

We note that, conditional upon the value of  $K_-$ , the random variable  $\mathcal{N}_t - \mathcal{N}'$  is the sum of  $N - K_-$  i.i.d.  $\mathcal{G}(e^{-\theta})$  random variables. On the event  $\{K_- \leq N^{1-\beta}\}$ ,  $\mathcal{N}_t - \mathcal{N}'$  is the sum of at least  $N - \lfloor N^{1-\beta} \rfloor$  such random variables, so that we have

$$\mathbb{P}(B) \leq \mathbb{P}(\Xi_{r_1, r_2}) + \mathbb{P}\left(\sum_{i=\lfloor N^{1-\beta} \rfloor + 1}^N G_i < N\right) + \mathbb{P}(K_- > N^{1-\beta}). \quad (\text{I.35})$$

We will now deal separately with each of these three terms.

- The rightmost term is dealt with using [\(I.33\)](#)
- The leftmost term is bounded thanks to Ray's estimate combined with the many-to-one lemma. Let us denote by  $\mathfrak{N}_{r_1, r_2}$  the number of particles that go from above  $r_1$  to below  $r_2$  in the time interval, namely

$$\mathfrak{N}_{r_1, r_2} = \#\{(i, j) \in \llbracket 1, N \rrbracket \times \mathbb{N}^*, j \leq G_i, X_{t-\theta}^{ij} \geq r_1 \text{ and } \exists s \in [t - \theta, t] X_s^{ij} < r_2\}, \quad (\text{I.36})$$

so that  $\Xi_{r_1, r_2} = \{\mathfrak{N}_{r_1, r_2} \geq 1\}$ . Thanks to Markov's inequality, we then have

$$\mathbb{P}(\Xi_{r_1, r_2}) \leq \mathbb{E}(\mathfrak{N}_{r_1, r_2}). \quad (\text{I.37})$$

We also have

$$\mathfrak{N}_{r_1, r_2} = \sum_{i=1}^N \sum_{j=1}^{G_i} \mathbf{1}_{\{X_{t-\theta}^{ij} \geq r_1, \exists s \in [t-\theta, t] \text{ s.t. } X_s^{ij} < r_2\}}. \quad (\text{I.38})$$

Now, if a particle is below  $r_1$  at time  $t - \theta$ , the whole sum for the corresponding index  $i$  is zero in the sum in (I.38); and for the families whose initial particle is above  $r_1$  at time  $t - \theta$ , the corresponding term is controlled by the many-to-one lemma (Lemma I.A): indeed, the family is a BMP starting with one particle at time  $t - \theta$ , so that we have at time  $t$

$$\mathbb{E} \left( \sum_{j=1}^{G_i} \mathbf{1}_{\{X_{t-\theta}^{ij} \geq r_1, \exists s \in [t-\theta, t] \text{ s.t. } X_s^{ij} < r_2\}} \middle| \mathcal{X}_{t-\theta} \right) = e^\theta \mathbb{Q}_{X_{t-\theta}^{i1}}(\tau_{r_2} \leq \theta) \mathbf{1}_{\{X_{t-\theta}^{i1} \geq r_1\}}.$$

Now this quantity is bounded by (I.31), and we have at most  $N$  such non-zero terms in the sum in the l.h.s. of (I.38), so that in the end we have

$$\mathbb{E}(\mathfrak{N}_{r_1, r_2}) \leq N e^\theta C_\kappa \theta^\kappa. \quad (\text{I.39})$$

Now recall (I.37) and that  $\theta = 2N^{-\beta}$ , so that we have  $e^\theta \leq 2$  for  $N$  large enough (depending on  $\beta$ ), and thus

$$\mathbb{P}(\Xi_{r_1, r_2}) \leq C'_\kappa N^{1-\beta\kappa}, \quad (\text{I.40})$$

for a certain constant  $C'_\kappa > 0$ .

- Finally we estimate the central term in (I.35),  $\mathbb{P} \left( \sum_{i=\lfloor N^{1-\beta} \rfloor + 1}^N G_i < N \right)$ , which is again a large deviation probability for a sum of i.i.d.  $\mathcal{G}(e^{-\theta})$  random variables. Indeed we have

$$\mathbb{P} \left( \sum_{i=\lfloor N^{1-\beta} \rfloor + 1}^N G_i < N \right) = \mathbb{P} \left( \frac{1}{N - \lfloor N^{1-\beta} \rfloor} \sum_{i=\lfloor N^{1-\beta} \rfloor + 1}^N G_i < \frac{1}{1 - N^{-\beta}} \right).$$

For  $\theta$  small enough, and thus  $N$  large enough (depending on  $\beta$ ),  $\frac{1}{1 - \frac{\theta}{2}} < e^\theta$ , so we can indeed use Cramér's theorem to study the large deviations below the mean for this sum of random variables: Lemma B.1 (iv) ensures that

$$\mathbb{P} \left( \sum_{i=\lfloor N^{1-\beta} \rfloor + 1}^N G_i < N \right) \leq e^{-(N - N^{1-\beta}) \Lambda_{e^{-\theta}}^* \left( \frac{1}{1 - N^{-\beta}} \right)}. \quad (\text{I.41})$$

Since  $N^{-\beta} = \theta/2$ , the expansion (B.2) yields that  $\Lambda_{e^{-\theta}}^* \left( \frac{1}{1 - N^{-\beta}} \right) = (1 - \ln(2))N^{-\beta} + o(N^{-\beta})$ , so that, choosing a constant  $c_6$  such that  $0 < c_6 < 1 - \ln(2)$ , (I.41) implies that, for  $N$  large enough (depending on  $\beta$ ):

$$\mathbb{P} \left( \sum_{i=\lfloor N^{1-\beta} \rfloor + 1}^N G_i < N \right) \leq e^{-c_6 N^{1-\beta}}.$$

Looking back at [\(I.35\)](#), we have obtained the bound, valid for  $N$  large enough (depending on  $\beta$ ):

$$\mathbb{P}(B) \leq C'_\kappa N^{1-\beta\kappa} + e^{-c_6 N^{1-\beta}} + c_2 e^{-c_1 N^{\frac{1-2\beta}{3}}}. \quad (\text{I.42})$$

Now observe that, on the event  $B^c$ , there must be a time between  $t - \theta$  and  $t$  at which no particle in the  $N$ -BMP lies below  $r_2$ . If, in addition, between  $t - \theta$  and  $t$ , no particle is allowed to go below  $r_3$  once it has gone above  $r_2$ , one must then have that  $m_t^N \geq r_3$ . Thus, using [\(I.32\)](#) and arguing in exactly the same way as we did to obtain the bound [\(I.40\)](#) on the probability of  $\mathbb{P}(\Xi_{r_1, r_2})$ , we deduce that, for all large enough  $N$  (depending on  $\beta$ ), we have that

$$\mathbb{P}(B^c \cap \{m_t^N < r_3\}) \leq C'_\kappa N^{1-\beta\kappa}. \quad (\text{I.43})$$

Using the fact that  $b_t - 4\varepsilon < r_3$ , we have

$$\mathbb{P}(m_t^N < b_t - 4\varepsilon) \leq \mathbb{P}(m_t^N < r_3) \leq \mathbb{P}(B^c \cap \{m_t^N < r_3\}) + \mathbb{P}(B),$$

and combining [\(I.43\)](#) and [\(I.42\)](#), in which the value of  $\kappa$  can be taken as large as we want, we deduce that, for every  $\kappa > 0$ , there exists a constant  $C^{(1)}$  (depending on  $\beta, \kappa, \varepsilon, b, s_0, t_0, T$ , but not on the integer  $k$  for which  $t = k\theta$ ) such that, for all  $N \geq 1$ ,

$$\mathbb{P}(m_t^N < b_t - 4\varepsilon) \leq C^{(1)} N^{-\kappa}. \quad (\text{I.44})$$

*Step 3. Lower bound on the whole time interval.*

We will use the regularity of  $b$  and  $m^N$  to deduce the same kind of bound as [\(I.44\)](#), but on whole time intervals. Indeed, even though  $m^N$  can have some upward jumps (when the particle at the lowest position is killed while being far away from all the other particles), it can never go down too fast, thanks to Ray's estimate. Suppose that for some time  $s \in [t, t + \theta]$ ,  $m_s^N$  is below  $b_s - 7\varepsilon$ . Since  $\theta$  is an  $\varepsilon$ -modulus of continuity for  $b$ , we have that  $b_s - 7\varepsilon < b_t - 6\varepsilon < r_5$ . This means that on the event  $\Xi_{r_4, r_5}^c$  with no downwards crossing of the interval  $[r_5, r_4]$ , a particle that lies below  $b_s - 7\varepsilon$  at time  $s$  had to be below  $r_4$  at time  $t$ , so that

$$m_t^N \leq r_4 < b_t - 4\varepsilon.$$

Hence we have

$$\mathbb{P}(\exists s \in [t, t + \theta], m_s^N < b_s - 7\varepsilon) \leq \mathbb{P}(\Xi_{r_4, r_5}) + \mathbb{P}(m_t^N < b_t - 4\varepsilon).$$

Since the bound [\(I.40\)](#) holds as well for  $\mathbb{P}(\Xi_{r_4, r_5})$ , we deduce, using [\(I.44\)](#), that for every  $\kappa' > 0$ , there exists a constant  $C^{(2)}$  (depending on  $\beta, \kappa', \varepsilon, b, s_0, t_0, T$ ) such that, for all  $N \geq 1$ ,

$$\mathbb{P}(\exists s \in [t, t + \theta], m_s^N < b_s - 7\varepsilon) \leq C^{(2)} N^{-\kappa'}. \quad (\text{I.45})$$

Taking the union bound over every possible  $t = k\theta$  in  $[t_0 - \theta, T]$ , where there are at most  $\frac{T}{\theta}$  of these intervals in  $[t_0, T]$ , we have

$$\mathbb{P}(\exists s \in [t_0, T], m_s^N < b_s - 7\varepsilon) \leq \frac{T}{\theta} C^{(2)} N^{-\kappa} = 2TC^{(2)} N^{\beta-\kappa'}.$$

Given  $\kappa > 0$ , we can set  $\kappa' = \beta + \kappa$ , and deduce a bound, valid for all  $N \geq 1$ , of the form

$$\mathbb{P}(\exists s \in [t_0, T], m_s^N < b_s - 7\varepsilon) \leq C^{(3)} N^{-\kappa}, \quad (\text{I.46})$$

with a constant  $C^{(3)}$  depending on  $\beta, \kappa, \varepsilon, b, s_0, t_0, T$ .

Step 4. *Upper bound and conclusion.*

Now we prove that  $b$  is an approximate upper bound for the minimum for large  $N$ . We start by proving (I.48), which bounds  $\mathbb{P}(m_t^N > b_t + 3\varepsilon)$  for  $t = k\theta$  in  $[t_0, T]$ . Let us first write  $p := e^t U(t, b_t + 3\varepsilon)$ , consider  $\beta < \beta' < 1/2$ , and let us prove that, for large enough  $N$  (depending on  $\beta, \beta', \varepsilon, t_0, T, b$ ), we have

$$p \geq 2N^{-\beta'}. \quad (\text{I.47})$$

For  $i \in \mathbb{Z}$ , let  $r'_i = \min\{x_j \text{ such that } x_j > b_t\} + i\varepsilon$  and note that  $r'_0, \dots, r'_4$  are elements of the space grid subject to (I.30). We then have  $r'_2 \leq b_t + 3\varepsilon$  and thus  $e^t U(t, r'_2) \leq p$  since  $r \mapsto U(t, r)$  is a non-decreasing function. Since  $\tau_\beta$  follows the exponential distribution  $\mathcal{E}(1)$  under  $\mathbb{Q}_{\mu_0}$ , we have that  $\mathbb{Q}_{\mu_0}(\tau_\beta \leq t + \theta | \tau_\beta > t) = 1 - e^{-\theta}$ . On the other hand,

$$\mathbb{Q}_{\mu_0}(\tau_\beta \leq t + \theta | \tau_\beta > t) = \mathbb{Q}_{\mu_0}(X_t > r'_2, \tau_\beta \leq t + \theta | \tau_\beta > t) + \mathbb{Q}_{\mu_0}(X_t \leq r'_2, \tau_\beta \leq t + \theta | \tau_\beta > t).$$

With our assumptions, we have that  $b_s \leq r'_1$  for every  $s \in [t, t + \theta]$ . Using the fact that that  $\tau^b$  is a stopping time, the Markov property, and stochastic monotonicity, we deduce that  $\mathbb{Q}_{\mu_0}(X_t > r'_2, \tau_\beta \leq t + \theta | \tau_\beta > t) \leq \mathbb{Q}_{r'_2}(\tau_{r'_1} \leq \theta)$ , and Ray's estimate (I.30) ensures that this probability is bounded above by  $C_\kappa \theta^\kappa$  (for any  $\kappa > 0$ ). On the other hand, we have that  $\mathbb{Q}_{\mu_0}(X_t \leq r'_2, \tau_\beta \leq t + \theta | \tau_\beta > t) \leq e^t U(t, r'_2) \leq p$ . Putting together the previous results, we have proved that  $1 - e^{-\theta} \leq C_\kappa \theta^\kappa + p$ , so that  $p \geq 1 - e^{-\theta} - C_\kappa \theta^\kappa$ . Remembering that  $\theta = 2N^{-\beta}$ , and choosing  $\kappa > 1$ , (I.47) is proved.

Using the fact that  $p = e^t U(t, b_t + 3\varepsilon)$ , we have that

$$\mathbb{P}(m_t^N > b_t + 3\varepsilon) = \mathbb{P}(F^N(t, b_t + 3\varepsilon) = 0) \leq \mathbb{P}\left(\|F^N(t, \cdot) - e^t U(t, \cdot)\|_\infty > \frac{p}{2}\right). \quad (\text{I.48})$$

Thanks to (I.48), (I.47) and (I.1), we have that

$$\mathbb{P}(m_t^N > b_t + 3\varepsilon) \leq \mathbb{P}\left(\|F^N(t, \cdot) - e^t U(t, \cdot)\|_\infty > N^{-\beta'}\right) \leq c_2 e^{-c_1 N^{\frac{1-2\beta'}{3}}}. \quad (\text{I.49})$$

Then we argue in a similar way as in Step 3. Consider  $t = k\theta$  in  $[t_0, T - \theta]$ . Using the fact that  $b_s \geq b_t - \varepsilon \geq r'_{-2}$  for all  $s \in [t, t + \theta]$ , we see that, on the event  $\{\exists s \in [t, t + \theta], m_s^N > b_s + 7\varepsilon\}$ , the event  $\{m_{t+\theta}^N \leq b_{t+\theta} + 3\varepsilon\}$  implies that at least one particle lies above  $r'_3$  at a certain time  $s \in [t, t + \theta]$  then crosses  $r'_4$  before time  $t + \theta$ . Using Ray's estimate (I.32) and arguing exactly as in the proof of (I.40), we deduce that the probability of the intersection of these two events is bounded above by  $C'_\kappa N^{1-\beta\kappa}$ .

In view of (I.49), we deduce that

$$\mathbb{P}(\exists s \in [t, t + \theta], m_s^N > b_s + 7\varepsilon) \leq c_2 e^{-c_1 N^{\frac{1-2\beta'}{3}}} + C'_\kappa N^{1-\beta\kappa}.$$

Taking the union bound over every possible  $t = k\theta$  in  $[t_0, T - \theta]$  (whose number is bounded above e.g. by  $T/\theta$ ), we obtain an upper bound of the form

$$\mathbb{P}(\exists s \in [t_0, T], m_s^N > b_s + 7\varepsilon) \leq C^{(4)} N^{-\kappa}, \quad (\text{I.50})$$

valid for all  $N \geq 1$  and  $\kappa > 0$ .

Combining (I.50) and with (I.46), we deduce (I.3).

The case  $t_0 = 0$ .

In the case  $t_0 = 0$ , the map  $b$  is continuous on the whole interval  $[0, T]$ , and we can take  $s_0 = 0$  in the definition of the grid in Step 1. The proof of (I.44) for  $t = k\theta$  with  $k \geq 2$  in Step 2 requires no extra argument from the case  $t_0 > 0$ . On the other hand, when  $k = 0$ , the proof is a consequence of the fact that, thanks to condition (i), we must have  $\mu_0((-\infty, b_0)) = 0$ , so that



$\mathbb{P}(m_0^N \geq b_0) = 1$ . And when  $k = 1$ , i.e. for  $t = \theta$ , this property also provides the control that is needed at time  $t - \theta$  to prove (I.44) following the proof given in Step 2. Step 3 is identical to the case  $t_0 > 0$ , and so is Step 4.

*Almost sure convergence.*

Thanks to the Borel-Cantelli lemma, the almost sure convergence on  $m_t^N$  to  $b_t$  in the supremum norm is a straightforward consequence of (I.3). □

**Remark I.2.** *As we mentioned earlier, the constants that appear in this proof are not universal: they depend strongly on the underlying Markov process  $X$  through Ray's estimate and on the boundary  $b$  through the size of the grid and the number of terms used in the union bounds.*



## Chapter II

# The comonotone flow of a stochastically monotone Feller process on the real line

This chapter contains the main part of [BF23b], that can be found at <https://hal.science/hal-04309782>. Appendix C contains some additional details to this work.

### II.1 Introduction

We start by recalling various necessary definitions and properties in Sections II.1.1 to II.1.4, then state our main results in Section II.1.5 and discuss them in Section II.1.6.

#### II.1.1 Markov kernels

A Markov kernel  $K$  on a measurable space  $(S, \mathcal{S})$  is a map from  $S \times \mathcal{S}$  to  $\mathbb{R}$  such that

- (i) for all  $x \in S$ ,  $K(x, \cdot)$  is a probability measure on  $(S, \mathcal{S})$ ;
- (ii) for all  $B \in \mathcal{S}$ ,  $K(\cdot, B)$  is a measurable real-valued function on  $S$ .

When there is no ambiguity regarding the choice of the  $\sigma$ -algebra  $\mathcal{S}$ , we simply say that  $K$  is a Markov kernel on  $S$ . Given a bounded measurable real-valued function  $f$  on  $S$ , we define a bounded measurable real-valued function  $Kf$  on  $S$  by  $Kf(x) = \int_S f(y) dK(x, y)$ . Denoting by  $\mathcal{B}_b(S)$  the vector space of bounded real-valued measurable functions on  $S$ , equipped with the sup-norm  $\|f\|_\infty = \sup_{x \in S} |f(x)|$ , the map  $f \mapsto Kf$  defines a linear operator from  $\mathcal{B}_b(S)$  into itself, and satisfies

- ( $\alpha$ )  $\|Kf\|_\infty \leq \|f\|_\infty$ ;
- ( $\beta$ )  $Kf \geq 0$  when  $f \geq 0$ ;
- ( $\gamma$ )  $K\mathbf{1} = \mathbf{1}$  (where  $\mathbf{1}$  denotes the constant function equal to 1).

#### II.1.2 Markov and Feller semigroups

Given two Markov kernels  $K, L$  on  $S$ , the composition of the two kernels is yet another Markov kernel  $KL$  defined by  $(KL)(x, B) = \int_S L(y, B) dK(x, y)$ . The composition is an associative (but in general non-commutative) operation on Markov kernels. Moreover, for  $f \in \mathcal{B}_b(S)$ , we have that  $(KL)f = K(Lf)$ . We say that a family of Markov kernels  $(K_t)_{t \in \mathbb{R}_+}$  on  $S$  is a *Markov semigroup* if

(I) for all  $x \in S$  and  $B \in \mathcal{S}$ ,  $K_0(x, B) = \delta_x(B)$ ;

(II) the Chapman-Kolmogorov equation holds: for all  $s, t \geq 0$ ,  $K_{s+t} = K_s K_t$ .

Moreover, we say that a family  $(X_t)_{t \in \mathbb{R}_+}$  of  $S$ -valued random variables defined on the same probability space  $(\Omega, \mathcal{F}, \mathbb{Q})$  is a Markov process governed by the Markov semigroup  $(K_t)_{t \in \mathbb{R}_+}$  when, for all  $s, t \in \mathbb{R}_+$ , and all  $B \in \mathcal{S}$ , one has

$$\mathbb{E}(\mathbf{1}_B(X_{s+t}) | \sigma(X_u; u \in [0, s])) \stackrel{\text{a.S.}}{=} K_t(X_s, B).$$

Now assume that  $S$  is a locally compact separable metric space (abbreviated *lcsm* in the sequel), equipped with the corresponding Borel  $\sigma$ -algebra. We denote by  $\mathcal{C}_0(S)$  the vector space of continuous real-valued functions on  $S$  vanishing at infinity. Note that  $\mathcal{C}_0(S)$  equipped with the sup-norm is a Banach space.

We say that a Markov semigroup  $(K_t)_{t \in \mathbb{R}_+}$  on  $S$  enjoys the *Feller property* (or, more succinctly, that it is a *Feller semigroup*) when:

(Fa)  $\forall t \geq 0, \forall f \in \mathcal{C}_0(S), K_t f \in \mathcal{C}_0(S)$ ;

(Fb)  $\forall f \in \mathcal{C}_0(S), \lim_{t \rightarrow 0^+} \|K_t f - f\|_\infty = 0$ .

Let us now denote by  $\mathcal{C}_b(S)$  the vector space of bounded continuous real-valued functions on  $S$ , and observe that property (Fa) implies [\[1\]](#), but is not in general equivalent [\[2\]](#) to

(Fa')  $\forall t \geq 0, \forall f \in \mathcal{C}_b(S), K_t f \in \mathcal{C}_b(S)$ .

Moreover, (Fb) may be replaced by the apparently weaker assumption of pointwise (instead of uniform) convergence

(Fb')  $\forall f \in \mathcal{C}_0(S), \forall x \in S, \lim_{t \rightarrow 0^+} K_t f(x) = f(x)$ ,

but it turns out [\[3\]](#) that a Markov semigroup satisfying (Fa) and (Fb') also satisfies (Fb).

### II.1.3 Consistent families

For every integer  $n \geq 2$ , equip the product space  $S^n$  with the product  $\sigma$ -algebra  $\mathcal{S}^{\otimes n}$ . Given integers  $1 \leq k \leq n$ , and  $i_1, \dots, i_k \in \llbracket 1, n \rrbracket$ , we denote by  $\pi_{i_1, \dots, i_k}^n$  the projection from  $S^n$  to  $S^k$  defined by  $\pi_{i_1, \dots, i_k}^n(x_1, \dots, x_n) = (x_{i_1}, \dots, x_{i_k})$ . We say that a family of Markov kernels  $(K^{(n)})_{n \geq 1}$ , where, for each integer  $n$ ,  $K^{(n)}$  is a Markov kernel on  $S^n$ , is a *consistent family* if, for all  $1 \leq k \leq n$ , all  $i_1, \dots, i_k \in \llbracket 1, n \rrbracket$ , all  $\mathbf{x} \in S^n$  and all measurable subset  $B$  of  $S^k$ ,

$$K^{(n)}(\mathbf{x}, (\pi_{i_1, \dots, i_k}^n)^{-1}(B)) = K^{(k)}(\pi_{i_1, \dots, i_k}^n(\mathbf{x}), B). \quad (\text{II.1})$$

We say that a family of Markov kernels  $(K^{(n)})_{n \geq 2}$  is a *consistent extension* of a Markov kernel  $K$  on  $S$ , if the family  $(K^{(n)})_{n \geq 1}$ , with  $K^{(1)} = K$ , is a consistent family. A family of Markov semigroups  $\left( (K_t^{(n)})_{t \in \mathbb{R}_+} \right)_{n \geq 1}$  is said to be consistent if, for every  $t \in \mathbb{R}_+$ ,  $(K_t^{(n)})_{n \geq 1}$  is a consistent family of Markov kernels; a consistent extension of a Markov semigroup on  $S$  is defined accordingly.

Note that equal indices are allowed in [\(II.1\)](#), which implies in particular that, for all  $x \in S$ , we have

$$K^{(2)}((x, x), \{(y, y); y \in S\}) = 1. \quad (\text{II.2})$$

<sup>1</sup>See [Proposition II.1](#) below.

<sup>2</sup>See e.g. [BCR22](#).

<sup>3</sup>See e.g. [Kal02](#) Theorem 19.6, or [RY99](#) Chapter III, Proposition (2.4).

### II.1.4 Stochastic orders

Now assume that  $S = \mathbb{R}$ , and denote by  $\mathcal{B}_b^\nearrow(\mathbb{R})$  the set of bounded non-decreasing real-valued functions on  $\mathbb{R}$ . The usual stochastic order  $\mu \leq_{\text{st}} \nu$  between (Borel) probability measures on  $\mathbb{R}$  (see [MS02](#)) is defined by the fact that, for all  $f \in \mathcal{B}_b^\nearrow(\mathbb{R})$ , one has  $\int_{\mathbb{R}} f(x) d\mu(x) \leq \int_{\mathbb{R}} f(x) d\nu(x)$ . We say that a Markov semigroup  $(K_t)_{t \in \mathbb{R}_+}$  on  $\mathbb{R}$  is *stochastically monotone* when:

$$(M) \quad \forall t \geq 0, \forall f \in \mathcal{B}_b^\nearrow(\mathbb{R}), K_t f \in \mathcal{B}_b^\nearrow(\mathbb{R}).$$

An immediately equivalent formulation of (M) in terms of  $\leq_{\text{st}}$  is that

$$\forall x, y \in \mathbb{R}, x \leq y \Rightarrow K_t(x, \cdot) \leq_{\text{st}} K_t(y, \cdot). \quad (\text{II.3})$$

Given  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$  and  $\mathbf{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$ , we say that  $\mathbf{y}$  is order-compatible with  $\mathbf{x}$  if, for all  $1 \leq i, j \leq n$ ,  $x_i \leq x_j \Rightarrow y_i \leq y_j$ . We then denote

$$\mathbb{R}_{\mathbf{x}}^n = \{\mathbf{y} \in \mathbb{R}^n \text{ such that } \mathbf{y} \text{ is order-compatible with } \mathbf{x}\},$$

and we say that a Markov kernel  $K^{(n)}$  on  $\mathbb{R}^n$  is *order-preserving* when, for all  $\mathbf{x} \in \mathbb{R}^n$ , we have that  $K^{(n)}(\mathbf{x}, \mathbb{R}_{\mathbf{x}}^n) = 1$ .

Denote by  $\mathcal{SM}_b(\mathbb{R}^n)$  the set of real-valued bounded Borel super-modular functions on  $\mathbb{R}^n$ , i.e. bounded Borel functions  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  such that, for all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ ,  $f(\mathbf{x}) + f(\mathbf{y}) \leq f(\mathbf{x} \vee \mathbf{y}) + f(\mathbf{x} \wedge \mathbf{y})$ , where, for  $\mathbf{x} = (x_1, \dots, x_n)$  and  $\mathbf{y} = (y_1, \dots, y_n)$ , we set  $\mathbf{x} \vee \mathbf{y} = (\max(x_1, y_1), \dots, \max(x_n, y_n))$  and  $\mathbf{x} \wedge \mathbf{y} = (\min(x_1, y_1), \dots, \min(x_n, y_n))$ .

Given  $n \geq 2$  and two (Borel) probability measures  $\mu, \nu$  on  $\mathbb{R}^n$ , we say that  $\mu \leq_{\text{sm}} \nu$  (see e.g. [MS02](#)) when

$$\forall f \in \mathcal{SM}_b(\mathbb{R}^n), \int_{\mathbb{R}^n} f(\mathbf{x}) d\mu(\mathbf{x}) \leq \int_{\mathbb{R}^n} f(\mathbf{x}) d\nu(\mathbf{x}). \quad (\text{II.4})$$

Note that  $\leq_{\text{sm}}$  defines a partial order on the set of probability measures on  $\mathbb{R}^n$ , and that two probability measures that are comparable with respect to  $\leq_{\text{sm}}$  must have the same marginal distributions.

### II.1.5 Statement of the main results

**Theorem II.1.** *If  $(P_t)_{t \in \mathbb{R}_+}$  is a stochastically monotone Feller semigroup on  $\mathbb{R}$ , there exists a consistent extension of  $(P_t)_{t \geq 0}$  by a family  $\left( (P_t^{(n)})_{t \in \mathbb{R}_+} \right)_{n \geq 1}$  of order-preserving Feller semigroups.*

For  $n \geq 2$ , denote by  $\mathfrak{M}_n$  the set of Feller semigroups  $(M_t^{(n)})_{t \in \mathbb{R}_+}$  on  $\mathbb{R}^n$  such that, for every  $t \geq 0$ , every  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ , and every  $1 \leq i \leq n$ ,  $M_t^{(n)}(\mathbf{x}, (\pi_i^n)^{-1}(B)) = P_t(x_i, B)$ .

**Theorem II.2.** *Let  $(P_t)_{t \in \mathbb{R}_+}$  be a stochastically monotone Feller semigroup on  $\mathbb{R}$ , and, for  $n \geq 2$ , denote by  $(\check{P}_t^{(n)})_{t \in \mathbb{R}_+}$  the [4](#) Markov semigroup constructed in the proof of [Theorem II.1](#). Then, for every  $(M_t^{(n)})_{t \in \mathbb{R}_+} \in \mathfrak{M}_n$ ,  $s \in \mathbb{R}_+$  and  $\mathbf{x} \in \mathbb{R}^n$ , one has that  $M_s^{(n)}(\mathbf{x}, \cdot) \leq_{\text{sm}} \check{P}_s^{(n)}(\mathbf{x}, \cdot)$ . Since by construction  $(\check{P}_t^{(n)})_{t \in \mathbb{R}_+} \in \mathfrak{M}_n$ , and since  $\leq_{\text{sm}}$  is a partial order, this shows that  $(\check{P}_t^{(n)})_{t \in \mathbb{R}_+}$  is uniquely characterized as the solution of the following maximization problem with respect to  $\leq_{\text{sm}}$ :*

$$\check{P}_s^{(n)}(\mathbf{x}, \cdot) = \max_{(M_t^{(n)})_{t \in \mathbb{R}_+} \in \mathfrak{M}_n} M_s^{(n)}(\mathbf{x}, \cdot).$$

<sup>4</sup>A priori, this may not be unique, but it turns out that it is in view of the present statement.

## II.1.6 Discussion

Broadly speaking, [Theorem II.1](#) is a monotonicity equivalence result in the sense of [\[EM01\]](#): starting from a monotonicity property within a family of probability measures, one deduces the existence of a monotone coupling, i.e. a family of random variables providing an effective realization of the monotonicity property. The archetype for such results is Strassen's theorem<sup>5</sup>: if  $\mu \leq_{\text{st}} \nu$ , there exists a pair of random variables  $(X, Y)$  such that  $\text{Law}(X) = \mu$ ,  $\text{Law}(Y) = \nu$ , and almost surely  $X \leq Y$ . In our context, the existence of the Feller semigroup  $(P_t^{(n)})_{t \in \mathbb{R}_+}$  shows that<sup>6</sup>, given  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ , one can define a family of real-valued random variables  $(X_t^{x_1}, \dots, X_t^{x_n})_{t \in \mathbb{R}_+}$  on the same probability space, such that:

- for  $1 \leq i \leq n$ ,  $X_t^{x_i}$  is a real-valued Markov process with càdlàg paths, starting at  $x_i$ , and governed by  $(P_t)_{t \in \mathbb{R}_+}$ ,
- for all  $t \in \mathbb{R}_+$ ,  $(X_t^{x_1}, \dots, X_t^{x_n})$  is order-compatible with  $(x_1, \dots, x_n)$ .

The fact that  $(X_t^{x_1}, \dots, X_t^{x_n})_{t \in \mathbb{R}_+}$  is itself a Markov process on  $\mathbb{R}^n$  governed by a Feller Markov semigroup comes as an important additional property (see [\[BF23a\]](#), which was indeed our original motivation for investigating the present question).

Results of this kind have been established in a variety of contexts, including<sup>7</sup> discrete-time Markov chains on partially ordered Polish spaces ([\[KKO77\]](#)), continuous-time Markov chains on countable partially ordered spaces ([\[LMS00\]](#)), jump processes on partially ordered Polish spaces ([\[Zha00\]](#)), interacting particle systems on  $\{0, 1\}^{\mathbb{Z}^d}$  ([\[Lig05\]](#)).

In the setting of continuous-time Markov processes on the real line, let us mention two important special cases for which the conclusion of [Theorem II.1](#) can be established relatively easily:

- Lévy processes (see e.g. [\[Ber96\]](#)): using the *parallel coupling*, obtained by setting, for all  $x \in \mathbb{R}$  and  $t \geq 0$ ,  $X_t^x = x + X_t^0$ , where  $(X_t^0)_{t \geq 0}$  is a version of the Lévy process starting at 0, see [Figure II.1a](#),
- Feller processes with continuous paths: one defines  $P_t^{(n)}$  as the distribution at time  $t$  of a family<sup>8</sup> of  $n$  trajectories that evolve independently until they meet, and stick together thereafter (this is often called the *Doebelin coupling*), see [Figure II.1b](#); the construction with  $n = 2$  is a nice way of proving the not-so-obvious fact that any Feller process on the real line with continuous paths is stochastically monotone.

In the case of jump processes ([\[Che04\]](#)), and more generally, of Lévy-type processes ([\[Wan13, Kol11\]](#)), necessary and sufficient explicit conditions for stochastic monotonicity, expressed in terms of the infinitesimal generator of the process, have been obtained. One nice feature of [Theorem II.1](#) is that it holds in a completely general setting, regardless of the specific structure

<sup>5</sup>Strassen's theorem is valid in the general context of probability measures on Polish spaces. See [\[Str65\]](#) for Strassen's original paper, and [\[Lin99\]](#) for useful additional elements.

<sup>6</sup>From classical results (see e.g. [\[RY99\]](#) or [\[Kal02\]](#)), we have the existence of an  $\mathbb{R}^n$ -valued Markov process  $(X_t^{x_1}, \dots, X_t^{x_n})_{t \in \mathbb{R}_+}$  with càdlàg paths, governed by  $(P_t^{(n)})_{t \in \mathbb{R}_+}$  and starting at  $\mathbf{x}$ . Then, by the order-preserving property of  $P_t^{(n)}$ , almost surely, for all  $t \in \mathbb{Q}_+$ ,  $(X_t^{x_1}, \dots, X_t^{x_n}) \in \mathbb{R}_{\mathbf{x}}^n$ . Then, using right-continuity of paths and the fact that  $\mathbb{R}_{\mathbf{x}}^n$  is a closed subset of  $\mathbb{R}^n$ , we have that, almost surely, for all  $t \in \mathbb{R}_+$ ,  $(X_t^{x_1}, \dots, X_t^{x_n}) \in \mathbb{R}_{\mathbf{x}}^n$ .

<sup>7</sup>For the sake of brevity, we limit ourselves to a small sample of references, and point the interested reader to the works quoted by, or quoting, these references for a more exhaustive view of this subject.

<sup>8</sup>A construction for general Feller processes (not necessarily with continuous paths) leading to a strong Markov process is given in [\[EMS13\]](#) Section 2.1. Continuity of paths easily shows that  $P_t^{(n)}$  is order-preserving. The Feller property of  $(P_t^{(n)})_{t \geq 0}$  can then be established using the fact that we have a consistent extension of a Feller semigroup on the real line by order-preserving semigroups, exactly as in the Proof of [Theorem II.1](#)

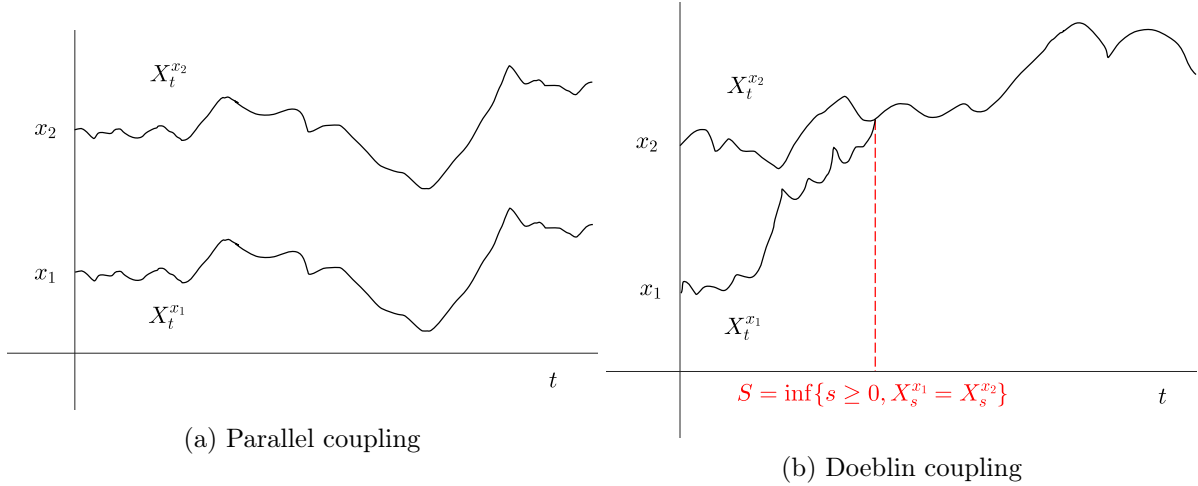


Figure II.1: Special cases of [Theorem II.1](#) with easier constructions

of the underlying process, as soon as it is stochastically monotone and enjoys the Feller property, which are also necessary conditions for the conclusion of the theorem to hold.

[Theorem II.2](#) shows that, for all  $n \geq 2$ ,  $s \in \mathbb{R}_+$  and  $\mathbf{x} = (x_1, \dots, x_n)$ ,  $\check{P}_s^{(n)}(\mathbf{x}, \cdot)$  produces an  $n$ -dimensional coupling of the distributions  $P_s(x_1, \cdot), \dots, P_s(x_n, \cdot)$ , enjoying a maximality property with respect to the super-modular order. Without the constraint that this coupling has to be the distribution at time  $s$  of a Feller process on  $\mathbb{R}^n$  whose marginals evolve according to  $(P_t)_{t \in \mathbb{R}_+}$ , the maximum would be achieved by the classical comonotone coupling of  $P_s(x_1, \cdot), \dots, P_s(x_n, \cdot)$  (see [\[MS02\]](#)). Loosely speaking,  $\check{P}^{(n)}$  is obtained (as a well-defined limit) by composing infinitely many such comonotone couplings associated with an infinitesimal time interval  $s$ , so we choose to call the resulting object the *comonotone flow* associated with  $(P_t)_{t \geq 0}$ . Strictly speaking, we have not defined a flow but only a consistent family  $\left( (P_t^{(n)})_{t \in \mathbb{R}_+} \right)_{n \geq 1}$  of Feller Markov semigroups. However, it is shown<sup>9</sup> in [\[LJR04\]](#) that such a family of semigroups corresponds to a stochastic flow, in a precise sense. We do not push this question further here, and postpone the study of the various properties enjoyed by this stochastic flow to future work.

### II.1.7 Organization of the paper

In [Section II.2](#), we collect various results related to the Feller property of Markov semigroups, on general (locally compact separable metric, or compact metric) spaces, then on  $\mathbb{R}$ . [Section II.3](#) is devoted to the proof of [Theorem II.1](#), and [Section II.4](#) to the proof of [Theorem II.2](#). [Section C.1](#) is an appendix where we explain why we think an element is missing in the proof of some results in [\[KKO77\]](#) which are related to the conclusion of [Theorem II.1](#) in the present paper. [Section C.3](#) shows some numerical illustrations.

## II.2 Feller-related results

This section is devoted to results related to the Feller property of Markov semigroups. In [Section II.2.1](#), we provide alternative formulations of the Feller property that turn out to be more suitable for the arguments developed in [Section II.3](#), namely, transferring the assumed

<sup>9</sup>See also the correction [\[LJR20\]](#) and the fully corrected version arXiv:math/0203221v6. Condition [\(II.1\)](#) and its consequence [\(II.2\)](#) precisely match the definition in [\[LJR04\]](#) of a compatible family of Feller semigroups satisfying  $P_t^{(2)} f^{\otimes 2}(x, x) = P_t f^2(x)$  for all  $x \in S$  and  $f \in C_0(S)$ .

regularity properties of the family of kernels  $(P_t)_{t \in \mathbb{R}_+}$  on  $\mathbb{R}$ , to families of kernels on  $\mathbb{R}^n$  that extend  $P_t$ . In [Section II.2.2](#), we show that the combination of stochastic monotonicity with the Feller property allows one to extend  $(P_t)_{t \in \mathbb{R}_+}$  to a Feller semigroup  $(\tilde{P}_t)_{t \in \mathbb{R}_+}$  on the extended real line  $\mathbb{R}$ , whose advantage (in our context) over the usual one-point compactification  $\mathbb{R} \cup \{\infty\}$  is its compatibility with the underlying order on  $\mathbb{R}$ .

## II.2.1 General properties

Given a lcsm space, we denote by  $\mathcal{P}(S)$  the set of (Borel) probability measures on  $S$ , equipped with the topology of weak convergence, which is metrizable (see below). To a Markov kernel  $K$  on  $S$ , we associate the map  $\check{K}$  from  $S$  to  $\mathcal{P}(S)$ , defined as  $x \mapsto K(x, \cdot)$ . Conversely, to a map  $\mathbf{K}$  from  $S$  to  $\mathcal{P}(S)$ , we associate the map  $\hat{\mathbf{K}}$  from  $S \times \mathcal{S}$  to  $\mathbb{R}$  defined as  $(x, B) \mapsto [\mathbf{K}(x)](B)$ , which may or may not be a Markov kernel, depending on whether the map  $\hat{\mathbf{K}}(\cdot, B)$  is measurable for every  $B$ .

**Lemma II.1.** *Given a lcsm space  $S$ , and a continuous map  $\mathbf{K} : S \rightarrow \mathcal{P}(S)$ ,  $\hat{\mathbf{K}}$  is a Markov kernel.*

*Proof.* Arguing as in the proof of [Proposition II.1](#) below (there is no circular reference, we just wanted to avoid writing the same proof twice), we have that, for all  $f \in \mathcal{C}_b(S)$ , the map  $x \mapsto \hat{\mathbf{K}}f(x) = \int_S f(y) d[\mathbf{K}(x)](y)$  is itself an element of  $\mathcal{C}_b(S)$ , and in particular is a Borel map.

We now apply Theorem 0.2.2 in [\[RY99\]](#), which is a functional version of the monotone class theorem: the set  $\mathcal{H}$  of functions  $f \in \mathcal{B}_b(S)$  such that  $x \mapsto \hat{\mathbf{K}}f(x)$  is Borel contains  $\mathcal{C}_b(S)$  – which plays the role of  $\mathcal{C}$  in [\[RY99\]](#) – and the assumptions of the theorem are met<sup>10</sup>, so that  $\mathcal{H}$  contains all  $\sigma(\mathcal{C}_b(S))$ -measurable functions, and thus all bounded real-valued Borel functions on  $S$ . In particular  $\hat{\mathbf{K}}\mathbf{1}_B$  is a Borel function for any Borel set  $B$  in  $S$ , so  $\hat{\mathbf{K}}$  is indeed a Markov kernel on  $S$ .  $\square$

**Proposition II.1.** *Given a Markov kernel  $K$  on a lcsm space  $S$ , the following three properties are equivalent:*

- (i)  $\check{K}$  is a continuous map;
- (ii)  $\forall f \in \mathcal{C}_b(S), Kf \in \mathcal{C}_b(S)$ ;
- (iii)  $\forall f \in \mathcal{C}_0(S), Kf \in \mathcal{C}_b(S)$ .

*Proof.* Remember that, for a Markov kernel  $K$  and  $f \in \mathcal{B}_b(S)$ , one always has  $Kf \in \mathcal{B}_b(S)$ , so that the only stake in (ii) and (iii) is the continuity of  $Kf$ . Now, by definition, continuity of the map  $x \mapsto K(x, \cdot)$  means that, for any sequence  $(x_k)_{k \geq 1}$  such that  $\lim_{k \rightarrow +\infty} x_k = x$ , we have

the weak convergence  $K(x_k, \cdot) \xrightarrow[k \rightarrow +\infty]{\mathbf{w}} K(x, \cdot)$ . In turn, weak convergence means that, for every  $f \in \mathcal{C}_b(S)$ ,  $\lim_{k \rightarrow +\infty} Kf(x_k) = Kf(x)$ , which reads as the continuity of  $Pf$  at  $x$ . As a consequence, (i) and (ii) are equivalent. The equivalence of (i) and (iii) stems from the fact that, since  $S$

is a lcsm space, the weak convergence  $K(x_k, \cdot) \xrightarrow[k \rightarrow +\infty]{\mathbf{w}} K(x, \cdot)$  is equivalent to the fact that  $\lim_{k \rightarrow +\infty} Kf(x_k) = Kf(x)$  for every  $f \in \mathcal{C}_0(S)$ .  $\square$

<sup>10</sup>The set  $\mathcal{H}$  is a vector space that contains the constant functions and the supremum of any bounded non-decreasing sequence of its non-negative elements (thanks to the monotone convergence theorem); the set  $\mathcal{C}$  is stable under pointwise multiplication.



**Proposition II.2.** *A Markov kernel  $K$  on a lcsm space  $S$  satisfies*

$$\forall f \in \mathcal{C}_0(S), Kf \in \mathcal{C}_0(S), \quad (\text{II.5})$$

if and only if the following two conditions are satisfied:

- (i)  $\check{K}$  is a continuous map;
- (ii) for every compact subset  $C$  of  $S$ ,  $\lim_{x \rightarrow \infty} K(x, C) = 0$ .

*Proof.* We start with the "if" part, assuming that (i) and (ii) are satisfied. Consider  $f \in \mathcal{C}_0(S)$ . Since  $\mathcal{C}_0(S) \subset \mathcal{C}_b(S)$ , Proposition II.1 and (i) show that  $Kf \in \mathcal{C}_b(S)$ , and it remains to prove that  $Kf$  goes to zero at infinity. Given  $\varepsilon > 0$  and a compact  $C$  such that  $|f| \leq \varepsilon$  outside  $C$ , we have  $|Kf(x)| \leq \|f\|_\infty K(x, C) + \varepsilon$ , so that, thanks to (ii),  $|Kf(x)| \leq 2\varepsilon$  as soon as  $x$  is outside a sufficiently large compact set  $C'$ .

Now for the "only if" part, assuming that (II.5) is satisfied. From Proposition (II.1), we have (i), since  $\mathcal{C}_0(S) \subset \mathcal{C}_b(S)$ . As for (ii), consider a compact set  $C$ . By compactness of  $C$  and local compactness of  $S$ , there exists<sup>11</sup> a continuous function  $f$  with compact support such that  $f \geq \mathbf{1}_C$ . Then  $K(x, C) = K\mathbf{1}_C(x) \leq Kf(x)$ , and, since  $f \in \mathcal{C}_0(S)$ , (II.5) implies that  $Kf \in \mathcal{C}_0(S)$  so that  $\lim_{x \rightarrow \infty} Kf(x) = 0$ .  $\square$

We now focus on the case where  $S$  is a compact (hence separable) metric space, denoting by  $\mathcal{C}(S)$  the space of continuous real-valued functions on  $S$  (due to the fact that  $S$  is compact, we have  $\mathcal{C}(S) = \mathcal{C}_b(S) = \mathcal{C}_0(S)$ ). Among the many possible metrics compatible with the weak convergence topology on  $\mathcal{P}(S)$ , we choose to use the Wasserstein (or Kantorovich-Rubinstein)  $W_1$  distance (see [Vil09a], Chapter 6). Specifically, consider the space  $\text{Lip}(S)$  of real-valued Lipschitz functions on  $S$ . Denoting by  $d_S$  the metric on  $S$ , for  $f \in \text{Lip}(S)$ , we let

$$\|f\|_{\text{Lip}} = \sup\{|f(y) - f(x)|/d_S(x, y); x, y \in S, x \neq y\}.$$

Note that, since  $S$  is compact, every  $f \in \text{Lip}(S)$  is bounded. The  $W_1$  distance on  $\mathcal{P}(S)$  is then given by

$$W_1(\mu, \nu) = \sup \left\{ \left| \int_S f(x) d\mu(x) - \int_S f(x) d\nu(x) \right|; \|f\|_{\text{Lip}} \leq 1 \right\}. \quad (\text{II.6})$$

**Lemma II.2.** *Let  $K, L, M$  be Markov kernels on  $S$ , then*

$$\sup_{x \in S} W_1(KL(x, \cdot), KM(x, \cdot)) \leq \sup_{x \in S} W_1(L(x, \cdot), M(x, \cdot)).$$

*Proof.* Consider  $f \in \text{Lip}(S)$  such that  $\|f\|_{\text{Lip}} \leq 1$ . By definition, for all  $x \in S$ , we have that  $|Lf(x) - Mf(x)| \leq W_1(L(x, \cdot), M(x, \cdot))$ , so that  $\|Lf - Mf\|_\infty \leq \sup_{x \in S} W_1(L(x, \cdot), M(x, \cdot))$ .

Using the contraction property ( $\alpha$ ) of Markov kernels,  $\|KLf - KMf\|_\infty \leq \|Lf - Mf\|_\infty$ , so that

$$\sup_{x \in S} \sup_{f \in \text{Lip}(S), \|f\|_{\text{Lip}} \leq 1} |KLf(x) - KMf(x)| \leq \sup_{x \in S} W_1(L(x, \cdot), M(x, \cdot)).$$

$\square$

<sup>11</sup>This is a rather standard result. Here is an elementary argument, using the notation of the proof of Proposition II.3. Since  $S$  is locally compact, for every  $x \in C$ , there exists a number  $r > 0$  such that  $\overline{B(x, r)}$  is compact. By compactness of  $C$ , we deduce that there exist  $m \geq 1$ ,  $x_1, \dots, x_m \in C$  and  $r_1, \dots, r_m > 0$  such that  $C \subset U$ , where  $U = \bigcup_{k=1}^m \overline{B(x_k, r_k)}$ . By compactness of  $C$  again, the continuous function  $x \mapsto d(x, U^c)$  has a minimum value  $\rho$  on  $C$ , and, since  $U^c$  is closed and  $C \cap U^c = \emptyset$ , we must have  $\rho > 0$ . Choosing  $\varepsilon \in ]0, \rho[$  ensures that  $C^\varepsilon \subset U$ . As a consequence,  $C^\varepsilon$  is included in the compact set  $\bigcup_{k=1}^m \overline{B(x_k, r_k)}$ , so that  $\overline{C^\varepsilon}$  is compact. One can then take  $f = I_{C^\varepsilon}$ .

**Proposition II.3.** *Given a family of Markov kernels  $(K_t)_{t \in J}$  on a compact metric space  $S$ , where  $J$  is a subset of  $]0, +\infty[$  such that  $\inf J = 0$ , the following properties are equivalent:*

(i)  $\forall f \in \mathcal{C}(S), \lim_{t \rightarrow 0^+} \|K_t f - f\|_\infty = 0$

(ii) *for all  $\varepsilon > 0$ , and all  $x \in S$ ,  $K_t(x, B_S(x, \varepsilon)^c)$  goes 12 to 0 as  $t$  goes to 0, uniformly over  $x \in S$ ;*

(iii) *for all  $x \in S$ ,  $W_1(K_t(x, \cdot), \delta_x)$  goes to 0 as  $t$  goes to 0, uniformly over  $x \in S$ .*

*Proof.* Before we start the proof, let us introduce a few notation and definitions. For a non-empty set  $A \subset S$  and  $x \in S$ , we let  $d_S(x, A) = \inf\{d_S(x, y); y \in A\}$ , and, for  $r > 0$ , we denote  $A^r = \{x \in S; d_S(x, A) < r\}$ . We then define the function  $I_A^r$  on  $S$  by  $I_A^r(x) = 1 - \frac{\min(d_S(x, A), r)}{r}$ . We have that  $0 \leq I_A^r \leq 1$ ,  $I_A^r = 1$  on  $A$ ,  $I_A^r = 0$  outside  $A^r$ , and  $\|I_A^r\|_{\text{Lip}} \leq 1/r$ .

We start with (i)  $\Rightarrow$  (ii). Assume that (i) holds, and consider  $\varepsilon > 0$ ,  $x \in S$ , and the function  $f = I_{B_S(x, \varepsilon/3)}^{\varepsilon/3}$ . For all  $y \in B_S(x, \varepsilon/3)$  one has that  $f(y) = 1$ , and, for all  $z \notin B_S(x, \varepsilon/3)$ , one has that  $z \notin B_S(x, \varepsilon/3)^{\varepsilon/3}$  so that  $f(z) = 0$ . Since we also have  $f \leq 1$ , we deduce that  $K_t(y, B_S(y, \varepsilon)^c) \leq 1 - K_t f(y)$ . From (i),  $K_t f(y)$  goes to  $f(y) = 1$  as  $t$  goes to 0, uniformly with respect to  $y \in B_S(x, \varepsilon/3)$ , so that  $K_t(y, B_S(y, \varepsilon)^c)$  goes to 0 as  $t$  goes to 0, uniformly with respect to  $y \in B_S(x, \varepsilon/3)$ . Covering the compact set  $S$  by a finite number of balls of the form  $B_S(x, \varepsilon/3)$ , we have proved that (ii) holds.

We now prove that (ii)  $\Rightarrow$  (i). Assume that (ii) holds. Consider  $f \in \mathcal{C}(S)$  and  $\varepsilon > 0$ . Since  $S$  is compact,  $f$  is uniformly continuous and there is a  $\delta > 0$  such that, whenever  $d_S(x, y) \leq \delta$ , one has  $|f(x) - f(y)| \leq \varepsilon$ . As a consequence, for all  $x \in S$ ,

$$|K_t f(x) - f(x)| \leq \varepsilon + 2 \|f\|_\infty K_t(x, B_S(x, \varepsilon)^c),$$

so (ii) leads to the desired conclusion.

The proof that (ii)  $\Rightarrow$  (iii) is similar, where Lipschitz continuity replaces uniform continuity. Indeed, for  $f \in \text{Lip}(S)$ , we have, for all  $\varepsilon > 0$ , and for all  $x \in S$ , that

$$|K_t f(x) - f(x)| \leq \|f\|_{\text{Lip}} \varepsilon + \Delta_S \|f\|_{\text{Lip}} K_t(x, B_S(x, \varepsilon)^c),$$

using the fact that, for all  $x, y \in S$ ,  $|f(y) - f(x)| \leq \|f\|_{\text{Lip}} \Delta_S$ , where  $\Delta_S = \sup_{x, y \in S} d_S(x, y)$  is

finite thanks to the fact that  $S$  is compact.

We now prove that (iii)  $\Rightarrow$  (ii). Given  $x \in S$  and  $\varepsilon > 0$ , consider the function  $f = I_{B_S(x, \varepsilon/3)}^{\varepsilon/3}$  used in the proof that (i)  $\Rightarrow$  (ii). We have that  $\|f\|_{\text{Lip}} \leq 3/\varepsilon$ , so that  $g = (3/\varepsilon)^{-1} f$ , satisfies  $\|g\|_{\text{Lip}} \leq 1$ , and as a consequence, for all  $y \in S$ ,  $|K_t f(y) - f(y)| \leq (3/\varepsilon) W_1(K_t(y, \cdot), \delta_y)$ , and we can argue as in the proof that (i)  $\Rightarrow$  (ii), thanks to (iii). □

## II.2.2 Extension of $(P_t)$ to $\overline{\mathbb{R}}$

For  $u, v \in \mathbb{R}$ , denote by  $\mathcal{C}_{u,v}(\mathbb{R})$  the set of continuous functions  $f : \mathbb{R} \rightarrow \mathbb{R}$  such that we have  $\lim_{x \rightarrow -\infty} f(x) = u$  and  $\lim_{x \rightarrow +\infty} f(x) = v$ .

**Proposition II.4.** *A Markov kernel  $K$  on  $\mathbb{R}$  satisfying II.5 and*

$$\forall f \in \mathcal{B}_b^\uparrow(\mathbb{R}), Kf \in \mathcal{B}_b^\uparrow(\mathbb{R}), \tag{II.7}$$

*also satisfies*

$$\forall f \in \mathcal{C}_{u,v}(\mathbb{R}), Kf \in \mathcal{C}_{u,v}(\mathbb{R}). \tag{II.8}$$

---

<sup>12</sup>We denote by  $B_S(x, r)$  the open ball of radius  $r$  centered at  $x$ .

*Proof.*

We first prove that, for all  $a > 0$ ,

$$\lim_{x \rightarrow -\infty} K(x, [-a, +\infty]) = 0 \text{ and } \lim_{x \rightarrow +\infty} K(x, ]-\infty, a]) = 0. \quad (\text{II.9})$$

Given  $a > 0$ , consider  $\varepsilon > 0$ . Then choose  $b \geq a$  such that  $K(0, [-b, +b]^c) \leq \varepsilon$ . Thanks to (II.5) and Proposition (II.2), there exists  $x_0 \leq 0$  such that, for all  $x \leq x_0$ ,  $K(x, [-b, +b]) \leq \varepsilon$ . By the monotonicity assumption (II.7),  $K(x, ]b, +\infty]) \leq K(0, ]b, +\infty]) \leq \varepsilon$ . Combining the two inequalities for  $K(x, \cdot)$ , we deduce that  $K(x, [-b, +\infty]) \leq 2\varepsilon$ , and, since  $b \geq a$ , we have that, for all  $x \leq x_0$ ,  $K(x, [-a, +\infty]) \leq K(x, [-b, +\infty]) \leq 2\varepsilon$ . We have proved the first part of (II.9). The second part is proved symmetrically.

Now consider  $f \in \mathcal{C}_{u,v}(\mathbb{R})$ . Given  $\varepsilon > 0$ , there exists  $a > 0$  such that, for all  $z < -a$ ,  $|f(z) - u| \leq \varepsilon$ . As a consequence,  $|Kf(x) - u| \leq \varepsilon + (\|f\|_\infty + u)K(x, [-a, +\infty])$ . Using the first part of (II.9), we deduce that  $\limsup_{x \rightarrow -\infty} |Kf(x) - u| \leq \varepsilon$ . Since  $\varepsilon$  is arbitrary, we have thus proved that  $\lim_{x \rightarrow -\infty} Kf(x) = u$ . A similar argument proves that  $\lim_{x \rightarrow +\infty} Kf(x) = v$ .

Finally, since  $f \in \mathcal{C}_b(\mathbb{R})$  and  $K$  satisfies (II.5), Proposition II.2 shows that  $Kf$  is a continuous function. □

We equip the extended real line  $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\} = [-\infty, +\infty]$  with the metric  $d_1$  defined by  $d_1(x, y) = |\phi(y) - \phi(x)|$ , where  $\phi(x) = \tanh(x)$  (with  $\tanh(+\infty) = 1$  and  $\tanh(-\infty) = -1$ ), which makes it a separable compact metric space. For  $n \geq 2$ , we equip  $\overline{\mathbb{R}}^n$  with the metric  $d_n$  defined by  $d_n(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n |\phi(y_i) - \phi(x_i)|$ .

Given a Borel probability measure  $\mu$  on  $\mathbb{R}$ , we extend it to a probability measure  $\tilde{\mu}$  on  $\overline{\mathbb{R}}$  by letting, for every Borel set  $B$  of  $\overline{\mathbb{R}}$ ,  $\tilde{\mu}(B) = \mu(B \cap \mathbb{R})$ .

**Lemma II.3.** *Given Borel probability measures  $\mu, \mu_1, \mu_2, \dots$  on  $\mathbb{R}$ ,  $\mu_k \xrightarrow[k \rightarrow +\infty]{\mathbf{w}} \mu$  if and only if  $\tilde{\mu}_k \xrightarrow[k \rightarrow +\infty]{\mathbf{w}} \tilde{\mu}$ .*

*Proof.* Assume that  $\tilde{\mu}_k \xrightarrow[k \rightarrow +\infty]{\mathbf{w}} \tilde{\mu}$ . For  $f \in \mathcal{C}_0(\mathbb{R})$ ,  $f$  can be extended to a function  $\tilde{f} \in \mathcal{C}(\overline{\mathbb{R}})$  by letting  $\tilde{f}(\pm\infty) = 0$ . With this definition, for any probability measure  $\nu$  on  $\mathbb{R}$ , we have that  $\int_{\mathbb{R}} f(x) d\nu(x) = \int_{\overline{\mathbb{R}}} \tilde{f}(x) d\nu(x)$ . The fact that  $\lim_{k \rightarrow +\infty} \int_{\overline{\mathbb{R}}} \tilde{f}(x) d\tilde{\mu}_k(x) = \int_{\overline{\mathbb{R}}} \tilde{f}(x) d\tilde{\mu}(x)$  thus implies that  $\lim_{k \rightarrow +\infty} \int_{\mathbb{R}} f(x) d\mu_k(x) = \int_{\mathbb{R}} f(x) d\mu(x)$ . So we have proved that  $\mu_k \xrightarrow[k \rightarrow +\infty]{\mathbf{w}} \mu$ .

Now assume that  $\mu_k \xrightarrow[k \rightarrow +\infty]{\mathbf{w}} \mu$ . For  $f \in \mathcal{C}(\overline{\mathbb{R}})$ , the restriction of  $f$  to  $\mathbb{R}$  is a bounded continuous function, so we have that  $\lim_{k \rightarrow +\infty} \int_{\mathbb{R}} f(x) d\mu_k(x) = \int_{\mathbb{R}} f(x) d\mu(x)$ , whence

$$\lim_{k \rightarrow +\infty} \int_{\overline{\mathbb{R}}} f(x) d\tilde{\mu}_k(x) = \int_{\overline{\mathbb{R}}} f(x) d\tilde{\mu}(x).$$

So we have proved that  $\tilde{\mu}_k \xrightarrow[k \rightarrow +\infty]{\mathbf{w}} \tilde{\mu}$ . □

We extend  $(P_t)_{t \in \mathbb{R}_+}$  to a family of Markov kernels  $(\tilde{P}_t)_{t \in \mathbb{R}_+}$  on  $\overline{\mathbb{R}}$  by setting, for all  $x \in \mathbb{R}$ :  $\tilde{P}_t(x, \cdot) = \widetilde{P_t(x, \cdot)}$ ,  $\tilde{P}_t(+\infty, \cdot) = \delta_{+\infty}(\cdot)$  and  $\tilde{P}_t(-\infty, \cdot) = \delta_{-\infty}(\cdot)$ .

**Proposition II.5.** *If  $(P_t)_{\mathbb{R}_+}$  is stochastically monotone and enjoys the Feller property, then  $(\tilde{P}_t)_{t \geq 0}$  has the Feller property.*

*Proof.* We first prove (Fa), i.e.  $\forall f \in \mathcal{C}(\overline{\mathbb{R}})$  and  $t \geq 0$ ,  $\tilde{P}_t f \in \mathcal{C}(\overline{\mathbb{R}})$ . Given  $f \in \mathcal{C}(\overline{\mathbb{R}})$ , we denote by  $f|_{\mathbb{R}}$  the restriction of  $f$  to  $\mathbb{R}$ , and observe that  $f|_{\mathbb{R}} \in \mathcal{C}_{u,v}(\mathbb{R})$ , where  $u = f(-\infty)$  and  $v = f(+\infty)$ , so, by [Proposition II.4](#),  $P_t f|_{\mathbb{R}} \in \mathcal{C}_{u,v}(\mathbb{R})$ . For  $x \in \mathbb{R}$ ,  $\tilde{P}_t f(x) = P_t f|_{\mathbb{R}}(x)$ , while  $\tilde{P}_t f(-\infty) = u$  and  $\tilde{P}_t f(+\infty) = v$ . We deduce that  $\tilde{P}_t f$  is continuous on  $\overline{\mathbb{R}}$ .

We now prove (Fb'), i.e.  $\forall f \in \mathcal{C}(\overline{\mathbb{R}})$  and  $x \in \overline{\mathbb{R}}$ ,  $\lim_{t \rightarrow 0} \tilde{P}_t f(x) = f(x)$ . This is immediate for  $x = \pm\infty$  since then  $\tilde{P}_t f(x) = f(x)$  for all  $t$ . For all  $x \in \mathbb{R}$ , we have that  $P_t(x, \cdot)$  converges weakly to  $\delta_x$  as  $t \rightarrow 0$ , so that, since  $f|_{\mathbb{R}} \in \mathcal{C}_b(\mathbb{R})$ ,  $\lim_{t \rightarrow 0} P_t f|_{\mathbb{R}}(x) = f|_{\mathbb{R}}(x)$ , whence  $\lim_{t \rightarrow 0} \tilde{P}_t f(x) = f(x)$ .  $\square$

**Remark II.1.** *Observe that it is not true that, without additional assumptions, the extension to  $\overline{\mathbb{R}}$  of a Feller Markov semigroup on  $\mathbb{R}$  always inherits the Feller property (this is true nonetheless for the extension to the one-point compactification  $\mathbb{R} \cup \{\infty\}$ ). Consider for instance the semigroup  $(P_t)$  associated with a standard Brownian motion  $(B_t)_{t \in \mathbb{R}_+}$  on  $\mathbb{R}$  whose sign is reversed at each occurrence of an independent rate 1 Poisson process  $(N_t)_{t \in \mathbb{R}_+}$ , i.e.  $X_t = B_t \cdot (-1)^{N_t}$ . Indeed, for  $f \in \mathcal{C}_{a,b}(\mathbb{R})$  with  $a \neq b$ ,  $\lim_{x \rightarrow -\infty} P_t f(x) = a\mathbb{Q}(N_t \text{ is even}) + b\mathbb{Q}(N_t \text{ is odd})$ , so  $P_t f \notin \mathcal{C}_{a,b}(\mathbb{R})$  for  $t > 0$ .*

## II.3 Proof of [Theorem II.1](#)

This section is devoted to the proof of [Theorem II.1](#). First, in [Section II.3.1](#), we define the kernels  $Q_t^{(n)}$  by using, for  $\mathbf{x} = (x_1, \dots, x_n)$  and  $t \geq 0$ , the classical comonotone coupling of  $P_t(x_1, \cdot), \dots, P_t(x_n, \cdot)$ , which turns out to be order-preserving thanks to the stochastic monotonicity of  $(P_t)_{t \geq 0}$  (more precisely, we work with the extended semigroup  $(\tilde{P}_t)_{t \geq 0}$  on  $\overline{\mathbb{R}}$ ). Then, in [Section II.3.2](#), we construct the kernels  $P_t^{(n)}$  by iterating the composition of  $Q_s^{(n)}$  a large number of times for smaller and smaller  $s$ , then taking the limit. One key tool to prove that this procedure leads to a well-behaved limit is [Proposition II.7](#), which enables the transfer of regularity properties of the kernels  $P_t$  to the kernels  $Q_t^{(n)}$ , thanks to the interplay between order-preserving properties of kernels and estimates in the  $W_1^{(n)}$  distance.

### II.3.1 Discrete-time flow

For  $x \in \mathbb{R}$  and  $t \geq 0$ , let  $F_{x,t}^{[-1]} : [0, 1] \rightarrow \overline{\mathbb{R}}$  denote the quantile function of the probability distribution  $P_t(x, \cdot)$ , i.e.  $F_{x,t}^{[-1]}(u) = \inf\{y \in \mathbb{R}; P_t(x, ] - \infty, y]) \geq u\}$ . We extend the definition to  $x \in \overline{\mathbb{R}}$  by letting  $F_{-\infty,t}^{[-1]} \equiv -\infty$  and  $F_{+\infty,t}^{[-1]} \equiv +\infty$ .

**Lemma II.4.** *The family of functions  $F_{x,t}^{[-1]}$  enjoys the following properties:*

- (a) for all  $x \in \overline{\mathbb{R}}$  and  $t \geq 0$ , the map  $F_{x,t}^{[-1]}$  is non-decreasing;
- (b) for all  $x \in \mathbb{R}$ ,  $t \geq 0$  and  $u \in ]0, 1[$ ,  $F_{x,t}^{[-1]}(u) \in \mathbb{R}$ ;
- (c) for all  $t \geq 0$  and  $u \in ]0, 1[$ , the map  $x \mapsto F_{x,t}^{[-1]}(u)$  is non-decreasing;
- (d) If  $U$  is a random variable following the uniform distribution on the interval  $[0, 1]$ , then, for all  $x, t$ , the distribution of the random variable  $F_{x,t}^{[-1]}(U)$  is  $\tilde{P}_t(x, \cdot)$ ;
- (e) for all  $t \in \mathbb{R}$ , and  $x \in \overline{\mathbb{R}}$ , except for an at most countable set of values of  $u \in [0, 1]$ , the map  $y \mapsto F_{y,t}^{[-1]}(u)$  is continuous at  $x$ .

*Proof.* Properties (a) and (b) are immediate consequences of the definition and of the fact that the cdf of a real-valued random variable is non-decreasing, with limits 0 and 1 at  $-\infty$  and  $+\infty$ , respectively. Property (c) is a consequence of stochastic monotonicity: if  $x_1, x_2 \in \mathbb{R}$  are such that  $x_1 \leq x_2$ , then, for all  $y \in \mathbb{R}$ ,  $P_t(x_1, ] - \infty, y]) \geq P_t(x_2, ] - \infty, y])$ . Extension to  $\overline{\mathbb{R}}$  is immediate in view of the definition of  $F_{x,t}^{[-1]}$  when  $x = \pm\infty$ . Property (d) is obvious when  $x = \pm\infty$ , and classical when  $x \in \mathbb{R}$ , we e.g. refer to the proof of Theorem 25.6 in [Bil95]. We now deal with property (e). Given  $t \geq 0$ ,  $x \in \mathbb{R}$  and a sequence  $(x_k)_{k \geq 1}$  of real numbers such that  $\lim_{k \rightarrow +\infty} x_k = x$ , we have the weak convergence  $P_t(x_k, \cdot) \xrightarrow[k \rightarrow +\infty]{\mathbf{w}} P_t(x, \cdot)$ , so that, referring again to the proof of Theorem 25.6 in [Bil95], whenever  $F_{x,t}^{[-1]}$  is continuous at a certain  $u \in ]0, 1[$ , we have that  $\lim_{k \rightarrow +\infty} F_{x_k,t}^{[-1]}(u) = F_{x,t}^{[-1]}(u)$ . Since  $F_{x,t}^{[-1]}$  is non-decreasing, the number of points of discontinuity is at most countable. Assume now that  $x = +\infty$ . From the proof of Proposition II.4, we have that, for any  $a > 0$ ,  $\lim_{x \rightarrow +\infty} P_t(x, ] - \infty, a]) = 0$ , so that, for any  $u \in ]0, 1[$ ,  $\lim_{x \rightarrow +\infty} F_{x,t}^{[-1]}(u) = +\infty = F_{+\infty,t}^{[-1]}(u)$ . The case  $x = -\infty$  is treated symmetrically.  $\square$

Given  $t \geq 0$  and  $\mathbf{x} = (x_1, \dots, x_n) \in \overline{\mathbb{R}}^n$ , define  $Q_t^{(n)}(\mathbf{x}, \cdot)$  as the distribution on  $\overline{\mathbb{R}}^n$  of the random vector  $Z_{\mathbf{x}} = \left( F_{x_i,t}^{[-1]}(U) \right)_{1 \leq i \leq n}$ , where  $U$  is a random variable following the uniform distribution on the interval  $[0, 1]$ .

**Proposition II.6.** *For all  $t \geq 0$ , and  $n \geq 2$ ,  $Q_t^{(n)}$  is an order-preserving<sup>13</sup> Markov kernel on  $\overline{\mathbb{R}}^n$ . Moreover, for all  $t \geq 0$ , the family  $(Q_t^{(n)})_{n \geq 2}$  is a consistent extension of  $\tilde{P}_t$ .*

*Proof.* By Lemma II.4 (c), we have that, for any  $u \in [0, 1[$ ,  $\left( F_{x_i,t}^{[-1]}(u) \right)_{1 \leq i \leq n}$  is order-compatible with  $x_1, \dots, x_n$ . To check that  $Q_t^{(n)}$  is a Markov kernel, we note that, by definition, for all  $\mathbf{x}$ ,  $Q_t^{(n)}(\mathbf{x}, \cdot)$  is a probability measure on  $\overline{\mathbb{R}}^n$ . Now consider a function  $f \in \mathcal{C}(\overline{\mathbb{R}}^n)$ , and define  $Q_t^{(n)} f(\mathbf{x}) = \int_{\overline{\mathbb{R}}^n} f(\mathbf{y}) dQ_t^{(n)}(\mathbf{x}, \mathbf{y})$ . Given  $\mathbf{x} \in \overline{\mathbb{R}}^n$  and a sequence  $(\mathbf{x}_k)_{k \geq 1}$  such that  $\lim_{k \rightarrow +\infty} \mathbf{x}_k = \mathbf{x}$ , we deduce from Lemma II.4 (e) that, almost surely,  $\lim_{k \rightarrow +\infty} Z_{\mathbf{x}_k} = Z_{\mathbf{x}}$ , so that, by continuity of  $f$ , almost surely,  $\lim_{k \rightarrow +\infty} f(Z_{\mathbf{x}_k}) = f(Z_{\mathbf{x}})$ . We thus obtain by dominated convergence that  $\lim_{k \rightarrow +\infty} Q_t^{(n)} f(\mathbf{x}_k) = Q_t^{(n)} f(\mathbf{x})$ . We deduce that  $Q_t^{(n)} f \in \mathcal{C}(\overline{\mathbb{R}}^n)$ , then invoke Lemma II.1 to deduce that  $Q_t^{(n)}$  is a Markov kernel.

That  $Q_t^{(n)}$  is an extension of  $\tilde{P}_t$  is a direct consequence of Lemma II.4 (d). Consistency is a consequence of the fact that, by definition,  $\pi_{i_1, \dots, i_k}^n(Z_{\mathbf{x}}) = \left( F_{x_{i_j}, t}^{[-1]}(U) \right)_{1 \leq j \leq k} = Z_{\pi_{i_1, \dots, i_k}^n(\mathbf{x})}$ .  $\square$

### II.3.2 Construction of the limit $R_t^{(n)}$

Now we call  $D_+$  the set of positive dyadic rational numbers. For any fixed  $t \in D_+$ , we write  $t = k2^{-m_0} \in D_+$  where  $k \geq 1$  and  $m_0 \geq 0$  are integers, and  $m_0$  has the minimum possible value in such an expression. Then, for every integer  $m \geq m_0$ , we let

$$Q_t^{(n),m} = \left[ Q_{2^{-m}}^{(n)} \right]^{k2^{m-m_0}},$$

<sup>13</sup>We extend the definition of an order-preserving Markov kernel on  $\mathbb{R}^n$  to that of an order-preserving Markov kernel on  $\overline{\mathbb{R}}^n$  in the obvious way.

i.e.  $Q_t^{(n),m}$  is the repeated composition of kernels  $Q_{2^{-m}}^{(n)} \cdots Q_{2^{-m}}^{(n)}$  with a total of  $k2^{m-m_0}$  kernels in the composition.

As a result,  $(Q_t^{(n),m})_{m \geq m_0}$  is a sequence of Markov kernels on  $\overline{\mathbb{R}}^n$ . Moreover, from [Proposition II.6](#) for all  $t \in D_+$ ,  $n \geq 2$ , and  $m \geq m_0$ ,  $Q_t^{(n),m}$  is order-preserving<sup>14</sup>, and, for all  $t \in D_+$  and  $m \geq m_0$ , the family  $(Q_t^{(n),m})_{n \geq 2}$  is a consistent extension<sup>15</sup> of  $\tilde{P}_t$ . Now let  $\mathbf{Q}_t^{(n),m} = \overline{Q_t^{(n),m}}$ .

Denote by  $W_1^{(n)}$  the Wasserstein (or Kantorovich-Rubinstein) distance on  $\mathcal{P}(\overline{\mathbb{R}}^n)$ , where  $\overline{\mathbb{R}}^n$  is equipped with the distance  $d_n$  defined in [Section II.2.2](#).

**Proposition II.7.** *For all  $m \geq m_0$ , the map  $\mathbf{Q}_t^{(n),m}$  from  $\overline{\mathbb{R}}^n$  to  $\mathcal{P}(\overline{\mathbb{R}}^n)$  satisfies:*

$$W_1^{(n)}\left(\mathbf{Q}_t^{(n),m}(\mathbf{x}), \mathbf{Q}_t^{(n),m}(\mathbf{y})\right) \leq \sum_{i=1}^n |\tilde{P}_t \phi(y_i) - \tilde{P}_t \phi(x_i)|.$$

*Proof.* Let  $\mathbf{z} = (\mathbf{x}, \mathbf{y}) = (x_1, \dots, x_n, y_1, \dots, y_n)$ . Let  $(X_1, \dots, X_n, Y_1, \dots, Y_n)$  be a random vector on  $\overline{\mathbb{R}}^n$  whose distribution is  $Q_t^{(2n),m}(\mathbf{z}, \cdot)$ . By the consistency property, the distribution of  $(X_1, \dots, X_n)$  is  $Q_t^{(n),m}(\mathbf{x}, \cdot)$  and the distribution of  $(Y_1, \dots, Y_n)$  is  $Q_t^{(n),m}(\mathbf{y}, \cdot)$ . Moreover, the order-preserving property implies that, with probability 1, for all  $1 \leq i \leq n$ , the relative order of  $X_i$  and  $Y_i$  is the same<sup>16</sup> as the relative order between  $x_i$  and  $y_i$ . Now consider a function  $f \in \text{Lip}(\overline{\mathbb{R}}^n)$  such that  $\|f\|_{\text{Lip}} \leq 1$ . Let  $\mathbf{X} = (X_1, \dots, X_n)$  and  $\mathbf{Y} = (Y_1, \dots, Y_n)$ . We have that

$$\left|Q_t^{(n),m}(\mathbf{y}, \cdot) - Q_t^{(n),m}(\mathbf{x}, \cdot)\right| = |\mathbb{E}f(\mathbf{Y}) - \mathbb{E}f(\mathbf{X})| \leq \mathbb{E}|f(\mathbf{Y}) - f(\mathbf{X})|.$$

Since  $\|f\|_{\text{Lip}} \leq 1$ ,

$$\mathbb{E}|f(\mathbf{Y}) - f(\mathbf{X})| \leq \mathbb{E} \sum_{i=1}^n |\phi(Y_i) - \phi(X_i)| = \sum_{i=1}^n \mathbb{E}|\phi(Y_i) - \phi(X_i)|.$$

Since  $\phi$  is non-decreasing and the relative order of  $X_i$  and  $Y_i$  is non-random, we have that

$$\mathbb{E}|\phi(Y_i) - \phi(X_i)| = |\mathbb{E}(\phi(Y_i) - \phi(X_i))| = |\mathbb{E}\phi(Y_i) - \mathbb{E}\phi(X_i)|.$$

Since  $Q_t^{(n),m}$  is an extension of  $\tilde{P}_t$ , we have that  $\mathbb{E}\phi(X_i) = \tilde{P}_t \phi(x_i)$  and  $\mathbb{E}\phi(Y_i) = \tilde{P}_t \phi(y_i)$ .  $\square$

**Corollary II.1.** *For all  $n \geq 2$  and  $t \in D_+$ , the family  $(\mathbf{Q}_t^{(n),m})_{m \geq m_0}$  of maps from  $\overline{\mathbb{R}}^n$  to  $\mathcal{P}(\overline{\mathbb{R}}^n)$  is equicontinuous.*

*Proof.* Note that  $\phi \in \mathcal{C}(\overline{\mathbb{R}})$ , so the Feller property of  $\tilde{P}_t$  implies that  $\tilde{P}_t \phi$  is continuous.  $\square$

Denote by  $\mathfrak{C}_n$  the set of continuous functions from  $\overline{\mathbb{R}}^n$  to  $\mathcal{P}(\overline{\mathbb{R}}^n)$ , equipped with the distance

$$d_{\mathfrak{C}_n}(\mathbf{K}, \mathbf{L}) = \sup_{x \in \overline{\mathbb{R}}^n} W_1^{(n)}(\mathbf{K}(x), \mathbf{L}(x)).$$

Remember that  $\overline{\mathbb{R}}^n$  is a compact metric space, so that  $\mathcal{P}(\overline{\mathbb{R}}^n)$  is also a compact metric space, hence a complete metric space. As a consequence, for each  $t \in D_+$  and  $n \geq 2$ , in view of [Corollary II.1](#), we can invoke the Arzelà-Ascoli theorem (see e.g. [Dix84](#)) to show the convergence

<sup>14</sup>Using the fact that the composition of order-preserving Markov kernels is still an order-preserving Markov kernel, see [Lemma C.1](#) in [Section C.2](#).

<sup>15</sup>Using the fact that, if  $(K^{(n)})_{n \geq 2}$  and  $(L^{(n)})_{n \geq 2}$  are consistent extensions, respectively of  $K$  and  $L$ , then  $(K^{(n)}L^{(n)})_{n \geq 2}$  is a consistent extension of  $KL$ , see [Lemma C.2](#) in [Section C.2](#).

<sup>16</sup>To be explicit: if  $x_i \leq y_i$ , then almost surely  $X_i \leq Y_i$ , while, if  $y_i \leq x_i$ , then almost surely  $Y_i \leq X_i$ .

along a subsequence: the sequence of maps  $(\mathbf{Q}_t^{(n),m_k})_{k \geq 1}$  converges, as  $k \rightarrow +\infty$ , to a limiting map  $\mathbf{R}_t^{(n)}$ , in the sense of uniform convergence of continuous maps from  $\overline{\mathbb{R}}^n$  to  $\mathcal{P}(\overline{\mathbb{R}}^n)$ . Moreover, by diagonal extraction, we can assume that convergence occurs simultaneously for every  $t$  in (the countable set)  $D_+$  and every  $n \geq 2$ . To sum up, we have a strictly increasing sequence of integers  $(m_k)_{k \geq 1}$  and, for every  $t \in D_+$  and  $n \geq 2$ , an element  $\mathbf{R}_t^{(n)}$  of  $\mathfrak{C}_n$ , such that<sup>17</sup>

$$\lim_{k \rightarrow +\infty} d_{\mathfrak{C}_n}(\mathbf{Q}_t^{(n),m_k}, \mathbf{R}_t^{(n)}) = 0. \quad (\text{II.10})$$

Now let  $R_t^{(n)} = \widehat{\mathbf{R}_t^{(n)}}$ , and note that  $R_t^{(n)}$  is a Markov kernel (e.g. since  $\mathbf{R}_t^{(n)}$  is a continuous map from  $\overline{\mathbb{R}}^n$  to  $\mathcal{P}(\overline{\mathbb{R}}^n)$ , see [Lemma II.1](#)), and that, by continuity of the projection maps  $\pi_{i_1, \dots, i_k}^n$ , for each  $t \in D_+$ , the family of Markov kernels  $(R_t^{(n)})_{n \geq 2}$  inherits the property of being a family of order-preserving kernels forming a consistent extension of  $\tilde{P}_t$ .

Let us now check that the semigroup property (restricted to  $D_+$ ) holds for our family of Markov kernels:

$$\forall s, t \in D_+, R_{s+t}^{(n)} = R_s^{(n)} R_t^{(n)}. \quad (\text{II.11})$$

By construction, given  $s, t \in D_+$ , we have that  $Q_s^{(n),m_k} Q_t^{(n),m_k} = Q_{s+t}^{(n),m_k}$  for all large enough  $k$ . Now consider  $f \in \text{Lip}(\overline{\mathbb{R}}^n)$  such that  $\|f\|_{\text{Lip}} \leq 1$ , and write

$$Q_s^{(n),m_k} Q_t^{(n),m_k} f - R_s^{(n)} R_t^{(n)} f = Q_s^{(n),m_k} Q_t^{(n),m_k} f - Q_s^{(n),m_k} R_t^{(n)} f + Q_s^{(n),m_k} R_t^{(n)} f - R_s^{(n)} R_t^{(n)} f.$$

Using [Lemma II.2](#), we have that, for all  $\mathbf{x} \in \overline{\mathbb{R}}^n$ ,

$$\left| Q_s^{(n),m_k} Q_t^{(n),m_k} f(\mathbf{x}) - Q_s^{(n),m_k} R_t^{(n)} f(\mathbf{x}) \right| \leq d_{\mathfrak{C}_n}(\mathbf{Q}_t^{(n),m_k}, \mathbf{R}_t^{(n)}). \quad (\text{II.12})$$

On the other hand,  $R_t^{(n)} f \in \mathcal{C}(\overline{\mathbb{R}}^n)$  since  $f \in \mathcal{C}(\overline{\mathbb{R}}^n)$  and  $\mathbf{R}_t^{(n)}$  is a continuous map from  $\overline{\mathbb{R}}^n$  to  $\mathcal{P}(\overline{\mathbb{R}}^n)$  (see [Proposition II.1](#)). Since, for every  $\mathbf{x} \in \overline{\mathbb{R}}^n$ , we have the weak convergence  $Q_s^{(n),m_k}(\mathbf{x}, \cdot) \xrightarrow[k \rightarrow +\infty]{\mathbf{w}} R_s^{(n)}(\mathbf{x}, \cdot)$ , we deduce that

$$\lim_{k \rightarrow +\infty} Q_s^{(n),m_k} R_t^{(n)} f(\mathbf{x}) = R_s^{(n)} R_t^{(n)} f(\mathbf{x}). \quad (\text{II.13})$$

Combining [\(II.10\)](#), [\(II.12\)](#) and [\(II.13\)](#), we have that  $\lim_{k \rightarrow +\infty} Q_s^{(n),m_k} Q_t^{(n),m_k} f(\mathbf{x}) = R_s^{(n)} R_t^{(n)} f(\mathbf{x})$ , for all  $\mathbf{x} \in \overline{\mathbb{R}}^n$ . Now the weak convergence  $Q_{s+t}^{(n),m_k}(\mathbf{x}, \cdot) \xrightarrow[k \rightarrow +\infty]{\mathbf{w}} R_{s+t}^{(n)}(\mathbf{x}, \cdot)$  implies that we have  $\lim_{k \rightarrow +\infty} Q_{s+t}^{(n),m_k} f(\mathbf{x}) = R_{s+t}^{(n)} f(\mathbf{x})$ . Using the identity  $Q_s^{(n),m_k} Q_t^{(n),m_k} = Q_{s+t}^{(n),m_k}$ , we deduce that  $\lim_{k \rightarrow +\infty} Q_s^{(n),m_k} Q_t^{(n),m_k} f(\mathbf{x}) = R_{s+t}^{(n)} f(\mathbf{x})$ , so that  $R_s^{(n)} R_t^{(n)} f(\mathbf{x}) = R_{s+t}^{(n)} f(\mathbf{x})$ , and [\(II.11\)](#) is established<sup>18</sup>.

Now for  $\mathbf{x} = (x_1, \dots, x_n)$  and  $\varepsilon > 0$ , observe that  $B_{\overline{\mathbb{R}}^n}(\mathbf{x}, \varepsilon)^c \subset \bigcup_{i=1}^n B_{\overline{\mathbb{R}}}(x_i, \varepsilon/n)^c$ , so that, using the fact that  $R_t^{(n)}$  is an extension of  $\tilde{P}_t$  and the union bound,

$$R_t^{(n)}(\mathbf{x}, B_{\overline{\mathbb{R}}^n}(\mathbf{x}, \varepsilon)^c) \leq \sum_{i=1}^n \tilde{P}_t(x_i, B_{\overline{\mathbb{R}}}(x_i, \varepsilon/n)^c).$$

<sup>17</sup>Note that  $\mathbf{Q}_t^{(n),m_k}$  is defined as soon as  $m_k \geq m_0$ , where  $m_0$  depends on  $t$ .

<sup>18</sup>Indeed, we have established that, for all  $\mathbf{x} \in \overline{\mathbb{R}}^n$ ,  $W_1^{(n)}(R_s^{(n)} R_t^{(n)}(\mathbf{x}, \cdot), R_{s+t}^{(n)}(\mathbf{x}, \cdot)) = 0$ .

Using the Feller property of  $(\tilde{P}_t)_{t \geq 0}$  and Proposition II.3 (ii), we deduce that  $R_t^{(n)}(\mathbf{x}, B_{\mathbb{R}^n}(\mathbf{x}, \varepsilon)^c)$  goes to 0 as  $t$  goes to 0, uniformly over  $\mathbf{x} \in \overline{\mathbb{R}^n}$ , which, thanks to Proposition II.3 again, shows that

$$\lim_{\substack{t \rightarrow 0 \\ t \in D_+}} \sup_{x \in \overline{\mathbb{R}^n}} W_1^{(n)}(R_t^{(n)}(x, \cdot), \delta_x) = 0. \quad (\text{II.14})$$

Using Lemma II.2, we have that

$$\sup_{x \in \overline{\mathbb{R}^n}} W_1^{(n)}(R_{s+t}^{(n)}(x, \cdot), R_s^{(n)}(x, \cdot)) \leq \sup_{x \in \overline{\mathbb{R}^n}} W_1^{(n)}(R_t^{(n)}(x, \cdot), \delta_x). \quad (\text{II.15})$$

Combining (II.14) and (II.15) shows that the map  $s \mapsto \mathbf{R}_s^{(n)}$  from  $D_+$  to  $\mathfrak{C}_n$  is uniformly continuous. Since  $\mathfrak{C}_n$  is a complete metric space, and  $D_+$  is a dense subset of  $\mathbb{R}_+$ , there is a unique extension (see Dix84) to a uniformly continuous map  $s \mapsto \mathbf{R}_s^{(n)}$  from  $\mathbb{R}_+$  to  $\mathfrak{C}_n$ . This allows us to extend the definition  $R_t^{(n)} = \widehat{\mathbf{R}_t^{(n)}}$  to every  $t \in \mathbb{R}_+$ .

We now check that the semigroup property holds for  $(R_t)_{t \in \mathbb{R}_+}$ , i.e.

$$\forall s, t \in \mathbb{R}_+, R_{s+t}^{(n)} = R_s^{(n)} R_t^{(n)}. \quad (\text{II.16})$$

Let  $(s_k)_{k \geq 1}$  and  $(t_k)_{k \geq 1}$  be sequences of elements in  $D_+$  which converge to  $s$  and  $t$  respectively. By (II.11), we have that  $R_{s_k}^{(n)} R_{t_k}^{(n)} = R_{s_k+t_k}^{(n)}$ . Now, for  $f \in \text{Lip}(\overline{\mathbb{R}^n})$  such that  $\|f\|_{\text{Lip}} \leq 1$ , we write

$$R_{s_k}^{(n)} R_{t_k}^{(n)} f - R_s^{(n)} R_t^{(n)} f = R_{s_k}^{(n)} R_{t_k}^{(n)} f - R_{s_k}^{(n)} R_t^{(n)} f + R_{s_k}^{(n)} R_t^{(n)} f - R_s^{(n)} R_t^{(n)} f.$$

Using Lemma II.2, we have that, for all  $\mathbf{x} \in \overline{\mathbb{R}^n}$ ,

$$\left| R_{s_k}^{(n)} R_{t_k}^{(n)} f(\mathbf{x}) - R_{s_k}^{(n)} R_t^{(n)} f(\mathbf{x}) \right| \leq d_{\mathfrak{C}_n}(\mathbf{R}_{t_k}^{(n)}, \mathbf{R}_t^{(n)}).$$

Arguing exactly as in the proof of (II.11), we have that  $R_t^{(n)} f \in \mathcal{C}(\overline{\mathbb{R}^n})$ . Moreover, by continuity of the extension to  $\mathbb{R}_+$ , we have the weak convergence  $R_{s_k}^{(n)}(\mathbf{x}, \cdot) \xrightarrow[k \rightarrow +\infty]{\mathbf{w}} R_s^{(n)}(\mathbf{x}, \cdot)$ , so that

$$\lim_{k \rightarrow +\infty} R_{s_k}^{(n)} R_t^{(n)} f(\mathbf{x}) = R_s^{(n)} R_t^{(n)} f(\mathbf{x}),$$

and similarly

$$\lim_{k \rightarrow +\infty} R_{s_k+t_k}^{(n)} f(\mathbf{x}) = R_{s+t}^{(n)} f(\mathbf{x}),$$

so that  $R_{s+t}^{(n)} f(\mathbf{x}) = R_s^{(n)} R_t^{(n)} f(\mathbf{x})$ , and (II.16) is established.

We now observe that, by continuity of the projection maps  $\pi_{i_1, \dots, i_k}^n$ , and the Feller property of  $(\tilde{P}_t)_{t \in \mathbb{R}_+}$ , the property that  $(R_t^{(n)})_{n \geq 2}$  is a family of order-preserving kernels forming a consistent extension of  $\tilde{P}_t$ , already established for every  $t \in D_+$ , holds for every  $t \in \mathbb{R}_+$ . Also, thanks to the continuity of  $t \mapsto \mathbf{R}_t^{(n)}$  at  $t = 0$ , and by Proposition II.3, the family of Markov kernels  $(R_t^{(n)})_{n \geq 2}$  satisfies property (Fb).

For  $n \geq 2$ ,  $t \in \mathbb{R}_+$ ,  $\mathbf{x} \in \mathbb{R}^n$  and Borel set  $B$  of  $\mathbb{R}^n$ , we define  $P_t^{(n)}(\mathbf{x}, B) = R_t^{(n)}(\mathbf{x}, B)$ . Since  $R_t^{(n)}$  is an extension of  $\tilde{P}_t$  and  $\tilde{P}_t(x_i, \mathbb{R}) = 1$ , we have that  $R_t^{(n)}(\mathbf{x}, \mathbb{R}^n) = 1$ , so  $P_t^{(n)}$  is indeed a Markov kernel on  $\mathbb{R}$ , and the fact that, for all  $t$ ,  $(P_t^{(n)})_{n \geq 2}$  is a consistent extension of  $P_t$  by order-preserving Markov kernels is an immediate consequence of the fact that  $(R_t^{(n)})_{n \geq 2}$  is a consistent extension of  $\tilde{P}_t$  by order-preserving Markov kernels. Since a function  $f \in \mathcal{C}_0(\mathbb{R}^n)$  immediately extends to a function  $\tilde{f} \in \mathcal{C}(\overline{\mathbb{R}^n})$ , we deduce the fact that  $P_t^{(n)} f \in \mathcal{C}_b(\mathbb{R}^n)$  and



the convergence  $\lim_{t \rightarrow 0} \left\| P_t^{(n)} f - f \right\|_\infty = 0$  from the corresponding properties of  $R_t^{(n)}$ . To complete the proof of the Feller property of  $(P_t^{(n)})_{t \in \mathbb{R}_+}$ , it remains to prove that  $P_t^{(n)} f \in \mathcal{C}_0(\mathbb{R}^n)$ , i.e.  $\lim_{\mathbf{x} \rightarrow \infty} P_t^{(n)} f(\mathbf{x}) = 0$ , since we already know that  $P_t^{(n)} f \in \mathcal{C}_b(\mathbb{R}^n)$ . To this end, using the fact that  $P_t^{(n)}$  is an extension of  $P_t$ , we have that, for all  $a > 0$ :

$$P_t^{(n)}(\mathbf{x}, [-a, +a]^n) \leq \min_{1 \leq i \leq n} P_t(x_i, [-a, +a]).$$

Thanks to [Proposition II.2](#) and the Feller property of  $P_t$ , we deduce that

$$\lim_{\mathbf{x} \rightarrow \infty} P_t^{(n)}(\mathbf{x}, [-a, +a]^n) = 0,$$

and, thanks to [Proposition II.2](#) again, that  $\lim_{\mathbf{x} \rightarrow \infty} P_t^{(n)} f(\mathbf{x}) = 0$ .

## II.4 Proof of [Theorem II.2](#)

This (short) section is devoted to the proof of [Theorem II.2](#). We start with a general result showing that the  $\leq_{\text{sm}}$  order is preserved by consistent families of order-preserving Markov kernels. Combined with the maximal property of the comonotone coupling with respect to  $\leq_{\text{sm}}$  and the fact that  $(\check{P}_t^{(n)})_{t \in \mathbb{R}_+}$  is obtained as a limit of the composition of order-preserving kernels constructed using the comonotone coupling, we deduce [Theorem II.2](#).

**Proposition II.8.** *If  $(K^{(n)})_{n \geq 1}$  is a consistent family of order-preserving Markov kernels on the successive powers of  $\mathbb{R}$ , then, for all  $n \geq 2$  and  $f \in \mathcal{SM}_b(\mathbb{R}^n)$ ,  $K^{(n)} f \in \mathcal{SM}_b(\mathbb{R}^n)$ .*

*Proof.* Consider  $\mathbf{x} = (x_1, \dots, x_n)$  and  $\mathbf{y} = (y_1, \dots, y_n)$ . Let  $\mathbf{z} = (\mathbf{x}, \mathbf{y}) = (x_1, \dots, x_n, y_1, \dots, y_n)$ . Let  $(X_1, \dots, X_n, Y_1, \dots, Y_n)$  be a random vector on  $\overline{\mathbb{R}}^n$  whose distribution is  $K^{(2n)}(\mathbf{z}, \cdot)$ , and let  $\mathbf{X} = (X_1, \dots, X_n)$  and  $\mathbf{Y} = (Y_1, \dots, Y_n)$ . By the consistency property, the distribution of  $\mathbf{X}$  is  $K^{(n)}(\mathbf{x}, \cdot)$  and the distribution of  $\mathbf{Y}$  is  $K^{(n)}(\mathbf{y}, \cdot)$ , so that  $K^{(n)} f(\mathbf{x}) = \mathbb{E} f(\mathbf{X})$  and  $K^{(n)} f(\mathbf{y}) = \mathbb{E} f(\mathbf{Y})$ . Now, for  $1 \leq i \leq n$ , let  $Z_i^+ = X_i$  and  $Z_i^- = Y_i$  when  $x_i \leq y_i$ ,  $Z_i^+ = Y_i$  and  $Z_i^- = X_i$  when  $x_i > y_i$ , so that, thanks to the consistency property again, the distribution of  $(Z_1^+, \dots, Z_n^+)$  is  $K^{(n)}(\mathbf{x} \vee \mathbf{y}, \cdot)$  and the distribution of  $(Z_1^-, \dots, Z_n^-)$  is  $K^{(n)}(\mathbf{x} \wedge \mathbf{y}, \cdot)$ , so that  $K^{(n)} f(\mathbf{x} \vee \mathbf{y}) = \mathbb{E} f(Z_1^+, \dots, Z_n^+)$  and  $K^{(n)} f(\mathbf{x} \wedge \mathbf{y}) = \mathbb{E} f(Z_1^-, \dots, Z_n^-)$ . Moreover, by the order-preserving property of  $K^{(2n)}$ , we have that  $\mathbf{X} \vee \mathbf{Y} = (Z_1^+, \dots, Z_n^+)$  and  $\mathbf{X} \wedge \mathbf{Y} = (Z_1^-, \dots, Z_n^-)$ . By the super-modularity property of  $f$ ,  $f(\mathbf{X} \vee \mathbf{Y}) + f(\mathbf{X} \wedge \mathbf{Y}) \geq f(\mathbf{X}) + f(\mathbf{Y})$ , and we thus deduce that  $f(Z_1^+, \dots, Z_n^+) + f(Z_1^-, \dots, Z_n^-) \geq f(\mathbf{X}) + f(\mathbf{Y})$ . Taking expectations, we deduce that  $K^{(n)} f(\mathbf{x} \vee \mathbf{y}) + K^{(n)} f(\mathbf{x} \wedge \mathbf{y}) \geq K^{(n)} f(\mathbf{x}) + K^{(n)} f(\mathbf{y})$ , so we have proved that  $K^{(n)} f$  is super-modular.  $\square$

**Proposition II.9.** *Let  $(M_t^{(n)})_{t \geq 0} \in \mathfrak{M}_n$ , and let  $(Q_t^{(n)})_{t \geq 0}$  be the family of Markov kernels introduced in [Section II.3.1](#). Then, for all  $t \geq 0$  and  $\mathbf{x} \in \mathbb{R}^n$ ,  $M_t^{(n)}(\mathbf{x}, \cdot) \leq_{\text{sm}} Q_t^{(n)}(\mathbf{x}, \cdot)$ .*

*Proof.* By construction (see Theorem 3.1.1 in [\[MS02\]](#)),  $Q_t^{(n)}(\mathbf{x}, \cdot)$  is a comonotone probability distribution on  $\mathbb{R}^n$ . Moreover,  $Q_t^{(n)}(\mathbf{x}, \cdot)$  and  $M_t^{(n)}(\mathbf{x}, \cdot)$  have the same marginals, namely the distributions  $P_t(x_1, \cdot), \dots, P_t(x_n, \cdot)$ . By Theorem 3.9.8 in [\[MS02\]](#) (using property (P5)), we deduce that  $M_t^{(n)}(\mathbf{x}, \cdot) \leq_{\text{sm}} Q_t^{(n)}(\mathbf{x}, \cdot)$ .  $\square$

*Proof of [Theorem II.2](#)* We now prove by induction that, for all  $f \in \mathcal{SM}_b(\mathbb{R}^n)$ ,  $s \geq 0$ , and integer  $k \geq 0$ ,

$$M_{sk}^{(n)} f \leq \left[ Q_s^{(n)} \right]^k f. \quad (\text{II.17})$$

For  $k = 0$ , this is obvious since the property to be proved is that  $f \leq f$ . Now write  $\left[Q_s^{(n)}\right]^{k+1} f = Q_s^{(n)} \left[Q_s^{(n)}\right]^k f$ . Thanks to Proposition II.8,  $\left[Q_s^{(n)}\right]^k f$  is super-modular. We deduce from Proposition II.9 that  $M_s^{(n)} \left[Q_s^{(n)}\right]^k f \leq Q_s^{(n)} \left[Q_s^{(n)}\right]^k f = \left[Q_s^{(n)}\right]^{k+1} f$ . Assuming that  $M_{s_k}^{(n)} f \leq \left[Q_s^{(n)}\right]^k f$ , positivity implies that  $M_{s_{(k+1)}}^{(n)} f = M_s^{(n)} M_{s_k}^{(n)} f \leq M_s^{(n)} \left[Q_s^{(n)}\right]^k f$ , so that, combining the previous inequalities, we deduce that  $M_{s_{(k+1)}}^{(n)} f \leq \left[Q_s^{(n)}\right]^{k+1} f$ , and (II.17) is proved by induction.

Assume that  $f$  is continuous in addition to being in  $\mathcal{SM}_b(\mathbb{R}^n)$ . Given  $t \in D_+$ , remember from Section II.3.2 that, for  $m \geq m_0$ ,  $Q_t^{(n),m} = \left[Q_{2^{-m}}^{(n)}\right]^{k2^{m-m_0}}$ , where  $t = k2^{-m_0} \in D_+$ , and  $k \geq 1$  and  $m_0 \geq 0$  are integers. As a consequence, (II.17) implies that  $M_t^{(n)} f \leq Q_t^{(n),m} f$ . Remember that  $\check{P}_t^{(n)}$  is defined as the limit for  $k \rightarrow +\infty$  of a subsequence  $\left(Q_t^{(n),m_k}\right)_{k \geq 1}$ . Since  $f$  is continuous, we have, by weak convergence, that, as  $k \rightarrow +\infty$ ,  $Q_t^{(n),m_k} f$  converges pointwise to  $\check{P}_t^{(n)} f$ , so that, taking the limit in our inequality, we have that  $M_t^{(n)} f \leq \check{P}_t^{(n)} f$ . Given  $t \in \mathbb{R}_+$ , consider a sequence  $(s_m)_{m \geq 1}$  of elements of  $D_+$  such that  $\lim_{m \rightarrow +\infty} s_m = t$ . By the Feller property of  $(M_s^{(n)})_{s \in \mathbb{R}_+}$  and  $(\check{P}_s^{(n)})_{s \in \mathbb{R}_+}$ , we have that, as  $m \rightarrow +\infty$ ,  $M_{s_m}^{(n)} f$  converges pointwise to  $M_t^{(n)} f$ , and  $\check{P}_{s_m}^{(n)} f$  converges pointwise to  $\check{P}_t^{(n)} f$ , so that, taking the limit in the inequality  $M_{s_m}^{(n)} f \leq \check{P}_{s_m}^{(n)} f$ , we have that  $M_t^{(n)} f \leq \check{P}_t^{(n)} f$ . Using [MS02] (Theorems 3.9.10 and 3.9.11), Theorem II.2 is proved.  $\square$

## Chapter III

# Deep Learning methods for Mean Field Optimal Transport problem

This chapter contains the main part of [BFL+23], that can be found at <https://arxiv.org/abs/2302.14739>. Appendix D contains some additional details to this work.

### III.1 Introduction

Mean field games (MFGs) have been introduced by Lasry and Lions [LL06a, LL06b, LL07] and Caines, Huang and Malhamé [HMC06, HCM07] to approximate Nash equilibria in games with a very large number of players. At a high level, the main idea is to use a mean field approximation to represent the state of the population, and then to focus on the interactions between a single representative player and the distribution of the states of the other players. Mean field control (MFC) [BFY13] relies on a similar approximation but aims at representing situations in which a large number of agents cooperate to minimize a common social cost. The problem can be interpreted as an optimal control problem for a McKean-Vlasov (MKV) stochastic differential equation (SDE) or an optimal control for a Kolmogorov-Fokker-Planck (KFP) partial differential equation (PDE). In the past decade, the analysis of both MFGs and MFC problems has been extensively developed, see e.g. [BFY13] for an introduction to this topic, and [CD18] for a probabilistic viewpoint.

In the most common setup, the players try to minimize a total cost which is composed of a running cost integrated over time and a terminal cost. These costs generally account for the efforts made to control the dynamics as well as the preferences for some states over others. Another class of models has been introduced, in which there is no terminal cost and instead the terminal distribution of the population is imposed as a constraint. Nash equilibria have been studied under the name of planning problem for mean field game. This class of problems has been analyzed mostly using PDE-based techniques [ACCD12, Por13, Por14, OPS19, GMST19, BLL21]. In the special case of linear dynamics and a quadratic running cost in the control, the problem is related to the Schrödinger bridge problem, and equilibrium conditions can be phrased in terms of ordinary differential equations (ODEs) [CGPI5a, CGPI6, CGPI5b, CGPI8a, CGPI8b]. Mean-field Schrödinger bridge problems have also received a growing interest from the stochastic viewpoint, see e.g. [BCGL20, HT23].

This research direction is tightly connected to optimal transport (OT). Benamou and Brenier proposed in [BB00] a fluid mechanics framework for the  $L^2$  Monge-Kantorovich mass transport problem. Many works built on this approach to relate optimal transport and optimal control problems for continuity equations. Of particular interest for MFGs is the work [CCN13], which

clarified the link between geodesics for a class of distances between probability measures and a PDE system similar to the one arising in MFGs. For more background on OT, we refer the interested reader to the monographs [Vil09b, Vil21, San15, PC19, AGS05]. However, the solutions of MFGs correspond to Nash equilibria, and hence, in general, MFGs do not admit a variational structure. Furthermore, in many applications, it is not immediately clear to us why selfish players caring only about their individual costs would manage to agree and reach a target terminal distribution. Imposing a fixed terminal distribution seems more natural in the MFC setting, where the agents behave in a cooperative way to minimize the social cost. In the present work, we focus on such MFC with planning problems, in which a mean field of agents try to collectively minimize a social cost while ensuring that a fixed distribution is attained at the terminal time.

Since the work of Benamou and Brenier [BB00], several numerical methods have been investigated for similar problems, including MFGs with planning. Achdou *et al.* have proposed in [ACCD12] a method based on finite differences and Newton method to solve the PDE system of MFG with planning. Benamou and Carlier have used in [BC15, BCL16] an Augmented Lagrangian method approach with the alternating direction method of multipliers to solve OT and MFG (without planning). Similar methods have been used in [And17, AL16b] to solve MFGs and MFC problems (still without planning). Benamou *et al.* proposed in [BCDMN19] a method to solve MFG with planning through entropy regularization and Sinkhorn algorithm.

Recently, several deep learning methods have been proposed to solve high-dimensional optimal control problems and PDEs, such as the DeepBSDE method [HJW18, HJW17, HHL22], the Deep Galerkin Method [SS18] and physics-informed neural networks [RPK19]. Some of these methods have been extended to MFGs and MFC problems. In particular, [AACN<sup>+</sup>18, CL19] proposed deep learning methods to solve the PDE systems arising in mean field problems, [CL21a, GMW22, FZ20] introduced deep learning methods for differential MFC problems. Ruthotto *et al.* introduced a deep learning method for variational MFG with degenerate diffusion in [ROL<sup>+</sup>20]. Lin *et al.* introduced a deep learning method in [LFL<sup>+</sup>20] that utilizes the primal-dual relationship of variational MFG. Cao *et al.* noticed a connection between MFGs, generative adversarial networks and OT in [CGL20]. We refer the interested reader to e.g. [CL21b, GPW21, HL22] for recent surveys on this topic. The work most related to ours is the work of Liu *et al.* in [LCST22], where they considered the planning problems in a class of MFGs based on a generalized version of the Schrödinger bridge problem and proposed a neural network-based numerical method to solved it.

The main goal of this paper is to propose numerical methods based on deep learning to solve MFC problems with planning constraint, that we will call mean field optimal transport problems. To the best of our knowledge, the theory remains to be investigated in detail, and this is beyond the scope of the present work. Here, we proceed formally when needed, and we focus on the numerical aspects using machine learning tools. The rest of the paper is organized as follows. In [Section III.2](#), we introduce the problem and discuss several examples. In [Section III.3](#) we describe three numerical methods, each based on a different approach for the problem. In [Section III.4](#), we present numerical results on several benchmark problems.

## III.2 Definition of the problem

Before presenting the mean field optimal transport problem, let us first recall the definition of a typical mean field control problem. Let  $T$  be a time horizon. Let  $\mathcal{Q} = \mathbb{R}^d$  and  $\mathcal{Q}_T = [0, T] \times \mathcal{Q}$  denote the space domain and the time-space domain. Denote by  $\mathcal{P}_2(\mathcal{Q})$  the set of square-integrable probability measures on  $\mathcal{Q}$ . Let  $f : \mathcal{Q} \times \mathcal{P}_2(\mathcal{Q}) \times \mathbb{R}^k \rightarrow \mathbb{R}$  be a running cost function,  $g : \mathcal{Q} \times \mathcal{P}_2(\mathcal{Q}) \rightarrow \mathbb{R}$  be a terminal cost function,  $b : \mathcal{Q} \times \mathcal{P}_2(\mathcal{Q}) \times \mathbb{R}^k \rightarrow \mathbb{R}^d$  be a drift function and  $\sigma \in \mathbb{R}$  be a non-negative constant diffusion coefficient. In a classical MFC problem with given initial distribution  $\rho_0$  in  $\mathcal{P}_2(\mathcal{Q})$ , the goal is to find a feedback control  $v^* : \mathcal{Q}_T \rightarrow \mathbb{R}^k$  minimizing:

$$J^{MFC} : v \mapsto \mathbb{E} \left[ \int_0^T f(X_t^v, \mu^v(t), v(t, X_t^v)) dt + g(X_T^v, \mu^v(T)) \right] \quad (\text{III.1})$$

where  $\mu^v(t)$  is the distribution of  $X_t^v$ , under the constraint that the process  $X^v = (X_t^v)_{t \geq 0}$  solves the SDE

$$\begin{cases} X_0^v \sim \rho_0 \\ dX_t^v = b(X_t^v, \mu^v(t), v(t, X_t^v)) dt + \sigma dW_t, \quad t \geq 0, \end{cases} \quad (\text{III.2})$$

where  $W$  is a standard  $d$ -dimensional Brownian motion. It would also be interesting to consider open-loop controls, but since we are motivated by numerical applications, we restrict our attention to feedback controls. The cost [\(III.1\)](#) can be interpreted either as the expected cost for a single representative player, or as the average cost for the whole population, which we refer to as the social cost.

In this work, we are interested in a modified version of the above problem, where instead of having a terminal cost, a terminal distribution is imposed. This type of problem encompasses optimal transport as a special case, but it may incorporate mean field interactions in the drift and the running cost. For this reason, we will refer to this class of problems as mean field optimal transport (MFOT for short)<sup>1</sup>. Given two distributions  $\rho_0$  and  $\rho_T \in \mathcal{P}_2(\mathcal{Q})$ , the goal is to find a feedback control  $v^* : \mathcal{Q}_T \rightarrow \mathbb{R}^k$  minimizing

$$J^{MFOT} : v \mapsto \mathbb{E} \left[ \int_0^T f(X_t^v, \mu^v(t), v(t, X_t^v)) dt \right], \quad (\text{III.3})$$

where  $\mu^v(t)$  is the distribution of  $X_t^v$ , under the constraint that the process  $X^v = (X_t^v)_{t \geq 0}$  solves the SDE

$$\begin{cases} X_0^v \sim \rho_0, \quad X_T^v \sim \rho_T \\ dX_t^v = b(X_t^v, \mu^v(t), v(t, X_t^v)) dt + \sigma dW_t, \quad t \geq 0. \end{cases} \quad (\text{III.4})$$

---

<sup>1</sup>By analogy with MFG of planning type, we could also call such problems ‘‘MFC of planning type’’. Yet another possible terminology would be ‘‘mean field control problems with a terminal constraint’’. But referring to ‘‘optimal transport’’ seems clearer so we will stick to the MFOT terminology. We note that we present the problem with a stochastic dynamics while standard optimal transport problems are usually presented without noise in the dynamics, see [Vil09b](#), [San15](#). However, the topic of optimal transport with stochastic dynamics has recently attracted interest [TT13](#), [Mik21](#), which justifies our terminology.

We stress that the terminal constraint implicitly restricts the class of admissible controls since we are interested in minimizing only over controls  $v$  that make  $X_T^v$  have distribution  $\rho_T$ .

We now present a few useful examples, some of which will be revisited in the numerical experiments (see [Section III.4](#)).

**Example 1** (Optimal transport). *When  $b(x, \mu, a) = a$ ,  $f(x, \mu, a) = \frac{1}{2}a^\top a$  and  $\sigma = 0$ , the MFOT problem reduces to a standard OT problem. See e.g. [\[BB00\]](#).*

**Example 2** (Linear-quadratic). *Take  $b(x, \mu, a) = Ax + \bar{A}\bar{\mu} + Ba$ ,  $f(x, \mu, a) = x^\top Qx + \bar{\mu}^\top \bar{Q}\bar{\mu} + a^\top Ra$ , and  $g(x, \mu) = x^\top Q_T x + \bar{\mu}^\top \bar{Q}_T \bar{\mu}$ , where  $\bar{\mu} = \int \xi \mu(d\xi)$ , where  $A, \bar{A}, B, Q, \bar{Q}, R, Q_T$  and  $\bar{Q}_T$  are matrices of suitable sizes. In this setting, the MFC problem has an explicit solution, up to solving a forward-backward system of ODEs. Furthermore, if the initial distribution is Gaussian, then the optimal flow of distribution remains Gaussian. See e.g. [\[BFY13, Chapter 6\]](#). To the best of our knowledge, in the MFOT setting, a similar result is available in the literature only when  $Q = \bar{Q} = 0$ , which corresponds to the Schrödinger bridge problem. See [\[CGP18b, Section 7.1\]](#).*

**Example 3** (Crowd motion with congestion). *Take  $b(x, \mu, a) = a$ ,  $f(x, \mu, a) = (c + \rho \star \mu(x))^\gamma |a|^2 + \ell(x, \mu(x))$ , where  $c \geq 0$  is a constant,  $\rho$  is a regularizing kernel and  $\star$  denotes the convolution. For  $\gamma = 0$ , the model is linear-quadratic in the control. If  $\gamma > 0$ , the cost of moving increases with the density surrounding the agent, which represents the fact that the “energy” spent to move is higher in regions with higher density. This models a congestion effect. The last term in  $f$  can be used to represent crowd aversion if  $\ell$  is increasing with respect to  $\mu(x)$ , and it can be used to represent spatial preferences by taking for instance  $\ell(x, \mu(x)) = |x_* - x|^2$ , where  $x_*$  is a preferred position. The terminal cost  $g$  can also be used to represent crowd aversion or spatial preferences. See e.g. [\[AL15, AL16a\]](#) for more details on the analysis of the MFC PDE system for this class of models and [\[AL16b\]](#) for numerical aspects. When  $\ell = 0$  and  $\sigma = 0$ , the corresponding MFOT problem has been studied e.g. in [\[CCNT13\]](#). Similar models have also been studied in the context of MFGs, see e.g. [\[AP18, AL19\]](#).*

### III.3 Numerical methods

In this section, we introduce three different numerical methods to solve MFOT. [Section III.3.1](#) introduces a direct approach to solve a MFC problem that approximates the MFOT problem. [Section III.3.2](#) discusses the Deep Galerkin Method (DGM) to solve the underlying PDE system that characterizes the optimal solution to MFOT, which is composed of a coupled Hamilton-Jacobi-Bellman equation and a Kolomogrov-Fokker-Planck equation. [Section III.3.3](#) introduces the DeepADMM algorithm that solves a variational reformulation of the MFOT problem based on an augmented Lagrangian approach.

#### III.3.1 Direct approach for the optimal control formulation

We first introduce the direct approach, which does not require any derivation of optimality conditions. In order to make the problem numerically tractable, we make approximations on several levels. Motivated by the deep learning method for MFC problems proposed in [\[CL19\]](#)

(see also the first algorithm in [CL21b]), we first approximate the MFOT problem (III.3) by an MFC problem in which a terminal penalty is incurred based on the distance between the terminal distribution and the target distribution. We can then apply the algorithm of [CL19], which trains a neural network to learn the optimal control of the MFC problem. This method itself relies on three approximations.

### III.3.1.1 Problem Approximation

Instead of directly tackling the MFOT problem (III.3), we first consider the following MFC problem as an approximation of the original problem: Find a feedback control  $v^* : \mathcal{Q}_T \rightarrow \mathbb{R}^k$  minimizing (III.1) under the constraint (III.2) when the terminal cost is:

$$g(x, \mu) = G(\mathcal{W}_2(\mu, \rho_T)), \quad \mu \in \mathcal{P}_2(\mathcal{Q}). \quad (\text{III.5})$$

where  $G : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is an increasing function and  $\mathcal{W}_2$  denotes the Wasserstein distance on  $\mathcal{P}_2(\mathcal{Q})$ . A typical example that we will use in the experiments is a linear function. The purpose of introducing  $G(\mathcal{W}_2(\mu, \rho_T))$  is to add a penalty that enforces the planning constraint for the terminal distribution. Here we focus on the Wasserstein distance because of its connection with optimal transport, see e.g. [BB00, San15], although other similarity measures could be used. In our numerical experiments, we will take an increasing linear function for  $G$ .

Then, we use the following approximations:

- Since it is not possible to optimize over all feedback controls, we restrict the space of controls to the space of neural networks with a given architecture. We will denote by  $v_\theta$  a representative neural network of this class with parameter  $\theta$ . The problem becomes a finite-dimensional optimization problem, in which the goal is to find a value for the parameter  $\theta$  that minimizes the loss  $J(v_\theta)$ , i.e., the total cost of the MFC problem when using control  $v_\theta$ .
- Since it is not possible to represent the mean field state  $\mu^{v_\theta}(t)$  or to compute its evolution exactly, we approximate it by the empirical distribution  $\bar{\mu}^{N, v_\theta}(t) = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i, v_\theta}}$ , where each  $X_t^{i, v_\theta}$  is a solution of,

$$\begin{cases} X_0^{i, v_\theta} \sim \rho_0 & \text{i.i.d.} \\ dX_t^{i, v_\theta} = b(X_t^{i, v_\theta}, \bar{\mu}^{N, v_\theta}(t), v_\theta(t, X_t^{i, v_\theta}))dt + \sigma dW_t^i, & t \geq 0, \end{cases} \quad (\text{III.6})$$

where  $(W^i)_{i=1, \dots, N}$  is a family of  $N$  independent  $d$ -dimensional Brownian motions, which represent idiosyncratic noises affecting each particle independently. All the SDEs are based on the same control function  $v_\theta$ .

- Last, in order to be able to compute these dynamics using Monte Carlo simulations, we discretize the time variable  $t$ . Letting  $N_T$  be a number of time steps of length  $\Delta t = T/N_T$ , we replace the interval  $[0, T]$  by the time steps  $\{t_0 = 0, t_1 = \Delta t, \dots, t_{N_T} = N_T \Delta t\}$ . The time steps are  $t_n = n \Delta t$ ,  $n = 0, \dots, N_T$ . We then approximate the SDE system (III.6)

using an Euler-Maruyama scheme. The family of trajectories  $((X_t^{i,v_\theta})_{t \in [0,T]})_{i=1,\dots,N}$  is approximated by the family of sequences  $((X_{t_n}^{i,v_\theta,N_T})_{n=0,\dots,N_T})_{i=1,\dots,N}$  satisfying:

$$\begin{cases} X_0^{i,v_\theta,N_T} \sim \rho_0 & \text{i.i.d.} \\ X_{t_{n+1}}^{i,v_\theta,N_T} = X_{t_n}^{i,v_\theta,N_T} + b(X_{t_n}^{i,v_\theta,N_T}, \bar{\mu}_{t_n}^{N,v_\theta,N_T}, v_\theta(t_n, X_{t_n}^{i,v_\theta,N_T}))\Delta t + \sigma \Delta W_n^i, \end{cases} \quad (\text{III.7})$$

where  $\bar{\mu}_{t_n}^{N,v_\theta,N_T} = \frac{1}{N} \sum_{i=1}^N \delta_{X_{t_n}^{i,v_\theta,N_T}}$  is the empirical distribution associated with the samples  $X_{t_n}^{i,v_\theta,N_T}$ . Here,  $(\Delta W_n^i)_{i=1,\dots,N,n=0,\dots,N_T-1}$  are independent Gaussian random variables with variance  $\Delta t$ .

To summarize, the new problem is to find  $\theta^*$  minimizing:

$$J^{N,N_T}(\theta) = \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^N \sum_{n=0}^{N_T-1} f(X_{t_n}^{i,\theta,N_T}, \bar{\mu}_{t_n}^{N,\theta,N_T}, v_\theta(t_n, X_{t_n}^{i,\theta,N_T}))\Delta t + g(X_T^{N,\theta,N_T}, \bar{\mu}_T^{N,\theta,N_T}) \right]$$

subject to the dynamics [\(III.7\)](#). The full analysis of this problem and its rigorous connection with the original MFOT problem [\(III.3\)](#) is beyond the scope of this paper and is left for future work. We expect the control  $v_{\theta^*}$ , with the parameter value that is optimal for the above problem, to be approximately optimal for [\(III.3\)](#), under suitable assumptions on  $b$  and  $f$ . In particular,  $b$  and  $f$  should probably depend smoothly on the distribution so that they can be evaluated in a meaningful way at the empirical distribution  $\bar{\mu}_{t_n}^{N,\theta,N_T}$ .

### III.3.1.2 Description of the algorithm

**Optimization method.** To find an approximate minimizer, we use stochastic gradient descent (SGD) or one of its variants. At iteration  $k$ , we have a parameter  $\theta_k$  that we wish to update. We sample the initial positions  $(X_0^{i,v_{\theta_k},N_T})_{i=1,\dots,N}$  and the Brownian motion increments  $(\Delta W_n^i)_{i=1,\dots,N,n=0,\dots,N_T-1}$ . We then compute the empirical cost for this realization of the  $N$ -particle population, and use its gradient with respect to  $\theta_k$  to update the parameter. In other words, we apply SGD to the following loss function:

$$\mathcal{L}(\theta) = J^{N,N_T}(\theta) = \mathbb{E}_S[\mathcal{L}(\theta; S)],$$

with:

$$\mathcal{L}(\theta; S) = \frac{1}{N} \sum_{i=1}^N \sum_{n=0}^{N_T-1} f(X_{t_n}^{i,v_\theta,N_T}, \bar{\mu}_{t_n}^{N,v_\theta,N_T}, v_\theta(t_n, X_{t_n}^{i,v_\theta,N_T}))\Delta t + g(X_T^{N,v_\theta,N_T}, \bar{\mu}_T^{N,v_\theta,N_T})$$

where  $S = ((X_0^{i,v_\theta,N_T})_{i=1,\dots,N}, (\Delta W_n^i)_{i=1,\dots,N,n=0,\dots,N_T-1})$  denotes one random sample.

**Computation of the Wasserstein distance.** As shown in [\(III.5\)](#), the new problem we considered involves a Wasserstein distance between two continuous distributions, namely, the mean field distribution at terminal time  $\mu_T^v$  and the target distribution  $\rho_T$ . This is in general hard to compute. However, in our implementation, the mean field distribution is approximated by



an empirical distribution obtained by Monte Carlo simulations, as is explained above. We then sample the same number of points from the target distribution and compute the Wasserstein distance between the two empirical distributions. This is done in the following way. Let  $X$  and  $Y$  be two sets of  $N$  points each sampled from distributions  $\mu, \nu$ ,  $M_p$  the distance matrix,  $(M_p)_{ij} = |X_i - Y_j|^p$ , and the following set:

$$U_N = \left\{ A \in \mathbb{R}^{N \times N} \mid \sum_{j=1}^N A_{ij} = \sum_{i=1}^N A_{ij} = \frac{1}{N} \right\}. \quad (\text{III.8})$$

Then

$$\left( \mathcal{W}_p(\mu, \nu) \right)^p = \lim_{N \rightarrow \infty} \min_{T \in U_N} \langle T, M_p \rangle.$$

In order to efficiently compute the Wasserstein distance, we follow the algorithm proposed by Cuturi in [Cut13]. We consider an extra entropy regularization of the following form. Let  $\alpha > 0$ , we want to find  $T_\alpha^*$ , which is the solution to the following program:

$$\min_{T \in U_N} \langle T, M_p \rangle - \alpha \langle T \log(T), 1 \rangle.$$

Optimality conditions and Sinkhorn-Knopp [SK67] theorem give us the existence and uniqueness of the solution, as well as a unique decomposition of  $T_\alpha^*$  using two vectors  $u$  and  $v$  such that:

$$T_\alpha^* = \text{Diag}(u) \exp\left(-\frac{M_p}{\alpha}\right) \text{Diag}(v).$$

We can then compute  $u$  and  $v$  with Sinkhorn-Knopp algorithm. Further explanations on this algorithm can be found in [Cut13]. This method allows for fast computations and is easy to export to greater dimensions, at the cost of adding a layer of approximation due to the extra parameter  $\alpha$ . It can be noticed that as  $\alpha$  tends to zero, the regularized solution tends to the solution of discrete optimal transport. In practice, reducing  $\alpha$  to zero increases the number of iterations required for Sinkhorn algorithm to converge. However, in our numerical experiments we usually obtain good results with a small but non-zero  $\alpha$ .

**Remark 1.** *Notice that using our approach, we have one empirical distribution and one continuous distribution. Indeed, we have the empirical distribution obtained by Monte Carlo simulation and the target distribution  $\rho_T$ , which is generally given by a closed-form formula for its density. We could thus try to use the designated methods, such as Semi-discrete Optimal transport [MT21]. While being more accurate, these methods do not scale well in higher dimensions compared to Sinkhorn's alternative. Another interesting option, which might scale well in terms of dimension is [ESV<sup>+</sup>19, MNW19]. However, in our numerical experiments, we did not succeed to obtain equally good results using this alternative approach. Further investigation is left for future work.*

**Terminal penalty.** In our implementation, we take  $G$  as a linear function  $G(r) = C_W r$ , where  $C_W$  is a positive constant that weighs the importance of the terminal penalty in comparison with the running cost. This leads to a trade-off between minimizing the running cost

and satisfying the terminal constraint. We noticed that when  $C_W$  is too small, the algorithm minimizes the running cost without much consideration for the terminal condition and hence the terminal distribution is far from the target distribution. Therefore, the penalization has to be a significant component of the total loss if we want the terminal planning constraint to be approximately satisfied with good accuracy (see [Section III.4.3](#) for a discussion on the choice of this hyperparameter).

### III.3.2 Deep Galerkin Method for the PDE system

We now turn our attention to a method based on solving a forward-backward PDE system that characterizes the solution. We first discuss the PDE system and then use a deep learning method to solve this system.

#### III.3.2.1 PDE system for MFOT

As recalled above, in a standard MFC, the whole population uses a given feedback control  $v$ . Assuming that the distribution  $\mu^v(t) = \mathcal{L}(X_t^v)$  of a representative agent with dynamics [\(III.2\)](#) admits a smooth enough density  $m(t, \cdot)$ , the latter satisfies the Kolmogorov-Fokker-Planck (KFP) PDE:

$$\begin{cases} \frac{\partial m}{\partial t}(t, x) - \nu \Delta m(t, x) + \operatorname{div}(m(t, x)b(x, m(t, \cdot), v(t, x))) = 0 & t \in (0, T], x \in \mathcal{Q} \\ m(0, x) = m_0(x), & x \in \mathcal{Q}, \end{cases}$$

where  $m_0$  is the density of the initial distribution  $\rho_0$  and  $\nu = \frac{\sigma^2}{2}$ . The MFC problem [\(III.1\)](#) can then be viewed as an optimal control problem driven by the above KFP PDE. Under suitable conditions, the optimal control can be characterized through an adjoint PDE, which can be derived for instance via calculus of variations. See *e.g.* [\[BFY13, Chapter 4\]](#) for more details.

Let  $H : \mathcal{Q} \times L^2(\mathcal{Q}) \times \mathbb{R}^d \rightarrow \mathbb{R}$  be the Hamiltonian of the control problem faced by an infinitesimal agent in the first point above, which is defined by:

$$H : (x, m, p) \mapsto H(x, m, p) = \max_{v \in \mathbb{R}^k} \{-L(x, m, v, p)\}, \quad (\text{III.9})$$

where  $m$  denotes the density of  $\mu$ ,  $L : \mathcal{Q} \times L^2(\mathcal{Q}) \times \mathbb{R}^k \times \mathbb{R}^d \rightarrow \mathbb{R}$  is the Lagrangian, defined by:

$$L : (x, m, v, p) \mapsto L(x, m, v, p) = f(x, m, v) + \langle b(x, m, v), p \rangle. \quad (\text{III.10})$$

A necessary condition for the existence of an optimal control  $v^*$  is that:

$$v^*(t, x) = \operatorname{argmax}_{v \in \mathbb{R}^k} \{-L(x, m(t, \cdot), v, \nabla u(t, x))\},$$

where  $(u, m)$  solves the following system of partial differential equations:

$$\left\{ \begin{array}{l} 0 = -\frac{\partial u}{\partial t}(t, x) - \nu \Delta u(t, x) + H(x, m(t, \cdot), \nabla u(t, x)) \\ \quad + \int_{\mathcal{Q}} \frac{\partial H}{\partial m}(\zeta, m(t, \cdot), \nabla u(t, \zeta))(x) m(t, \zeta) d\zeta, \quad \text{in } (0, T] \times \mathcal{Q}, \quad (\text{III.11a}) \\ 0 = \frac{\partial m}{\partial t}(t, x) - \nu \Delta m(t, x) - \operatorname{div}\left(m(t, x) \partial_p H(x, m(t, \cdot), \nabla u(t, x))\right), \quad \text{in } [0, T) \times \mathcal{Q}, \quad (\text{III.11b}) \\ u(T, x) = g(x, m(T, \cdot)) + \int_{\mathcal{Q}} \frac{\partial g}{\partial m}(\zeta, m(T, \cdot))(x) m(T, \zeta) d\zeta, \quad \text{in } \mathcal{Q}, \quad (\text{III.11c}) \\ m(0, x) = m_0(x), \quad \text{in } \mathcal{Q}. \quad (\text{III.11d}) \end{array} \right.$$

The partial derivatives with respect to  $m$  appear in the backward PDE due to the fact that the population distribution changes when the control changes. These partial derivatives with respect to  $m$  should be understood in the following sense: if  $\varphi : L^2(\mathbb{R}^d) \rightarrow \mathbb{R}$  is differentiable,

$$\left. \frac{d}{d\varepsilon} \varphi(m + \varepsilon \tilde{m})(x) \right|_{\varepsilon=0} = \int_{\mathbb{R}^d} \frac{\partial \varphi}{\partial m}(m)(\zeta) \tilde{m}(\zeta) d\zeta.$$

We refer to e.g. [BFY13, Chapter 4] for more details and for the derivation using calculus of variations, which clarifies why the partial derivatives with respect to  $m$  appear. If the cost functions and the drift function depend on the density only locally (i.e., only on the density at the current position of the agent),  $\frac{\partial}{\partial m}$  becomes a derivative in the usual sense.

In this PDE system,  $m$  plays the role of the MFC problem's state. The forward equation is a Kolmogorov-Fokker-Planck (KFP) equation which describes the evolution of the mean field distribution. The other unknown function,  $u$ , plays the role of an adjoint state. Although the backward PDE has the form of a Hamilton-Jacobi-Bellman (HJB) equation,  $u$  cannot, in general, be interpreted as the value function associated to problem (III.1) because the value function depends on the population distribution, see e.g. [LP14, BFY15]. We refer the interested reader to e.g. [BFY13, Chapters 3 and 4] for the comparison with the MFG PDE system, in which the terms involving a derivative with respect to  $m$  are absent, and  $u$  can be interpreted as the value function of an infinitesimal player.

Now, for the MFOT problem, we can proceed formally in a similar way. We derive an analogous PDE system, except that the terminal condition for  $u$  disappears, and a terminal condition for  $m$  is added to the system. More precisely, we (formally) obtain the following PDE system:

$$\left\{ \begin{array}{l} 0 = -\frac{\partial u}{\partial t}(t, x) - \nu \Delta u(t, x) + H(x, m(t, \cdot), \nabla u(t, x)) \\ \quad + \int_{\mathcal{Q}} \frac{\partial H}{\partial m}(\zeta, m(t, \cdot), \nabla u(t, \zeta))(x) m(t, \zeta) d\zeta, \quad \text{in } (0, T] \times \mathcal{Q}, \quad (\text{III.12a}) \\ 0 = \frac{\partial m}{\partial t}(t, x) - \nu \Delta m(t, x) - \operatorname{div}\left(m(t, x) \partial_p H(x, m(t, \cdot), \nabla u(t, x))\right), \quad \text{in } [0, T) \times \mathcal{Q}, \quad (\text{III.12b}) \\ m(0, x) = m_0(x), \quad m(T, x) = m_T(x) \quad \text{in } \mathcal{Q}. \quad (\text{III.12c}) \end{array} \right.$$

where  $m_0$  and  $m_T$  are respectively the densities of  $\rho_0$  and  $\rho_T$ . To the best of our knowledge, this PDE system has not been derived nor analyzed in a general setting. Notice that even the existence of a solution is a non trivial question due to the fact that there is both an initial and a terminal constraint on the density. However, this system has been analyzed in special cases corresponding to optimal transport [BB00, CCN13] or to MFGs with planning [ACCD12, Por13, Por14, GMST19]. In the numerical examples of Section III.4, we will mostly focus on cases that have been previously studied, such as standard optimal transport or MFOT with congestion effects captured by the running cost.

### III.3.2.2 Description of the algorithm

To solve the PDE system (III.12), we follow the idea of the Deep Galerkin Method (DGM) introduced by Sigignano and Spiliopoulos [SS18] and adapted to the MFG and PDE systems in [AACN<sup>+</sup>18, CL21a, CL21b]. The main motivation underlying this approach is to learn the PDE solutions using parameterized functions. This avoids computing the functions on a mesh, which is not feasible in high dimensions. In the DGM, we replace the function(s) solving the PDE(s) with a neural network(s), which are trained to minimize the PDE residual(s) as well as the boundary condition(s).

To be specific, in our setting, we replace the functions  $m$  and  $u$  with neural networks, denoted by  $m_\theta$  and  $u_\omega$  and parameterized by  $\theta$  and  $\omega$  respectively. When the state  $x$  is in high dimension, i.e.,  $d$  is large, we expect  $m_\theta$  and  $u_\omega$  to provide good approximations of  $m$  and  $u$  using much fewer parameters than the number of points in a grid. Furthermore, for the numerical implementation, we restrict our attention to a compact domain  $\tilde{\mathcal{Q}}$ . We denote  $\tilde{\mathcal{Q}}_T = [0, T] \times \tilde{\mathcal{Q}}$ . We expect the density to have a negligible mass outside a compact set so that by solving the PDE system on a large enough compact set, we obtain a good approximation of the solution, at least in the region where the density is significantly positive. We then define the loss function:

$$\mathcal{L}(\theta, \omega) = \mathcal{L}^{(\text{KFP})}(m_\theta, u_\omega) + \mathcal{L}^{(\text{HJB})}(m_\theta, u_\omega),$$

where, for any  $(m, u) \in \mathcal{C}^{1,2}(\tilde{\mathcal{Q}}_T) \times \mathcal{C}^{1,2}(\tilde{\mathcal{Q}}_T)$ , the two losses are:

$$\begin{aligned} \mathcal{L}^{(\text{KFP})}(m, u) &= C^{(\text{KFP})} \left\| \frac{\partial m}{\partial t} - \nu \Delta m - \operatorname{div} \left( m \partial_p H(\cdot, m, \nabla u) \right) \right\|_{L^2(\tilde{\mathcal{Q}}_T)}^2 \\ &\quad + C_0^{(\text{KFP})} \|m(0, \cdot) - m_0\|_{L^2(\tilde{\mathcal{Q}})}^2 + C_T^{(\text{KFP})} \|m(T, \cdot) - m_T\|_{L^2(\tilde{\mathcal{Q}})}^2, \end{aligned} \quad (\text{III.13})$$

and (omitting the dependence on  $x$  and  $t$ , and writing explicitly only the dependence on the extra variable  $\zeta$  in the integral):

$$\mathcal{L}^{(\text{HJB})}(m, u) = C^{(\text{HJB})} \left\| -\frac{\partial u}{\partial t} - \nu \Delta u + H(\cdot, m, \nabla u) + \int_{\tilde{\mathcal{Q}}} \frac{\partial H}{\partial m}(\zeta, m, \nabla u(\zeta))(\cdot) m(\zeta) d\zeta \right\|_{L^2(\tilde{\mathcal{Q}}_T)}^2.$$

Here,  $C^{(\text{KFP})}$ ,  $C_0^{(\text{KFP})}$ ,  $C_T^{(\text{KFP})}$ ,  $C^{(\text{HJB})}$  are positive weights that give more or less importance to each component. If the space domain is bounded, we must include more penalty terms. Note that any smooth enough solution  $(m, u)$  to the PDE system (III.12) makes  $\mathcal{L}^{(\text{KFP})}$  and  $\mathcal{L}^{(\text{HJB})}$  vanish. The goal is to find two neural networks which approximately minimize these losses.

Since it is not possible to compute exactly the above residuals, we approximate the  $L^2$  norms using Monte Carlo samples. For example, we rewrite:

$$\begin{aligned} & \left\| \frac{\partial m}{\partial t} - \nu \Delta m - \operatorname{div} \left( m \partial_p H(\cdot, m, \nabla u) \right) \right\|_{L^2(\tilde{\mathcal{Q}}_T)}^2 \\ &= C(\tilde{\mathcal{Q}}_T) \mathbb{E}_{\tau, \xi} \left[ \left| \frac{\partial m}{\partial t}(\tau, \xi) - \nu \Delta m(\tau, \xi) - \operatorname{div} \left( m(\tau, \xi) \partial_p H(\xi, m(\tau), \nabla u(\tau, \xi)) \right) \right|^2 \right], \end{aligned}$$

where  $(\tau, \xi)$  follows a uniform distribution over  $\tilde{\mathcal{Q}}_T$ , and  $C(\tilde{\mathcal{Q}}_T)$  is a normalizing constant that depends on the domain. Likewise, for the other norms, it would also be possible to use different norms and different distributions to sample  $(\tau, \xi)$ . But for the sake of simplicity, we will stick to this setting for the present work. We obtain the following probabilistic formulation of the loss function  $\mathcal{L}$ :

$$\mathcal{L}(\theta, \omega) = \mathbb{E}_S [\mathcal{L}(\theta, \omega; S)], \quad \mathcal{L}(\theta, \omega; S) = \mathcal{L}^{(\text{KFP})}(m_\theta, u_\omega; S) + \mathcal{L}^{(\text{HJB})}(m_\theta, u_\omega; S),$$

where  $S = (\tau, \xi, \xi_0, \xi_T) \in [0, T] \times \tilde{\mathcal{Q}} \times \tilde{\mathcal{Q}} \times \tilde{\mathcal{Q}}$  denotes one sample, and for any  $m, u$  in  $\mathcal{C}^{1,2}(\mathcal{Q}_T)$ , the two losses at  $S$  are as:

$$\begin{aligned} \mathcal{L}^{(\text{KFP})}(m, u; S) &= C^{(\text{KFP})} \left| \frac{\partial m}{\partial t}(\tau, \xi) - \nu \Delta m(\tau, \xi) - \operatorname{div} \left( m(\tau, \xi) \partial_p H(\xi, m(\tau), \nabla u(\tau, \xi)) \right) \right|^2 \\ &\quad + C_0^{(\text{KFP})} |m(0, \xi_0) - m_0(\xi_0)|^2 + C_T^{(\text{KFP})} |m(T, \xi_T) - m_T(\xi_T)|^2, \end{aligned}$$

and

$$\begin{aligned} \mathcal{L}^{(\text{HJB})}(m, u; S) &= C^{(\text{HJB})} \left| - \frac{\partial u}{\partial t}(\tau, \xi) - \nu \Delta u(\tau, \xi) + H(\xi, m(\tau), \nabla u(\tau, \xi)) \right. \\ &\quad \left. + \int_{\mathcal{Q}} \frac{\partial H}{\partial m}(\zeta, m(\tau), \nabla u(\tau, \zeta))(\xi) m(\tau, \zeta) d\zeta \right|^2. \end{aligned}$$

Finally, to optimize over  $(\theta, \omega)$ , we use SGD (or one of its variants) on the loss  $\mathcal{L}$ . In practice, we use a mini-batch of samples at each iteration, which amounts to approximate the expectation by an empirical average over several samples.

### III.3.3 Augmented Lagrangian Method with Deep Learning

In this subsection, we present an approach based on a primal-dual formulation of the MFOT problem. We then introduce a deep learning adaptation of the alternating direction method of multipliers. We focus on the case when the interactions are local, and the drift is the control.

#### III.3.3.1 Primal and dual problems

Under suitable assumptions, the MFOT problem admits a variational formulation, which can be tackled using a direct optimization approach. As in the previous subsection, we assume that  $\rho_0$  and  $\rho_T$  have respectively density  $m_0$  and  $m_T$ .

We focus on a model with local interactions, meaning that an agent at state  $x$  interacts with the density of the population at  $x$ . To alleviate the presentation, we will use the same notation for the costs and the drift functions, but now their second input is a real number  $m$  instead of an element  $\mu \in \mathcal{P}_2(\mathcal{Q})$ . So we have  $f : \mathcal{Q} \times \mathbb{R} \times \mathbb{R}^k \rightarrow \mathbb{R}$ ,  $g : \mathcal{Q} \times \mathbb{R} \rightarrow \mathbb{R}$ , and  $b : \mathcal{Q} \times \mathbb{R} \times \mathbb{R}^k \rightarrow \mathbb{R}^d$ . We also modify accordingly the definition of the Hamiltonian  $H$  in (III.9) and the Lagrangian  $L$  in (III.10) in Section III.3.2. We further assume that  $f(x, m, v)$  is convex in  $v$  for every  $(x, m)$ , and  $mf(x, m, v)$  is convex in  $m$  for every  $(x, v)$ . For simplicity, we consider that  $b(x, m, v) = v$ , i.e., the drift is the control. Therefore, in this model, we take  $k = d$  (and we will later use  $k$  to denote the iteration index).

**Primal problem.** The MFOT problem (III.3) introduced in Section III.2 is formally equivalent to the following PDE-constrained optimization problem:

$$\begin{aligned} & \inf_{v: \mathcal{Q}_T \rightarrow \mathbb{R}^d} \int_{\mathcal{Q}_T} f(x, m(t, x), v(t, x)) m(x, t) dx dt \\ \text{subject to } & \frac{\partial m}{\partial t}(t, x) - \nu \Delta m(t, x) + \operatorname{div}(m(t, x)v(t, x)) = 0 \quad t \in (0, T], x \in \mathcal{Q} \\ & m(0, x) = m_0(x), \quad m(T, x) = m_T(x) \end{aligned} \quad (\text{III.14})$$

The PDE constraint is the KFP equation corresponding to the stochastic dynamics in (III.4). Note that the formulation in terms of  $(m, v)$ , while intuitive, is not convex in general. For this reason, we consider an equivalent formulation in terms of  $(m, z) = (m, mv)$ . We define:

$$\tilde{f}(x, m, z) = \begin{cases} mf(x, m, \frac{z}{m}) & \text{if } m > 0 \\ 0 & \text{if } (m, z) = (0, 0) \\ +\infty & \text{otherwise} \end{cases} \quad (\text{III.15})$$

Note that  $(m, z) \mapsto \tilde{f}(x, m, z)$  is lower semi-continuous on  $\mathbb{R} \times \mathbb{R}^d$ . Under suitable conditions, it can be proved that  $(m, z) \mapsto \tilde{f}(x, m, z)$  is convex on  $\mathbb{R} \times \mathbb{R}^d$ . We also define the space  $\mathbf{K}$ ,

$$\mathbf{K} = \left\{ (m, z) \mid \begin{aligned} & \frac{\partial m}{\partial t}(t, x) - \nu \Delta m(t, x) + \operatorname{div} z(t, x) = 0, \\ & m(0, x) = m_0(x), m(T, x) = m_T(x), m \geq 0 \end{aligned} \right\} \quad (\text{III.16})$$

With all these definitions, we are ready to present the primal problem:

$$\inf_{(m, z) \in \mathbf{K}} \mathcal{B}(m, z) = \inf_{(m, z) \in \mathbf{K}} \int_{\mathcal{Q}_T} \tilde{f}(x, m(t, x), z(t, x)) dx dt \quad (\text{III.17})$$

Assuming that problem (III.17) has a unique optimal solution  $(m^*, z^*)$  and that problem (III.14) has a unique optimal control  $v^*$ , then the following connection holds:  $v^*(t, x) = z^*(t, x)/m^*(t, x)$  if  $m^*(t, x) > 0$ ,  $v^*(t, x) = 0$  if  $m^*(t, x) = 0$ .

**Dual problem.** We now introduce a dual optimization problem. We define the following

functionals:

$$\begin{aligned} \mathcal{A}(u) = \inf_{m \geq 0} \int_{\mathcal{Q}_T} m(t, x) \left( \frac{\partial u}{\partial t}(t, x) + \nu \Delta u(t, x) - H(x, m(t, x), \nabla u(t, x)) \right) dx dt \\ + \int_{\mathcal{Q}} (m_0(x)u(0, x) - m_T(x)u(T, x)) dx \end{aligned} \quad (\text{III.18})$$

$$\mathcal{F}(u) = \int_{\mathcal{Q}} m_T(x)u(T, x) - m_0(x)u(0, x) dx \quad (\text{III.19})$$

$$\mathcal{G}(\mathbf{a}, \mathbf{b}) = - \inf_{m \geq 0} \int_{\mathcal{Q}_T} m(t, x) \left( \mathbf{a}(t, x) - H(x, m(t, x), \mathbf{b}(t, x)) \right) dx dt. \quad (\text{III.20})$$

Note that if we define the linear differential operator  $\Lambda u = \left( \frac{\partial u}{\partial t} + \nu \Delta u, \nabla u \right)$ , then  $\mathcal{A}(u) = -(\mathcal{G}(\Lambda u) + \mathcal{F}(u))$ . Similar variation formulation for first-order mean field game, mean field control, and first-order planning mean field game has been considered e.g. in the works [\[AL16a\]](#), [\[GMST19\]](#), [\[BC15\]](#) <sup>2</sup>. Here, we generalize the variation formulation to the problem of mean field optimal transport. The functional  $\mathcal{A}(u)$  could be considered as the Lagrangian dual function of the primal functional  $\mathcal{B}(m, z)$ , which can be further related to the sum of two functionals  $\mathcal{F}(u)$  and  $\mathcal{G}(\Lambda u)$ . Consider the following problem:

$$\inf_u \mathcal{F}(u) + \mathcal{G}(\Lambda u). \quad (\text{III.21})$$

Based on Fenchel-Rockafellar duality theorem (see Section 31, Theorem 31.1 in [\[Roc70\]](#)), we expect problems [\(III.17\)](#) and [\(III.21\)](#) to be in duality, meaning:

$$\inf_{(m, z) \in \mathbf{K}} \mathcal{B}(m, z) = \sup_u \mathcal{A}(u) = - \inf_u \mathcal{F}(u) + \mathcal{G}(\Lambda u) \quad (\text{III.22})$$

Note that this primal-dual relationship also plays an important role in demonstrating the uniqueness and existence of solutions to MFG and MFC PDE systems, see e.g. [\[LL07\]](#), [\[CG15\]](#), [\[AL16a\]](#). Here, we expect a similar result to hold for MFOT under suitable conditions. The rigorous definition of the two problems and the analysis of this duality relationship is left for future work. For now, we proceed formally.

We can at least formally establish a connection between the primal problem, the dual problem, and the optimal control in the following way. Let  $u^*$  be the optimal solution to the dual [\(III.21\)](#) and let  $(m^*, z^*)$  be the optimal solution to the primal problem [\(III.17\)](#). Then the optimal control for the original problem [\(III.14\)](#) is given by:

$$v^*(t, x) = \partial_p H(x, m^*(t, x), \nabla u^*(t, x)).$$

We notice that  $(u^*, m^*)$  forms a solution to the MFOT PDE system [\(III.12\)](#). This fact suggests that we can work on the dual problem [\(III.21\)](#) directly to solve the MFOT problem. Under suitable assumptions, it can be shown that the dual problem [\(III.21\)](#) is a strongly convex, unconstrained optimization problem over function space, which motivates the use of classic algorithms in convex optimization. However, the presence of the infinite dimensional linear

<sup>2</sup>See Lemma 2.2 in [\[AL16a\]](#), Lemma 2.1 in [\[GMST19\]](#)

operator  $\Lambda$  makes the problem hard to solve efficiently in general. Fortunately, the structure of the objective as a sum of two convex functionals makes the problem amenable to algorithms based on splitting schemes, such as the Alternating Direction Method of Multipliers (ADMM) [BPC<sup>+</sup>11].

---

**Algorithm 1:** Vanilla ADMM for MFOT

---

**Data:** Initial Guess  $(u^{(0)}, q^{(0)}, \lambda^{(0)})$ ; number of iterations  $N$ ; hyperparameter  $r > 0$   
**Result:** Function  $(u^{(N)}, q^{(N)}, \lambda^{(N)})$  that are close to the saddle point of  $\mathcal{L}_r$  defined in (III.24)

```

1 begin
2   for  $k = 1, \dots, N$ , do
3      $u^{(k)} = \arg \min_{u: \mathcal{Q}_T \rightarrow \mathbb{R}} \mathcal{F}(u) - \langle \lambda^{(k-1)}, \Lambda u \rangle + \frac{r}{2} \|\Lambda u - q^{(k-1)}\|^2$ 
4      $q^{(k)} = \arg \min_{q: \mathcal{Q}_T \rightarrow \mathbb{R}^{d+1}} \mathcal{G}(q) + \langle \lambda^{(k-1)}, q \rangle + \frac{r}{2} \|\Lambda u^{(k)} - q\|^2$ 
5      $\lambda^{(k)} = \lambda^{(k-1)} - r(\Lambda u^{(k)} - q^{(k)})$ 

```

---

### III.3.3.2 Description of the algorithm

Introducing a new variable  $q$  that will play the role of  $\Lambda u$ , we can rewrite problem (III.21) as the following constrained optimization program:

$$\inf_{u, q: q = \Lambda u} \mathcal{F}(u) + \mathcal{G}(q). \quad (\text{III.23})$$

The goal is now to find a saddle point of the associated Lagrangian. In fact, for numerical purposes, we will consider an augmented Lagrangian, defined as follows.

Let  $r > 0$  be a constant and introduce  $\lambda : \mathcal{Q}_T \rightarrow \mathbb{R}^{d+1}$ , the Lagrangian multiplier associated with the constraint  $q = \Lambda u$ . Let  $\langle \cdot, \cdot \rangle$  denote the inner product on  $L^2(\mathcal{Q}_T)$ . We introduce the augmented Lagrangian:

$$\mathcal{L}_r(u, q, \lambda) = \mathcal{F}(u) + \mathcal{G}(q) - \langle \lambda, \Lambda u - q \rangle + \frac{r}{2} \|\Lambda u - q\|^2 \quad (\text{III.24})$$

Now, the original MFOT problem is reduced to finding a saddle point of  $\mathcal{L}_r$ . Here, we state the original ADMM method in Algorithm 1 that finds the saddle point via an alternating optimization procedure.

This general procedure can be implemented, for example, when the functions  $(u, q, \lambda)$  are approximated by their values on a finite-difference grid. Such a procedure has been used for MFG and MFC problems, using finite elements [BC15, And17] or finite differences [AL16b]. Furthermore, Benamou et.al. proved that under the existence of solution and the full column rank condition of the finite-difference differential operator  $\Lambda$ , the ADMM method converges to the desired solution for mean field games, see Theorem 3.1 in [BC15]. However, as already mentioned, approximating functions by their values on a mesh is not feasible in high dimensions. We thus propose a different implementation of the ADMM based on neural network approximations.



In [Algorithm 1](#), the objectives in the steps are given by functionals to be minimized over functional spaces, which is not tractable in general. We restrict our attention to spaces of parameterized functions that can be expressed as neural networks, denoted by  $(u_\theta, q_\omega, \lambda_\psi)$  with parameter  $\theta, \omega, \psi$  respectively. We then follow the strategy introduced with the DGM [SS18](#) and already used in [Section III.3.2](#) to create computable loss functions that are stochastic approximations of the functionals.

Recall that the truncated space domain  $\tilde{\mathcal{Q}}$  and the associated time-space domain  $\tilde{\mathcal{Q}}_T$ . Let  $X \sim \mathcal{U}(\tilde{\mathcal{Q}}_T)$ ,  $Y \sim \mathcal{U}(\tilde{\mathcal{Q}})$  be two random variables with uniform distribution in the time-space domain and the space domain respectively. Let  $\rho_X, \rho_Y$  be the value of the uniform density on  $\tilde{\mathcal{Q}}_T$  and  $\tilde{\mathcal{Q}}$  respectively. Here, we overload the notation  $\langle \cdot, \cdot \rangle$  and  $\| \cdot \|$  to represent the Euclidean inner product and norm on both  $L_2(\tilde{\mathcal{Q}}_T)$  and  $\mathbb{R}^{d+1}$ :

$$\begin{aligned}\mathcal{L}^{(u)}(\theta; \omega, \psi) &= \mathcal{L}_1(u_\theta, q_\omega, \lambda_\psi), & \mathcal{L}^{(q)}(\omega; \theta, \psi) &= \mathcal{L}_2(u_\theta, q_\omega, \lambda_\psi), \\ \mathcal{L}^{(\lambda)}(\psi; \theta, \omega, \psi_{old}) &= \mathcal{L}_3(u_\theta, q_\omega, \lambda_{\psi_{old}}, \lambda_\psi),\end{aligned}$$

where

$$\mathcal{L}_1(u, q, \lambda) = \frac{1}{\rho_Y} \mathbb{E}_Y [u(T, Y)m_T(Y) - u(0, Y)m_0(Y)] \quad (\text{III.25})$$

$$+ \frac{1}{\rho_X} \mathbb{E}_X \left[ \frac{r}{2} \|\Lambda u(X) - q(X)\|^2 - \langle \Lambda u(X), \lambda(X) \rangle \right] \quad (\text{III.26})$$

$$\mathcal{L}_2(u, q, \lambda) = \frac{1}{\rho_X} \mathbb{E}_X \left[ \mathcal{G}(q(X)) + \langle \lambda(X), q(X) \rangle + \frac{r}{2} \|\Lambda u(X) - q(X)\|^2 \right] \quad (\text{III.27})$$

$$\mathcal{L}_3(u, q, \lambda_{old}, \lambda) = \frac{1}{\rho_X} \mathbb{E}_X \left[ \|\lambda_{old}(X) - r(\Lambda u(X) - q(X)) - \lambda(X)\|^2 \right]. \quad (\text{III.28})$$

Here, the subscript *old* is used to refer to the previous iteration: the loss for  $\lambda$  involves the previous estimate  $\lambda_{old}$ . When using a neural network, it amounts to using the previous neural network parameters  $\psi_{old}$ . The loss function aims at mimicking the effect of the direct update in the third step of standard ADMM ([Algorithm 1](#)) when  $\lambda$  is approximated by a neural network.

The algorithm DeepADMM is presented in [Algorithm 2](#).

We have several remarks regarding DeepADMM and the augmented Lagrangian formulation in order for readers to better understand this approach. First, compared with [Algorithm 1](#), the updates in [Algorithm 2](#) for functions  $u$  and  $q$  are quite straightforward to understand. Instead of searching optimizer over function space, we reduce the problem to a finite dimension through stochastic approximation of the objective and search in the parameter space instead. The computed stochastic gradient can be considered as an unbiased estimator of the population gradient with respect to the functional, and the variance of this stochastic gradient decreases as the batch size increases.

In [Section D.2](#), we discuss the computation of  $\mathcal{G}$  for several typical models.

## III.4 Numerical experiments

In this section, we present numerical experiments obtained with the three methods discussed in the previous section. For brevity, we refer to the three methods respectively introduced in

---

**Algorithm 2:** DeepADMM for MFOT

---

**Data:** Initial parameter  $\theta^{(0)}, \omega^{(0)}, \psi^{(0)}$ ; number of ADMM iterations  $K$ ; SGD parameters  
**Result:** Final parameter  $\theta^{(K)}, \omega^{(K)}, \psi^{(K)}$

```
1 begin
2   for  $k = 1, \dots, K$  do
3     Compute  $\theta^{(k)}$  using SGD to (approximately) minimize the loss
        $\mathcal{L}^{(u)}(\cdot; \omega^{(k-1)}, \psi^{(k-1)})$ 
4     Compute  $\omega^{(k)}$  using SGD to (approximately) minimize the loss
        $\mathcal{L}^{(g)}(\cdot; \theta^{(k)}, \psi^{(k-1)})$ 
5     Compute  $\psi^{(k)}$  using SGD to (approximately) minimize the loss
        $\mathcal{L}^{(\lambda)}(\cdot; \theta^{(k)}, \omega^{(k)}, \psi^{(k-1)})$ 
```

---

Sections [III.3.1](#), [III.3.2](#) and [III.3.3](#) as Method 1, Method 2 and Method 3 (and M1, M2, and M3 for short in the plots).

We first consider two test cases for which we have explicit solutions (up to solving ODE systems) and can thus be used to benchmark our algorithms in any dimension. We then consider two test cases that can be viewed as modifications of standard OT with crowd aversion or congestion effects.

### III.4.1 Case 1: Linear Quadratic Problem

The first class of models that we consider has a linear-quadratic structure, which falls in the setting discussed in Example [2](#).

#### III.4.1.1 Description of the problem

In this model, we take:

$$b(x, \mu, a) = Ax + Ba, \quad f(x, \mu, a) = a^\top Ra, \quad \rho_0 = \mathcal{N}(\bar{x}_0; \Sigma_0), \quad \rho_T = \mathcal{N}(\bar{x}_T; \Sigma_T),$$

where  $A, B, R, \Sigma_0, \Sigma_T$  are (constant) matrices of suitable sizes. The vectors  $\bar{x}_0$  and  $\bar{x}_T$  correspond to the initial and terminal means. We will consider two settings. In order to have a benchmark solution, we will take  $\sigma = B$ . This enables us to use the solution provided by [\[CGP18b\]](#), Section 7.1], which boils down to solving a system of ODEs. For the sake of completeness, we provide the details in [Section D.1](#).

#### III.4.1.2 Evaluation Metrics

In this model, since we have access to the optimal solution, we can evaluate the learnt solutions given by the three methods we proposed with respect to the ground-truth solution. We denote by  $v^*$  the optimal control and  $\hat{v}$  a learnt control. As explained below in detail, we use the following metrics: the total cost (namely  $J^{MFOT}(\hat{v})$ , with  $J^{MFOT}$  introduced in [\(III.3\)](#)), the relative

error between the achieved cost, and the optimal cost (namely  $J^{MFOT}(\hat{v})$  and  $J^{MFOT}(v^*)$ ), the deviation from the terminal distribution (i.e., the Wasserstein distance between the achieved terminal distribution and the target terminal distribution,  $\rho_T$ ), and the weighted  $L^2$  error between the learnt control  $\hat{v}$  and the optimal control  $v^*$ , weighted by the population distribution.

**Computation of the control.** The control is parameterized in different ways across different methods. For Method 1,  $\hat{v}(t, x) = v_\theta(t, x)$ . For Method 2 and Method 3,  $\hat{v}(t, x) = -\frac{1}{2}BR^{-1}\nabla\hat{u}_\theta(t, x)$ , where  $\hat{u}_\theta$  is the neural network that approximates the dual variable, solution to the HJB equation.

**Total cost.** Recall the definition of the objective  $J^{MFOT}$  defined in [\(III.1\)](#). Let  $v$  be a control. In the present Linear-Quadratic case, we have that,

$$J^{MFOT}(v) = \int_0^T \int_{\mathcal{Q}} m^v(t, x) f(x, m^v, v) dx dt = \int_0^T \int_{\mathcal{Q}} m^v(t, x) v(t, x)^\top R v(t, x) dx dt,$$

where  $m^v$  is the density of mean field distribution driven by  $v$ , which satisfies the KFP PDE [\(III.12\)](#). In order to evaluate  $J^{MFOT}(v)$ , we use Monte Carlo simulations. We discretize the time variable  $t$ . Let  $N_T$  be a number of time steps of length  $\Delta t = T/N_T$ . We consider an equi-distanced time discretization with time-steps  $\{t_0 = 0, t_1 = \Delta t, \dots, t_{N_T} = N_T \Delta t\}$ . Again, we simulate solutions to the underlying SDE using an Euler-Maruyama scheme similar to the one used in [\(III.7\)](#). We simulate a family of  $N$  sequences  $((X_{t_n}^{i,v})_{n=0, \dots, N_T})_{i=1, \dots, N}$  using the following update

$$\begin{cases} X_0^{i,v} \sim \rho_0 & \text{i.i.d.} \\ X_{t_{n+1}}^{i,v} = X_{t_n}^{i,v} + (AX_{t_n}^{i,v} + Bv(t_n, X_{t_n}^{i,v}))\Delta t + \sigma\sqrt{\Delta t}\Delta W_n^i, \end{cases} \quad (\text{III.29})$$

where  $\{\Delta W_n^i\}$  are i.i.d standard Gaussian random variables in  $\mathbb{R}^d$ . With these sampled sequences, we compute the objective as

$$J^{MFOT}(v) = \frac{1}{N} \sum_{i=1}^N \sum_{n=0}^{N_T-1} v(t_n, X_{t_n}^{i,v})^\top R v(t_n, X_{t_n}^{i,v}) \Delta t.$$

**Relative Error.** The relative error between  $J^{MFOT}(\hat{v})$  and  $J^{MFOT}(v^*)$  is defined as

$$\frac{|J^{MFOT}(\hat{v}) - J^{MFOT}(v^*)|}{|J^{MFOT}(v^*)|}.$$

The major reason to consider relative error instead of absolute error is because the scale of the running cost varies greatly across different problems. Also, we want to stress that even though  $v^*$  is the analytical optimal solution, it may happen that  $J^{MFOT}(\hat{v}) < J^{MFOT}(v^*)$  if  $\hat{v}$  does not satisfy exactly the constraint (in contrast with  $v^*$ ).

**Expected  $L^2$  error for control.** The expected  $L^2$  error between the learnt control  $\hat{v}$  and the ground-truth control  $v^*$  is defined as,

$$d_{L^2}(\hat{v}, v^*) = \int_0^T \int_{\mathcal{Q}} m^*(t, x) \|\hat{v}(t, x) - v^*(t, x)\|^2 dx dt$$

where  $m^*$  is the density of the optimal mean-field associated with  $v^*$ . For the LQ problem, the optimal mean field  $m^*$  is Gaussian for any  $t \in [0, T]$ , with mean  $\mu_t$  and variance  $\Sigma_t$  given by analytical formulas in [Section D.1](#). We can thus evaluate the  $L^2$  error again with Monte Carlo samples for each time step. As above, we discretize the time variable  $t$  with  $N_T + 1$  points and for  $t_n = n\Delta t, n = 0, \dots, N_T$ , we generate i.i.d. samples  $(X_{t_n}^{i,v})_{i=1,\dots,N} \sim \mathcal{N}(\mu_{t_n}, \Sigma_{t_n})$ . We then estimate the  $L^2$  error as:

$$d_{L^2}(\hat{v}, v^*) \approx \frac{1}{N} \sum_{i=1}^N \sum_{n=0}^{N_T-1} \|\hat{v}(t_n, X_{t_n}^{i,v}) - v^*(t_n, X_{t_n}^{i,v})\|^2 \Delta t.$$

**Deviation of distribution.** The deviation of the mean field  $\hat{\rho}_T$  from the terminal target distribution  $\rho_T$  is quantified by two different metrics: Wasserstein-2 distance  $\mathcal{W}_2(\hat{\rho}_T, \rho_T)$  and  $L^2$  distance  $d_{L^2}(\hat{m}_T, m_T)$ . Here,  $\hat{\rho}_T$  is the measure of the mean field distribution driven by the learnt control  $\hat{v}$  at time  $T$ .  $\hat{m}_T$  and  $m_T$  are the density of  $\hat{\rho}_T$  and  $\rho_T$  respectively.

- **Wasserstein-2 distance.** We adopt a similar method to compute the Wasserstein-2 distance as the one discussed in [Section III.3.1](#). We simulate  $N$  particles following the dynamics [\(III.29\)](#), and obtain a collection of  $N$  samples  $(X_T^{i,\hat{v}})_{i=1,\dots,N}$ , which forms an empirical distribution approximating  $\hat{\rho}_T$ . We also generate  $N$  samples directly from the target distribution  $\rho_T$ , denoted by  $(Y_T^i)_{i=1,\dots,N}$ . Then, we define the distance matrix  $\mathcal{M}$  by  $\mathcal{M}_{ij} = |X_T^{i,\hat{v}} - Y_T^j|^2$ , and we recall that the set  $U_N$  is defined by [\(III.8\)](#). We approximate the Wasserstein-2 distance between the two empirical distributions formed by  $X_T^{i,\hat{v}}$  and  $Y_T^i$  through the following linear program:

$$\mathcal{W}_2(\hat{\rho}_T, \rho_T) \approx \left( \min_{A \in U_N} \langle A, M \rangle \right)^{1/2}.$$

- **$L^2$  distance.** We will also use as a metric the  $L^2$  distance between  $\hat{m}_T$  and  $m_T$  on the truncated domain  $\tilde{\mathcal{Q}}$ . To evaluate the integral, in the absence of analytical formula for  $m_T$ , we again use a Monte Carlo approach. We uniformly sample  $N$  points in  $\tilde{\mathcal{Q}}$  denoted by  $(X_j)_{j=1,\dots,N}$ . Let  $C(\tilde{\mathcal{Q}})$  denotes the inverse of the value of the uniform density. Then we can approximate the  $L^2$  distance by:

$$\frac{C(\tilde{\mathcal{Q}})}{N} \sum_{i=1}^N \|\hat{m}_T(X_i) - m_T(X_i)\|^2.$$

### III.4.1.3 Numerical results

In the numerical tests, we take the values given in [Table III.1](#) for the parameters of the model, with the time horizon  $T = 1.0$ .

For LQ test 1, in dimension 1, [Figure III.2](#) displays the evolution of the density. For methods 1, 2, and 3, we obtain the control learnt using the neural networks and simulate  $N$  trajectories by the Monte Carlo method following the dynamics [\(III.29\)](#), with  $v$  replaced by the learnt control. We then estimate the mean field distribution using kernel density estimation (KDE). We see that

Test	$d$	$A$	$B$	$\sigma$	$R$	$\bar{x}_0$	$\Sigma_0$	$\bar{x}_T$	$\Sigma_T$
LQ Test 1	1	1	1	1	$\frac{1}{2}$	0.0	1	2.0	0.5
LQ Test 2	2	$I_d$	$I_d$	$I_d$	$\frac{1}{2}I_d$	[0.0, 0.0]	$I_d$	[2.0, 2.0]	$\frac{1}{2}I_d$

Table III.1: Parameters for the two linear-quadratic test cases. For LQ Test 2,  $\bar{x}_0$  and  $\bar{x}_T$  are 2-dimensional vectors with all coordinates equal to 0 or 2 respectively, and  $I_d$  denotes the  $2 \times 2$  identity matrix.

the distributions obtained with the three methods match well the ground-truth one obtained with ODEs. The distributions move towards the right and concentrate around the final mean. [Figure III.3](#) shows the evolution of the control. We see that the three methods provide good approximations of the true optimal control, at least in the region where the density is high. In regions where the density is very low, the control is not well approximated, but this is not an issue as far as the optimal behavior of the population is concerned. The first part of [Figure III.1](#) shows the results obtained for the metrics introduced above. We see that some of the methods achieve a smaller total cost than the true optimal control. This would not be possible for controls satisfying perfectly the terminal constraint, but it is possible here due to the fact that the methods satisfy only approximately the planning constraint. This also explains why the relative errors on the total cost are of the order of a few percents. We expect that it would be possible to reduce the relative errors by further fine-tuning the hyperparameters. Moreover, the optimal control is well approximated, as shown by the  $L^2$  distance to the true optimal control. Furthermore, we see that Methods 2 and 3 have a higher Wasserstein-2 distance between the terminal distribution and the target distribution, but the  $L^2$  distance is much lower.

Test case ( $\mu \pm \sigma$ )	Method	Total Cost	Relative Error	$d_{L^2}(\hat{v}, v^*)(\times 10^{-2})$	$W_2(\hat{\rho}_T, \rho_T)(\times 10^{-2})$	$d_{L^2}(\hat{m}_T, m_T)(\times 10^{-3})$
Linear Quadratic LQ Test 1	ODE ( $v^*$ )	$2.112 \pm 0.001$	-	-	-	-
	M1	$2.223 \pm 0.001$	5.25%	$2.701 \pm 0.872$	$0.6622 \pm 0.388$	$1.772 \pm 1.101$
	M2	$2.088 \pm 0.003$	1.13%	$0.038 \pm 0.024$	$7.119 \pm 3.643$	$0.0003 \pm 0.0002$
	M3	$1.998 \pm 0.056$	5.39%	$4.461 \pm 1.261$	$9.763 \pm 7.542$	$3.826 \pm 5.044$

Figure III.1: Comparison of three different methods v.s. the analytical solution on the LQ test case 1. The evaluation metrics are described in [Section III.4.1.2](#)

As for LQ test 2, in dimension 2, [Figure III.5](#) displays the evolution of the density for each of the methods. The densities move from the bottom left corner to the top right corner. Furthermore, the terminal distribution is more concentrated because the terminal variance is smaller than the initial variance. [Figure III.6](#) shows the evolution of the first dimension of the control (the second dimension is similar, so we omit it for brevity). The ground-truth control is linear in space for each time step. We see that the three methods manage to learn approximately linear controls, at least in the region where the density is significantly positive.

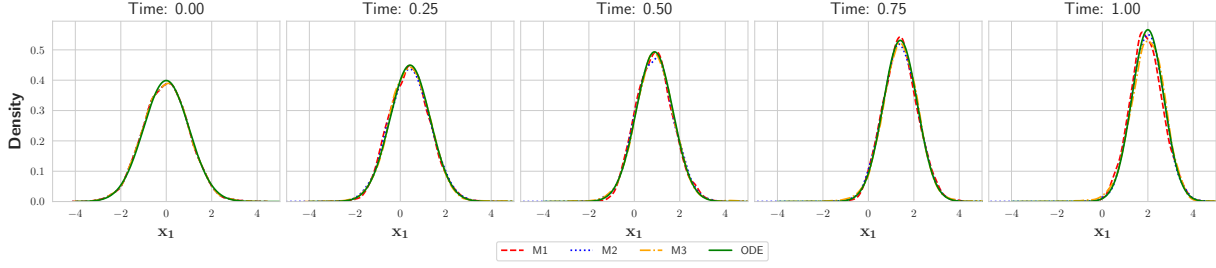


Figure III.2: Evolution of the density in the LQ Test case 1. Each plot corresponds to one time step and displays the densities as functions of the space variable,  $x$ . The densities are: The density obtained by applying the control learnt by each of the three deep learning methods as well as the ground-truth density given by the ODE method.

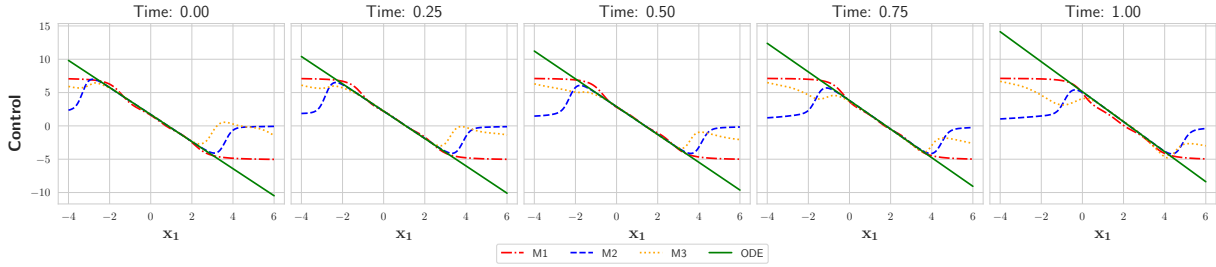


Figure III.3: Evolution of the control in the LQ Test case 1. Each plot corresponds to one time step and displays the controls as functions of the space variable,  $x$ . The controls are: The control learnt by each of the three deep learning methods as well as the ground-truth control given by the ODE method.

Figure III.4 shows the results obtained for the metrics introduced above. We see that here again, each of the three methods achieves a smaller total cost than the true optimal control due to the fact that the terminal constraint is not perfectly satisfied. The optimal control is well approximated, and the terminal distribution is matched with good accuracy.

Test case ( $\mu \pm \sigma$ )	Method	Total Cost	Relative Error	$d_{L^2}(\hat{v}, v^*)(\times 10^{-2})$	$\mathcal{W}_2(\hat{\rho}_T, \rho_T)(\times 10^{-2})$	$d_{L^2}(\hat{m}_T, m_T)(\times 10^{-3})$
Linear Quadratic LQ Test 2	ODE ( $v^*$ )	$4.242 \pm 0.039$	-	-	-	-
	M1	$4.338 \pm 0.049$	2.26%	$7.995 \pm 1.760$	$3.192 \pm 0.708$	$2.205 \pm 0.567$
	M2	$3.822 \pm 0.034$	9.90%	$5.799 \pm 0.329$	$56.86 \pm 17.35$	$0.013 \pm 0.002$
	M3	$4.079 \pm 0.083$	3.84%	$9.348 \pm 1.739$	$52.40 \pm 11.24$	$7.897 \pm 2.648$

Figure III.4: Comparison of three different methods v.s. the analytical solution on the LQ test case 2. The evaluation metrics are described in Section III.4.1.2.

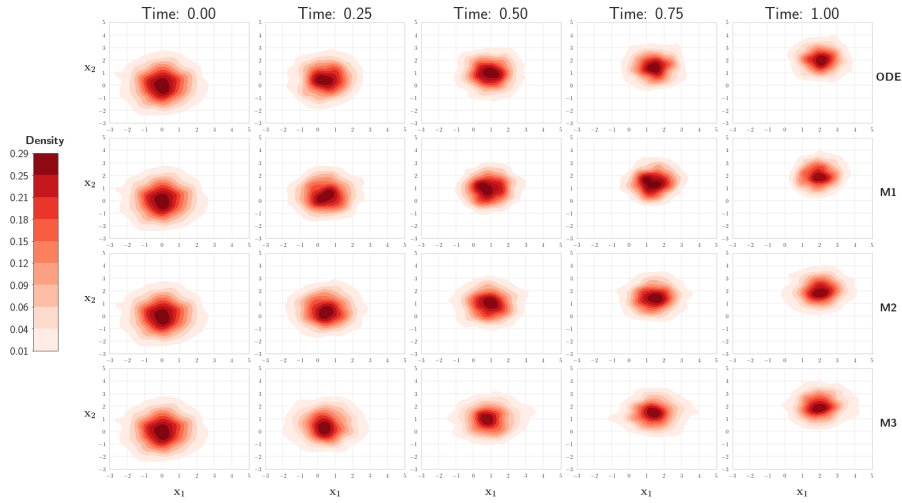


Figure III.5: Evolution of the density in the Linear Quadratic Test case 2. Each column corresponds to one time step, and each row corresponds to one of the methods. Each plot displays the density as a function of the space variable, i.e.,  $(m(t, x)_1)_{x \in [-4, 6]^2}$ . The first row corresponds to the solution obtained by the ground-truth ODE method. The second, third and fourth rows correspond respectively to methods 1, 2 and 3.

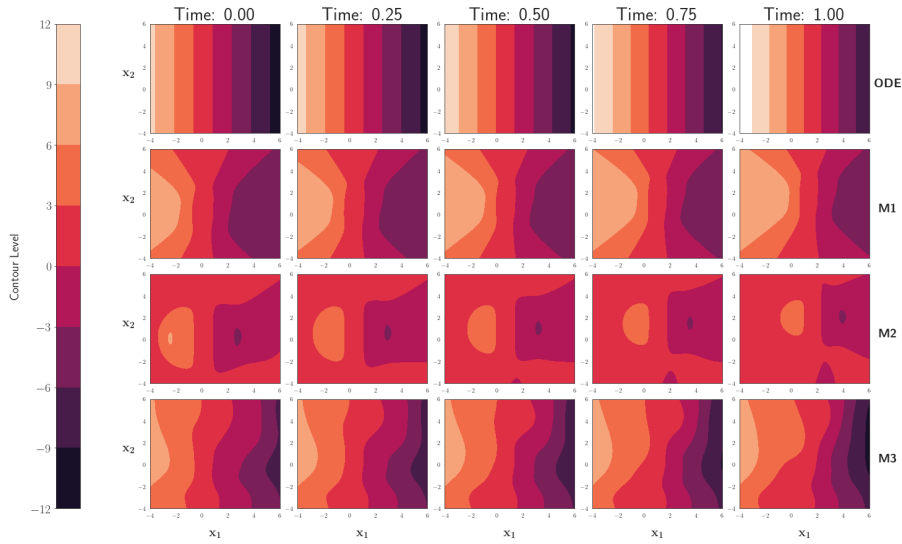


Figure III.6: Evolution of the control in the Linear Quadratic Test case 2. Each column corresponds to one time step, and each row corresponds to one of the methods. Each plot displays the first dimension of the control as a function of the space variable, i.e.,  $(v(t, x)_1)_{x \in [-4, 6]^2}$ . The first row corresponds to the solution obtained by the ground-truth ODE method. The second, third and fourth rows correspond respectively to methods 1, 2 and 3.

### III.4.2 Case 2: Transport with congestion effects

The second class of models that we consider is inspired by crowd motion and falls in the setting discussed in Example 3.

#### III.4.2.1 Description of the problem

In this model, intuitively, the cost is higher when moving through a crowded region, i.e., where the density is high. Specifically, we take:

$$b(x, \mu, a) = a, \quad f(x, \mu, a) = R|\ell(x, \mu)|^\gamma |a|^2, \quad \rho_0 = \mathcal{N}(\bar{x}_0; \Sigma_0), \quad \rho_T = \mathcal{N}(\bar{x}_T; \Sigma_T).$$

For the function  $\ell$ , we take two different models. We consider the following non-local dependence:

$$\ell(x, \mu) = c + \rho_\epsilon \star \mu(x),$$

where  $c > 0$  is a constant,  $\rho_\epsilon$  is a Gaussian kernel and  $\star$  denotes the convolution. We use Method 1 to solve the MFOT problem with this function  $\ell$ . Since it is based on Monte Carlo simulations of trajectories, it is straightforward to compute a convolution with the empirical distribution at a given time step.

We also consider a variation with a local dependence.

$$\ell(x, \mu) = c + m(x)$$

where  $c > 0$  is a constant and  $m$  denotes the density of  $\mu$ . For this type of model, Methods 2 and 3 are better suited since, in these methods, we directly have access to the approximate density in the form of a neural network.

#### III.4.2.2 Numerical results

We focus on one test case called "Congestion" below in dimension  $d = 1$ . In this model,  $\gamma = 1$ . For the sake of comparison, we also consider the corresponding model with the same choice of parameters except that  $\gamma = 0$ , i.e., there are no congestion effects in the running cost. The values that we take in the numerical tests are given in Table III.2 below.

Test case	$d$	$\gamma$	$c$	$\sigma$	$R$	$\bar{x}_0$	$\Sigma_0$	$\bar{x}_T$	$\Sigma_T$
Case 1, No congestion	1	0	0.1	0.1	0.5	0	0.04	2	0.04
Case 2, Congestion	1	1	0.1	0.1	0.5	0	0.04	2	0.04
Case 3, Congestion	5	1	1	$I_d$	0.5	$[0, \dots, 0]$	$0.1 I_d$	$[2, \dots, 2]$	$0.1 I_d$

Table III.2: Parameters for the test case with congestion and the benchmark model without congestion effects. For Case 3,  $\bar{x}_0$  and  $\bar{x}_T$  are 5-dimensional vectors with all coordinates equal to 0 or 2 respectively, and  $I_d$  denotes the  $5 \times 5$  identity matrix.



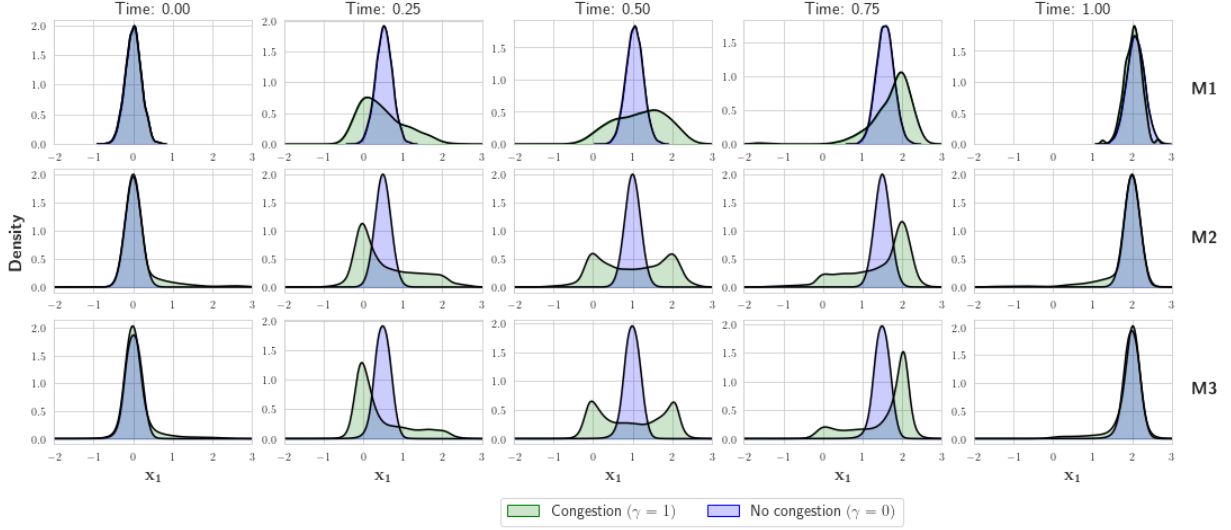


Figure III.7: Visualization of the mean field density  $\hat{m}(t, x)$  in Congestion test Case 1,2

In [Figure III.7](#), we present the evolution of the density under the control learnt by each of the three methods for congestion cases 1 and 2. Each row corresponds to one method. We see that, in the case where  $\gamma = 0$  (no congestion effect), the mass is transported directly towards the terminal distribution without much change in its shape. In contrast, in the case with  $\gamma = 1$ , the mass spreads in space and one part starts moving towards the target mean  $\bar{x}_T = 2$  whereas another part stays behind and catches up at later time steps. This is consistent with the idea that moving in congested regions is more expensive, so some agents would agree to wait until the density decreases before moving forward.

Finally, in [Figure III.8](#) we present the evolution of the density under the control learnt by each of the three methods for congestion case 3, which is in dimension 5. Each row corresponds to one method. To visualize density evolution in dimension 5, we plot the marginal distribution of the mean field distribution on the first and second dimensions. We see that, similarly to the congestion case 2, the mass spreads in space and gradually moves towards the target mean. Compared with congestion case 2, the difference in the moving pattern and extent of spreading is due to the difference of parameters in [Table III.2](#). With a larger value  $c$ , the behavior of the density would be closer to a direct transport to the terminal distribution without changes in the shape of the distribution.

### III.4.3 Remarks on the choice of hyperparameters

Each method has several hyperparameters, including the architecture of the neural networks. We provide below some remarks about the choice of hyperparameters in our implementation.

**Method 1.** In our implementation, we choose  $G(r) = C_W r$  where  $C_W$  is a hyperparameter that we adjust dynamically. We increase the constant  $C_W$  when we expect a higher running cost (for instance, in a higher dimension) in order to give enough importance to the penalty. The coefficient  $\alpha$  of regularization for the computation of the Wasserstein distance is also a

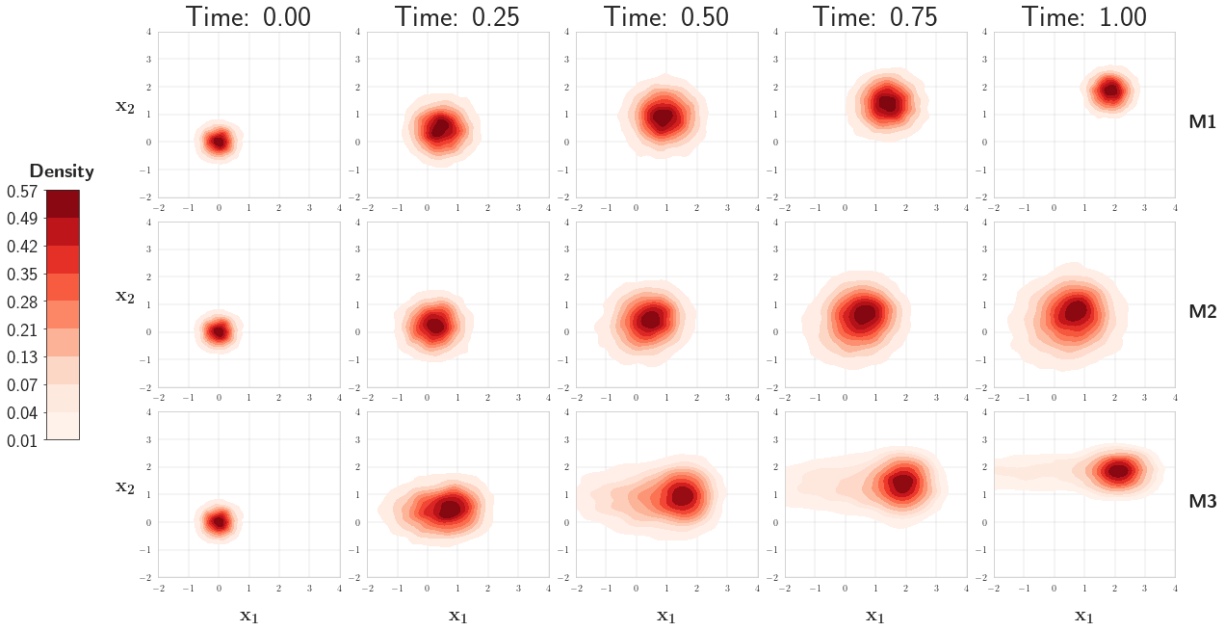


Figure III.8: Visualization of the mean field density  $\hat{m}(t, x)$  in Congestion test Case 3

hyperparameter that we adjust dynamically using the following heuristics. We start with a given value for  $\alpha$  and, when the estimated Wasserstein distance is small enough, we reduce the value of  $\alpha$ . The idea is that, as long as the terminal distribution does not match well enough the target distribution, we need a high level of regularization in order to estimate efficiently the Wasserstein distance between them. As the two distributions get closer, we can decrease the degree of regularization in order to have a more accurate estimation of the Wasserstein distance. The way we adjust  $C_W$  also depends on the dimension of the state variable. There is also a computational time aspect to take into account: as  $\alpha$  becomes smaller, the computations take more time (see [Section III.3.1.2](#) for more details). For the neural network, we take a feedforward fully connected neural network with 6 layers of 60 neurons each. The other hyperparameters are the number of particles  $N$  and the number of time steps  $N_T$ . We take  $N = 300$  and  $N_T = 20$ .

**Method 2.** In the second method, no time or space discretization is needed, and the density is directly approximated by a neural network, so we do not need to use a finite number of particles. However, we need to choose the values of the weights  $C_0^{(\text{KFP})}$ ,  $C_T^{(\text{KFP})}$ ,  $C^{(\text{KFP})}$ , and  $C^{(\text{HJB})}$  in the loss function. We used  $C_0^{(\text{KFP})} = 20$ ,  $C_T^{(\text{KFP})} = 50$ ,  $C^{(\text{KFP})} = 20$ ,  $C^{(\text{HJB})} = 1$ . As for the neural network, we used the architecture proposed in the DGM article [\[SS18\]](#), with 2 layers and a width equal to 40. During the training, at each iteration of SGD, we use a minibatch of 500 points in time and space, and 500 points in space for the initial and terminal conditions.

**Method 3.** The main hyperparameter in this method is  $r$ , which is used in the definition of the augmented Lagrangian [\(III.24\)](#). For the experiments, we select  $r = 0.1$ . Even though, in theory, the convergence of ADMM is independent of the choice of  $r$ , in practice, we often find

that a large  $r$  value could potentially increase numerical instability and lead the algorithm to diverge. Similarly, a small  $r$  value could slow down the convergence. As for the neural networks, we use the following architectures. For both  $u_\theta$ ,  $q_\omega$ , and  $\lambda_\psi$ , in general, we use a fully connected neural network with residual connections, sigmoid activation function, and appropriate output dimension. We use 6 layers and 100 neurons per layer. For LQ test cases, we further consider an extra quadratic correction in addition to the neural networks: the output of  $u_\theta$  is the sum of neural network output and a quadratic function with trainable weights. To effectively model the mean field density, a sigmoid activation function is applied to the first dimension of the output of the neural network  $\lambda_\psi$ , and then the result is multiplied by a constant  $C$ . In this way, the first dimension of  $\lambda_\psi$  takes values in  $(0, C)$ . In the experiments, we take  $C = 1$  for the LQ test cases and  $C = 5$  for the congestion test cases. During training, at each iteration of SGD, we use a minibatch of 512 points in time and space, and 512 points in space for the initial and terminal conditions.

### III.5 Conclusion and future directions

In this work, we have proposed three numerical methods based on deep learning for mean field optimal transport problems. The three methods can tackle a larger class of problems than deep learning methods proposed previously, which were mostly focusing on the Schrödinger bridge problem or MFGs with a specific structure. The first method replaces the terminal constraint with a penalty and then directly learns the optimal control using Monte Carlo trajectories. The second method solves a PDE system which is obtained as the optimality conditions for the MFOT problem. The third method relies on an augmented Lagrangian approach for the variational formulation of the problem. The numerical results show that the three methods match the analytical solution on an LQ problem, and that they are able to handle non-trivial mean field interactions modeling congestion effects.

From here, we can envision several research directions. First of all, the theoretical analysis of the MFOT problem remains to be tackled. For example, the existence and uniqueness of the solution to the PDE system have been proved only in relatively specific cases, see e.g. [\[ACCD12\]](#), [\[CCN13\]](#), [\[Por13\]](#), [\[GMST19\]](#). It would be interesting to extend the analysis to more general forms of dynamics and cost functions. From the numerical point of view, it would be interesting to scale-up the methods proposed in this work to a higher dimension, and to explore other deep learning methods. The numerical analysis and the convergence proof of the proposed methods also remain to be investigated in future work.



# Appendix A

## Construction of a coupling by pasting

### A.1 Pasting of Markov processes

In this section we sum up the tools we used to provide a formal construction of the  $N$ -BMP, the  $b$ -BMP and the couplings between them that appear in our proofs. We adapt the results of [INW68b] and we provide two additional results. In [INW68b], Ikeda, Nagasawa and Watanabe build a new strong Markov process by "piecing out" trajectories of a given strong Markov process (which we call "base process" in the following): the new process follows a trajectory of the base Markov process until a random time, and then undergoes a random renewal, before following another trajectory of that same Markov process from the new starting point; then the same mechanism is repeated iteratively to build the trajectory at all positive times.

This construction is presented in [Section A.1.1](#).

In [Section A.1.2](#), we give alternative definitions of the BMP,  $N$ -BMP and  $b$ -BMP using this pasting procedure, that fit into our proof strategy of [Proposition I.4](#), that appears as a corollary of [Proposition A.4](#).

In [Section A.1.3](#), we study what happens when several types of pasting are allowed (for instance in our case, we will distinguish branching times from death times). In that case, there are two points of view: one can either use the procedure once, considering all the pasting times, or use it several times, once for each type of pasting time. We prove in [Proposition A.2](#) that the resulting processes have the same distribution, meaning that this construction is, in a way, commutative (we can chose which type of pasting we use first), which turns out to be useful in the proof of [Proposition A.4](#).

Finally, we prove in [Section A.1.4](#) that, if the base process is a coupling of  $n$  strong Markov processes, and if one of the  $n$  coordinates is unaffected by renewal events, that coordinate in the pasted process has the same distribution as the same coordinate in the base process (see [Proposition A.3](#) for a precise statement).

#### A.1.1 Construction by piecing out

Let us first explain the construction presented in [INW68b], as well as the modification we use to adapt it to our context. In particular, their construction uses two elementary objects: the base Markov process and the renewal distribution; they paste versions of the base process by renewing it according to the renewal distribution at the end of its lifetime. We add a third elementary piece to their construction, because we would like to paste the process at a more general random time than its lifetime: we will paste versions of the base process according to a general pasting time and a renewal distribution.

*Base process.* Consider a locally compact Hausdorff space  $E$  with countable base, and denote by  $\mathcal{B}(E)$  its Borel  $\sigma$ -algebra. Let  $\overline{E} = E \cup \{\Delta\}$  the one-point compactification of  $E$ , where  $\Delta$  is an isolated cemetery point, and denote by  $\mathcal{B}(\overline{E})$  its Borel  $\sigma$ -algebra.

In order to characterize the distribution of a random variable with values in  $\overline{E}$ , we will use the set  $\mathcal{C}_0(\overline{E}) = \{f \in \mathcal{C}^0(\overline{E}, \mathbb{R}) \text{ s.t. } f(\Delta) = 0\}$ . Furthermore, we can also characterize the distribution of Markov processes with values in  $\overline{E}$  by their semigroup on  $\mathcal{C}_0(\overline{E})$ , and any Markov process in  $E$  with semigroup  $(\mathcal{P}_t)_{t \geq 0}$  on  $\mathcal{C}_0(E)$  can be seen as a Markov process in  $\overline{E}$  by extending its semigroup as follows

$$\forall f \in \mathcal{C}_0(\overline{E}) \forall y \in \overline{E}, \quad \overline{\mathcal{P}}_t f(y) = \begin{cases} \mathcal{P}_t f(y) & \text{if } y \in E \\ 0 & \text{if } y = \Delta \end{cases}$$

In the sequel we will omit the difference between those two objects and write  $\mathcal{P}_t$  instead of  $\overline{\mathcal{P}}_t$  for simplicity.

Let  $(W, \mathcal{W}, (\mathcal{W}_t)_{t \geq 0}, (\mathbb{P}_y)_{y \in \overline{E}}, (\mathcal{Y}_t)_{t \geq 0}, (\psi_t)_{t \geq 0})$  be a strong Markov process<sup>1</sup> on  $\overline{E}$  (here  $(W, \mathcal{W})$  is a measurable space,  $(\mathcal{W}_t)_{t \geq 0}$  a filtration on  $(W, \mathcal{W})$ ,  $(\mathbb{P}_y)_{y \in \overline{E}}$  a family of probability measures on  $(W, \mathcal{W})$ ,  $(\mathcal{Y}_t)_{t \geq 0}$  a family of random variables with values in  $E$  and for all  $t \geq 0$ ,  $\psi_t : W \rightarrow W$  is a shift operator) with the following assumptions:

- (a) for all  $w$  in  $W$ ,  $\mathcal{Y}(w)$  has right-continuous paths,
- (b) for all  $y$  in  $\overline{E}$ ,  $\mathcal{Y}$  has  $\mathbb{P}_y$ -almost surely left limits at every time,
- (c) for all  $y$  in  $\overline{E}$ ,  $\mathbb{P}_y(\mathcal{Y}_0 = y) = 1$ ,
- (d)  $\Delta$  is a trap point, meaning that for  $y \in \overline{E}$ ,

$$\mathbb{P}(\{w \in W, \exists t \geq s, \mathcal{Y}_s(w) = \Delta \text{ and } \mathcal{Y}_t(w) \neq \Delta\}) = 0.$$

- (e) the filtration  $(\mathcal{W}_t)_{t \geq 0}$  is complete<sup>2</sup> and right-continuous, contained in  $\mathcal{W}$ ,
- (f) the  $\psi_t$  are shift operators<sup>3</sup> defined on  $W$ , such that for all  $s, t \geq 0$  and  $w \in W$ , we have

$$\mathcal{Y}_s(\psi_t w) = \mathcal{Y}_{s+t}(w).$$

*Pasting time.* Our goal is to paste together countably many versions of that base process at successive random times. In [INW68b], the pasting time is the lifetime of the base process, but we would like to extend this to more general random times, such as hitting times. To that end, suppose given a stopping time  $\zeta^\dagger$  on  $W$ , meaning that for all  $t \geq 0$ ,  $\{\zeta^\dagger \leq t\}$  is in  $\mathcal{W}_t$ . Furthermore, we say that  $\zeta^\dagger$  is a *compatible stopping time* for  $\mathcal{Y}$  if it satisfies the following compatibility<sup>4</sup> equation with respect to the base Markov process

$$\forall t \geq 0 \forall w \in W \text{ s.t. } \zeta^\dagger(w) \geq t, \quad \zeta^\dagger(\psi_t w) = \zeta^\dagger(w) - t. \quad (\text{A.1})$$

We then make the following assumption:

<sup>1</sup>We call a strong Markov process a stochastic process that enjoys the strong Markov property, see for instance [Kal02], Proposition 19.17.

<sup>2</sup>Namely, the filtration is complete with respect to  $(W, \mathcal{W}, \mathbb{P}_y)$ , for all  $y \in E$ , in the sense that for all  $t \geq 0$ ,  $\mathcal{W}_t$  contains all the  $\mathbb{P}_y$ -null events of  $W$ .

<sup>3</sup>If  $T$  is random time on  $W$ , we define for  $w \in W$  the shift operator  $\psi_T w := \psi_{T(w)} w$ .

<sup>4</sup>Any hitting time for  $\mathcal{Y}$  is a compatible stopping time. These stopping times are natural extensions of hitting times.

(g)  $\mathbb{P}_\Delta(\zeta^\dagger = 0) = 1$  and  $\zeta^\dagger$  is a compatible stopping time.

In order to stick with the construction of [INW68b], we introduce the killed version of the process  $\mathcal{Y}$ : it is the process  $\bar{\mathcal{Y}}$  defined by

$$\forall w \in W \quad \forall t \geq 0 \quad \bar{\mathcal{Y}}_t(w) = \begin{cases} \mathcal{Y}_t(w) & \text{if } t < \zeta^\dagger(w), \\ \Delta & \text{otherwise,} \end{cases}$$

which, according to the assumptions on  $\mathcal{Y}$ , is a right-continuous strong Markov process which almost surely has left limits,  $\Delta$  as a terminal point<sup>5</sup> and whose lifetime is

$$\zeta := \zeta^\dagger \wedge \inf\{s \geq 0, \mathcal{Y}_s = \Delta\} = \inf\{s \geq 0, \bar{\mathcal{Y}}_s = \Delta\}. \quad (\text{A.2})$$

Indeed, if the process is not yet in the cemetery at time  $\zeta^\dagger$ , we kill it at this moment. Notice that since  $\inf\{s \geq 0, \bar{\mathcal{Y}}_s = \Delta\}$  is a hitting time for the process  $\bar{\mathcal{Y}}$ , which is also compatible with the shift operators (Assumption (f)),  $\zeta$  still satisfies (A.1) (Assumption (g)). Furthermore, if for a fixed  $y$  in  $E$ ,  $\mathcal{Y}$  has an infinite lifetime  $\mathbb{P}_y$ -almost surely, we also have  $\zeta^\dagger = \zeta$ .

**Remark A.1.** Notice that  $\zeta$  coincides with  $\zeta^\dagger$  when  $\mathcal{Y}$  has an infinite lifetime. In the following and most importantly in Section A.2, we will consider processes enjoying that property, so that we will drop the  $\dagger$  notation that distinguish them; for instance Assumption (h) below should be written with the  $\dagger$  symbol but since  $\mathcal{Y}_{\zeta_j} = \mathcal{Y}_{\zeta_j^\dagger}$ , we lighten the notation and forget it. The  $\dagger$  symbol is only introduced to allow us to fit precisely within the framework of [INW68b].

*Renewal distribution.* We finally have to explain how the renewal point is chosen: to that end, we shall use a particular case of so-called *instantaneous distributions* (see Definition 2.1 in [INW68b]), in the form of a Markov kernel  $\lambda : W \times \mathcal{B}(\bar{E}) \rightarrow \mathbb{R}$  satisfying the following assumptions:

(h) there exists an integer  $k \geq 1$ , and, for  $1 \leq j \leq k$ , Markov kernels  $\bar{\lambda}_j : \bar{E} \times \mathcal{B}(\bar{E}) \rightarrow \mathbb{R}$  and almost surely finite stopping times  $\zeta^j$  satisfying (g) such that  $\zeta^\dagger(w) = \min_j \zeta^j(w)$  and<sup>6</sup>

$$\lambda(w, \cdot) = \bar{\lambda}_j(\mathcal{Y}_{\zeta^j}(w), \cdot) \text{ if } \zeta^\dagger(w) = \zeta^j(w). \quad (\text{A.3})$$

(i) each  $\bar{\lambda}_j$  sends  $E$  to  $E$  and  $\Delta$  to  $\Delta$ , meaning that  $\bar{\lambda}_j(y, E) = 1$  for  $y \in E$  and  $\bar{\lambda}_j(\Delta, \cdot) = \delta_\Delta$ .

In the sequel, we denote by  $\lambda_j$  the Markov kernel  $\bar{\lambda}_j(\mathcal{Y}_{\zeta^j}, \cdot)$  and by  $M^{(\zeta^j)_j}((\lambda_j)_j)$  the Markov kernel  $\lambda$  defined in (A.3) (in order to emphasize its dependence on the  $\zeta^j$  and the  $\lambda_j$ ). Assumption (h) ensures that this is indeed an instantaneous distribution<sup>7</sup>.

<sup>5</sup>While there may be other trap points for  $\mathcal{Y}$ , we refer to  $\Delta$  as terminal point because we consider that the process has terminated its life when it has reached  $\Delta$ ; this is also the reason why we refer to  $\zeta$  as its lifetime.

<sup>6</sup>If several indices realize the minimum, we arbitrarily chose the smallest one.

<sup>7</sup>Since, for any stopping time  $T$ ,  $\mathcal{Y}_{\zeta^\dagger} \circ \psi_T = \mathcal{Y}_{\zeta^\dagger}$  on the event  $\{T < \zeta\} \subseteq \{T < \zeta^\dagger\}$  (thanks to the compatibility equation (A.1)), we have, looking at the definition (A.3) of  $\lambda$ , that

$$\mathbb{P}_y(\lambda(\cdot, dx) = \lambda(\psi_T(\cdot), dx), T < \zeta) = \mathbb{P}_y(T < \zeta),$$

which matches indeed Definition 2.1 in [INW68b].

*Definition of the pasted version of  $\mathcal{Y}$  with respect to  $(\zeta, \lambda)$ .* In a nutshell, using a renewal distribution satisfying Assumption (h) means that at time  $\zeta$  the process is randomly renewed to a position whose distribution depends only on its state at time  $\zeta$ : basically, the process follows a trajectory decided by a certain  $w_1 \in W$ , then jumps to a new position according to  $\lambda(w_1, \cdot)$  and then follows a new trajectory decided by  $w_2 \in W$ , jumps to a new position according to  $\lambda(w_2, \cdot)$ , and so on (see [Figure A.1](#) below for an example). This amounts to defining new probability measures on the measurable space  $\Omega = W \times \bar{E}$  equipped with the product  $\sigma$ -algebra  $\mathcal{W} \otimes \mathcal{B}(\bar{E})$ , where the  $W$ -coordinate contains the information for one trajectory, and the  $\bar{E}$ -coordinate for one renewal point; and then paste a sequence of such probability measures on the infinite product  $\mathfrak{P}W = \prod_{i=1}^{+\infty} \Omega = \prod_{i=1}^{+\infty} (W \times \bar{E})$  (equipped with the corresponding product  $\sigma$ -algebra that we denote by  $\mathfrak{P}\mathcal{W}$ ).

For  $y \in \bar{E}$ , we define a probability measure  $\mathbb{Q}_y^\lambda$  on  $\Omega$ , uniquely characterized by

$$\forall A \in \mathcal{W} \forall B \in \mathcal{B}(\bar{E}), \quad \mathbb{Q}_y^\lambda(A \times B) = \int_A \lambda(w, B) \mathbb{P}_y(dw). \quad (\text{A.4})$$

Now, we use Theorem 2.1 in [\[NW68b\]](#) to construct by concatenation a uniquely defined probability measure  $\mathfrak{P}^\zeta \mathbb{P}_y^\lambda$  on the product  $\mathfrak{P}W$  such that for any bounded real-valued random variable  $Z_i$  on  $\Omega^i$ , we have

$$\mathfrak{P}^\zeta \mathbb{E}_y^\lambda(Z) = \int_{\Omega^i} Z_i(\omega_1, \dots, \omega_i) \mathbb{Q}_{y_{i-1}}^\lambda(dw_i) \cdots \mathbb{Q}_{y_1}^\lambda(dw_2) \mathbb{Q}_y^\lambda(dw_1), \quad (\text{A.5})$$

where we have written  $\omega_i = (w_i, y_i)$ ,  $\mathfrak{P}^\zeta \mathbb{E}_y^\lambda$  the expectation under  $\mathfrak{P}^\zeta \mathbb{P}_y^\lambda$  and  $Z$  the extension of  $Z_i$  on  $\mathfrak{P}W$ , namely  $Z(\omega_1, \omega_2, \dots) = Z_i(\omega_1, \dots, \omega_i)$ .

We can finally build on  $\mathfrak{P}W$  the new process that corresponds to the pasting of trajectories as described above. We generically write  $\omega^i = (\omega_1, \dots, \omega_i)$  an element of  $\Omega^i$ , and  $\mathfrak{P}w = (\omega_i)_{i \geq 1} = (w_i, y_i)_{i \geq 1}$  an element of  $\mathfrak{P}W$ . Now set

$$\left\{ \begin{array}{l} \zeta_i(\mathfrak{P}w) = \zeta(w_i), \\ \bar{\zeta}_i(\mathfrak{P}w) = \sum_{k=1}^i \zeta(w_k), \\ \mathfrak{T}(\mathfrak{P}w) = \sum_{k=1}^{+\infty} \zeta(w_k), \end{array} \right.$$

as well as  $\bar{\zeta}_0(\mathfrak{P}w) = 0$ . We note that  $\bar{\zeta}_i$  depends only on the first  $i$  coordinates, which allows us to see  $\bar{\zeta}_i$  as a random variable on  $\Omega^i$ , for which we still use the same notation. Then we define the pasted process  $\mathfrak{P}^\zeta \mathcal{Y}$  on  $\mathfrak{P}W$  by<sup>8</sup>

$$\mathfrak{P}^\zeta \mathcal{Y}_t(\mathfrak{P}w) = \begin{cases} \mathcal{Y}_{t - \bar{\zeta}_{i-1}(\mathfrak{P}w)}(w_i) & \text{if } \bar{\zeta}_{i-1}(\mathfrak{P}w) \leq t < \bar{\zeta}_i(\mathfrak{P}w), i \geq 1, \\ \Delta & \text{otherwise.} \end{cases} \quad (\text{A.6})$$

**Remark A.2.** We emphasize that we have defined a process  $\mathfrak{P}^\zeta \mathcal{Y}$ , whose position at time  $t$  is  $\mathfrak{P}^\zeta \mathcal{Y}_t$ . We could have written  $[\mathfrak{P}^\zeta \mathcal{Y}]_t$  instead, but chose not to in order to alleviate the notation.

<sup>8</sup>In order to stick with the definitions of [\[NW68b\]](#), we should replace  $\mathcal{Y}$  by its killed version in this definition; however, since  $\mathcal{Y}_t = \bar{\mathcal{Y}}_t$  on  $\{t < \zeta\}$ , the corresponding quantities are equal. This allows us to lighten the notation in the sequel.



The results in [INW68b] and our additional assumptions then ensure<sup>9</sup> the following result.

**Proposition A.1.** *The process  $(\mathfrak{P}^\zeta \mathcal{Y}_t)_{t \geq 0}$  is a strong Markov process on  $\mathfrak{PW}$ , using a filtration  $(\mathfrak{P}^\zeta \mathcal{W}_t)_{t \geq 0}$  and shift operators  $\mathfrak{P}^\zeta \psi_t$  on  $\mathfrak{PW}$  defined as in [INW68b], and the probability measures  $(\mathfrak{P}^\zeta \mathbb{P}_y^\lambda)_{y \in \bar{E}}$ . Furthermore, it still enjoys Assumptions (a) to (f).*

We emphasize that, included in the fact it is a Markov process,  $\mathfrak{P}^\zeta \mathcal{Y}$  is adapted to the filtration (i.e.  $\mathfrak{P}^\zeta \mathcal{Y}_t$  is  $\mathfrak{P}^\zeta \mathcal{W}_t$ -measurable).

### A.1.2 Formal definition of some branching processes with selection

We now give equivalent definitions of the processes used in this paper, using the pasting procedure described in the last section.

We first recall the general context of this paper: we consider several branching processes built from an underlying Markov process  $X$ , which is assumed to be Feller. As we already mentioned, this implies that there exists a càdlàg realization of  $X$  which has the strong Markov property (see Theorems 19.15 and 19.17 in [Kal02]). Thus we can consider base Markov processes consisting in any finite number of independent copies of  $X$  satisfying Assumptions (a) to (f), and use the pasting procedure on such processes.

**Definition A.1.** *Set  $E = \bigsqcup_{n \geq 1} \mathbb{R}^n$  and  $\mathcal{X}$  a Markov process on  $E$  consisting in a fixed number of independent copies of the underlying Markov process starting from the initial positions given by an  $n$ -tuple  $x = (x^1, \dots, x^n)$  in a certain  $\mathbb{R}^n$ : we define  $\mathcal{X}_t = (X_t^1, \dots, X_t^n)$ , where  $(X_t^i)_{t \geq 0}$  has under  $\mathbb{P}_x$  the distribution of  $X$  starting at  $x^i$ , and the  $X^i$  are independent processes.*

*For  $x \in \mathbb{R}^n$ , we denote by  $\bar{\lambda}(x, \cdot)$  the distribution of  $(x^1, \dots, x^n, x^I)$ , where  $I$  has uniform distribution on  $\llbracket 1, n \rrbracket$ : thus the renewal state is uniformly chosen within the set of the  $(n+1)$ -tuples made by adding an  $(n+1)$ -th coordinate with the same value as one of the existing  $n$  coordinates of  $x$ .*

*On the same probability space, suppose<sup>10</sup> given a random time  $\zeta^\dagger$  with exponential  $\mathcal{E}(n)$  distribution<sup>11</sup> under  $\mathbb{P}_x$ , independent from  $\mathcal{X}$  and satisfying the compatibility equation [A.1].*

*Since  $\mathcal{X}$  has an infinite lifetime (the process takes values in  $E$ , so that it never hits the cemetery point), we have  $\zeta = \zeta^\dagger$ . We now write  $\lambda(w, \cdot) = \bar{\lambda}(\mathcal{X}_\zeta(w), \cdot)$  as in the former section, which corresponds to the fact that the renewal position is made by letting the chosen particle branch at time  $\zeta$ .*

*By using the pasting procedure with the base process  $\mathcal{X}$ , the pasting time  $\zeta$  and the renewal distribution  $\lambda$ , we obtain a strong Markov process  $\mathfrak{P}^\zeta \mathcal{X}$  as in [Proposition A.1], and we call branching Markov process (BMP) any process that has the same distribution as this pasted process  $\mathfrak{P}^\zeta \mathcal{X}$ .*

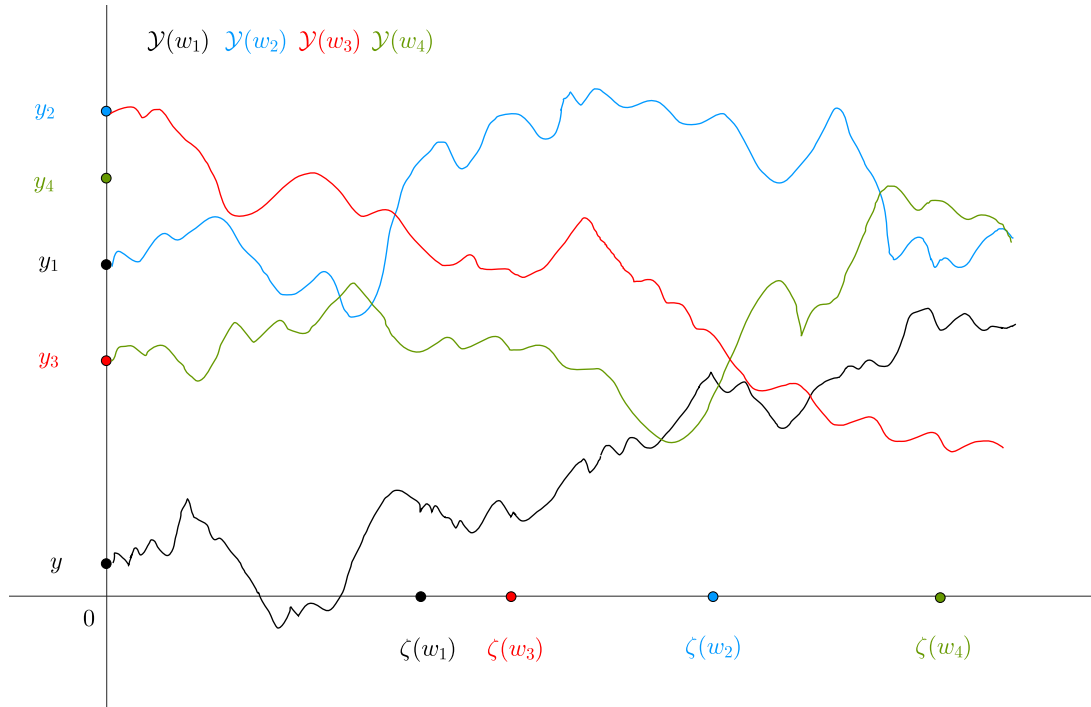
#### Remark A.3.

- *In order to use the many-to-one lemma ([Lemma I.A]), we need a notion of genealogy allowing us to follow trajectories of individual particles. To do so, we could keep track of*

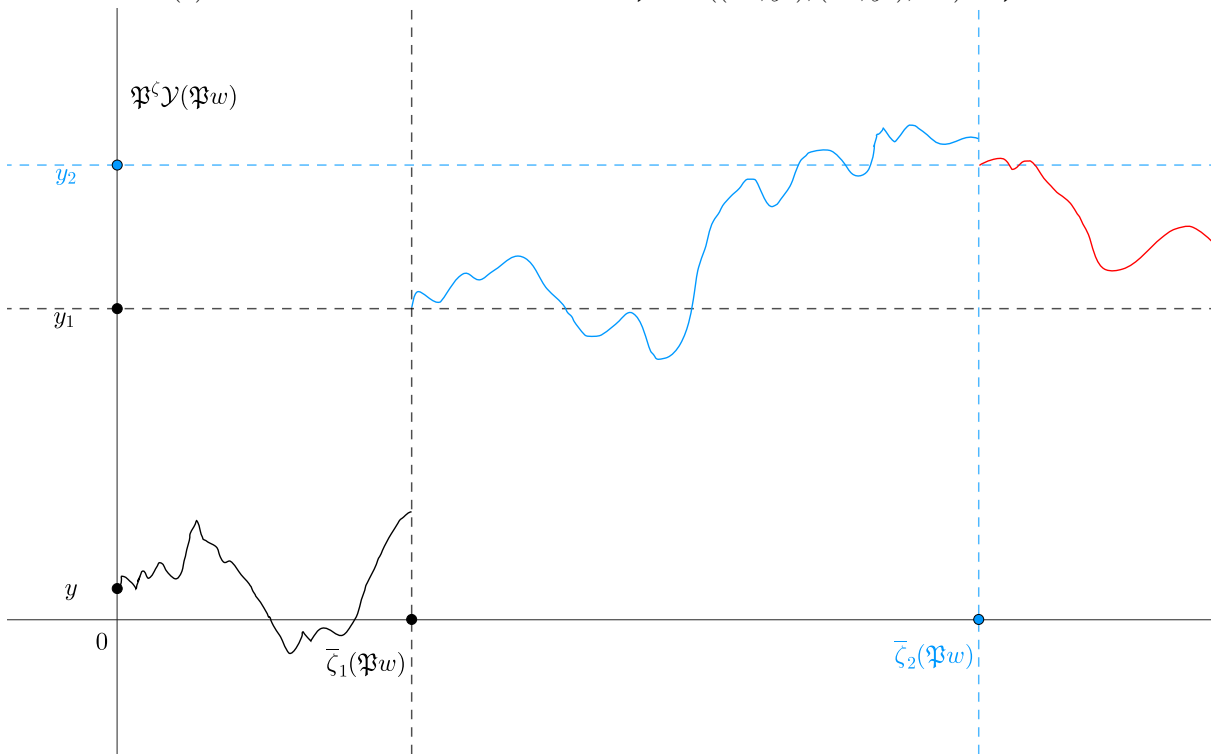
<sup>9</sup>The fact that  $\mathfrak{P}^\zeta \mathcal{Y}$  is a strong Markov process comes from Theorem 2.2 in [INW68b], which we can apply thanks to Assumptions (a),(c),(d),(e),(f),(g),(h) and (i). We give for each assumption the main reason why it still holds for the pasted process: (a):  $\mathfrak{P}^\zeta \mathcal{Y}$  has right-continuous paths by construction (because of (a) for  $\mathcal{Y}$ ); (b): it has left limits thanks to Theorem 2.3 in [INW68b] and (b) for  $\mathcal{Y}$ ; (c)  $\mathfrak{P}^\zeta \mathbb{P}_y^\lambda(\mathfrak{P}^\zeta \mathcal{Y}_0 = y) = \mathbb{P}_y(\mathcal{Y}_0 = y) = 1$ ; (d): Assumption (i) for  $\mathcal{Y}$  ensures that the pasted process remains in  $\Delta$ ; (e): the corollary of Theorem 2.2 in [INW68b] ensures that we have a complete and right-continuous filtration; (f): this comes from the definition (2.11) of  $\mathfrak{P}^\zeta \psi_t$  in [INW68b] and the Assumption (g) (compatibility equation).

<sup>10</sup>One possibility would be to consider the canonical realization as is done in [Section A.2.1.2] for the coupling of the  $N$ -BMP with a  $b$ -BMP.

<sup>11</sup>Notice that the distribution depends on the number of particles in the starting position  $x$ .



(a) The first coordinates of an element  $\mathfrak{P}w = ((w_1, y_1), (w_2, y_2), \dots) \in \mathfrak{P}W$



(b) The realization of the pasted process made with this element  $\mathfrak{P}w$

Figure A.1: Illustration of the pasting procedure

the ancestries of particles by adding a discrete coordinate to each of them containing the index of its parent. We chose not to do it here in order to alleviate the construction.

- Let us emphasize that if two or more particles have the same position at the branching event, the corresponding tuple appears several times in the set on which the uniform choice is performed: we decide uniformly which particle branches, not at which position the branching particle is.

We also construct the  $N$ -BMP in a similar way, using the pasting procedure at branching times, and performing the branching and selection simultaneously by making the lowest particle jump to the place of one of the particles, chosen uniformly, *via* the choice of the renewal distribution as follows.

**Definition A.2.** Set  $E_1 = \mathbb{R}^N$  and consider a Markov process  $\mathcal{X}$  consisting in  $N$  independent copies of the underlying Markov process  $X$  on a certain probability space  $W$ . As in the previous example, let  $\zeta$  be an independent and exponentially distributed random time with rate  $N$  satisfying (A.1). For  $x \in \mathbb{R}^N$  and  $m$  such that<sup>12</sup>  $x^m = \min x^i$ , we denote by  $\bar{\lambda}(x, \cdot)$  the distribution of  $(x^1, \dots, x^{m-1}, x^I, x^{m+1}, \dots, x^N)$ , where  $I$  has uniform distribution on  $\llbracket 1, N \rrbracket$ . We also define  $\lambda(w, \cdot) = \bar{\lambda}(\mathcal{X}_\zeta(w), \cdot)$ . We call  $N$ -BMP any stochastic process that has the same distribution as the pasted process  $\mathfrak{P}^\zeta \mathcal{X}$  on  $\mathfrak{PW}$  obtained within this setting.

Finally we construct the  $b$ -BMP by selecting the particles in a BMP *via* the renewal distribution. Notice that we add a time-coordinate<sup>13</sup>, in order to keep a time-homogeneous strong Markov process despite the fact that we kill particles below a generic curve  $b$ .

**Definition A.3.** Set  $E_2 = \left( \{\diamond\} \sqcup \bigsqcup_{n \geq 1} \mathbb{R}^n \right) \times \mathbb{R}_+$ , where  $\diamond$  is a point in which we put the process when all particle have gone below the curve. Let<sup>14</sup>  $\mathcal{Y} = (\mathcal{X}, \Theta)$  be a BMP  $(\mathcal{X}_t)_{t \geq 0}$  coupled with a deterministic time process  $(\Theta_t)_{t \geq 0}$  increasing with constant speed 1 (and always staying at  $\diamond$  if it starts from  $\diamond$ ). Let  $\zeta^\dagger$  be the first time when a particle goes below  $b$ , namely

$$\zeta^\dagger = \inf \{ s \geq 0, \exists i \in \llbracket 1, \#\mathcal{X}_{\Theta_s} \rrbracket, X_{\Theta_s}^i < b_{\Theta_s} \},$$

and  $\lambda$  the renewal distribution on  $E_2$  defined as follows: for  $y = (x, t) \in E_2$  with  $x \in \mathbb{R}$ , we set  $\bar{\lambda}(y, \cdot) = \delta_{(\diamond, t)}$ ; and for  $y = ((x^1, \dots, x^n), t) \in E_2$  with  $n \geq 2$ , let<sup>15</sup>  $m$  be s.t.  $x^m = \min_i x^i$ ; then we denote by  $\bar{\lambda}(y, \cdot)$  the Dirac mass on  $((x^1, \dots, x^{m-1}, x^{m+1}, \dots, x^n), t)$ . We also set, for  $y = (\diamond, t)$ ,  $\bar{\lambda}(y, \cdot) = \delta_{(\diamond, t)}$ , even though this will never be used in the pasting procedure because if the process starts with no particle, no particle ever goes below the curve (so that  $\zeta^\dagger = +\infty$ ).

We then define  $\lambda(w, \cdot) = \bar{\lambda}(\mathcal{Y}_\zeta(w), \cdot)$ . We call  $b$ -BMP any stochastic process that has the same distribution as the pasted process  $\mathfrak{P}^\zeta \mathcal{Y}$  on  $\mathfrak{PW}$  obtained within this setting. Notice that we kill indeed a particle which is below the curve, because by definition of  $\zeta$  and by right-continuity, its lowest particle has to be below the curve at time  $\zeta$ .

<sup>12</sup>If there are several indices  $m$  such that  $x^m = \min x^i$ , we arbitrarily chose  $m$  to be the smallest one.

<sup>13</sup>In the rest of the paper, we often refer to the space coordinate as  $b$ -BMP, even though strictly speaking, we should add the time coordinate.

<sup>14</sup>In the next section, we explain in a general setting why a pasted process can in turn be used as a base process.

<sup>15</sup>If there are several indices that realize the minimum, we chose the smallest one arbitrarily. Notice that this construction indeed kills instantaneously the particles that go below the curve. Although it might seem that, in the event that several particles go below the curve simultaneously, only one is killed at that time, when this happens, the new trajectory starts with particles below the curve already, and thus the corresponding  $\zeta$  time is zero; the renewal distribution is in fact used as many times as needed instantaneously, so that all particles below the boundary are killed.

**Remark A.4.**

- Notice that adding the time-coordinate preserves the fact that the process is Feller, and Assumptions (a) to (f) are still verified.
- The time  $\zeta^\dagger$  could a priori be infinite with positive probability in [Definition A.3](#), which we ruled out with Assumption (h) for the sake of simplicity. Since, in the main part of the article, we always consider the coupling up to a finite time horizon  $T > 0$ , we could always enforce the finiteness of  $\zeta^\dagger$  by replacing  $\zeta^1$  in Assumption (h) by  $\zeta^1 \wedge T$ , because we do not care about the distribution of the process after time  $T$ .

**A.1.3 Pasting of pasted processes**

In [Definition A.3](#) we use two levels of pasting: at the first level, we consider independent trajectories, and paste them to implement the branching (and obtain a BMP, as in [Definition A.1](#)); at the second level, we paste the resulting trajectories to implement the selection. One can wonder whether the distribution of the final process changes when pasting the other way round (starting with the selection procedure, and then adding the branching mechanism in a similar way as in [Definition A.1](#)).

We will show through [Proposition A.2](#) that these processes indeed share the same distribution, and that this is true in general when iterating the pasting procedure. Of course, we first have to explain carefully why (or rather how) we can iterate the procedure, meaning that we have to check the hypotheses (a) to (i) for the pasted process.

**A.1.3.1 Definition and notation for the pasting of pasted process**

Consider two stopping times  $\zeta$  and  $\theta$  on  $W$ , and two renewal distributions  $\lambda$  and  $\nu$ , both couples  $(\zeta, \lambda)$  and  $(\theta, \nu)$  satisfying Assumptions (g), (h) and (i). In order to lighten the notation, we will now use bold font to refer to the process defined in [Section A.1.1](#): we use the notation

$$(\mathbf{W}, \mathbf{W}, (\mathbf{W}_t)_{t \geq 0}, (\mathbf{P}_y)_{y \in \bar{E}}, (\mathbf{Y}_t)_{t \geq 0}, (\boldsymbol{\psi}_t)_{t \geq 0})$$

instead of

$$(\mathfrak{P}W, \mathfrak{P}W, (\mathfrak{P}W_t)_{t \geq 0}, (\mathfrak{P}^\zeta \mathbb{P}_y^\lambda)_{y \in \bar{E}}, (\mathfrak{P}^\zeta \mathcal{Y}_t)_{t \geq 0}, (\mathfrak{P}^\zeta \psi_t)_{t \geq 0}).$$

We now want to use this pasted process  $\mathcal{Y}$  as a new base process for a pasting procedure, with pasting times  $\theta$  and renewal distribution  $\nu$ . However, those are not defined on the same probability space as  $\mathcal{Y}$ . Thus we need to define a random time on  $\mathbf{W}$  which corresponds to  $\theta$ : define  $\tau = \zeta \wedge \theta$  and  $\Xi = \mathbf{1}_{\{\tau = \theta\}}$  on  $W$ , and define a new stopping time  $\boldsymbol{\theta}$  on  $\mathbf{W}$  by

$$\boldsymbol{\theta}(\mathbf{w}) = \begin{cases} \bar{\tau}_k(\mathbf{w}) & \text{if } \Xi(w_1) = \dots = \Xi(w_{k-1}) = 0, \Xi(w_k) = 1, \text{ for } k \geq 1, \\ \bar{\tau}_\infty(\mathbf{w}) & \text{if } \forall k \geq 1, \Xi(w_k) = 0, \end{cases} \quad (\text{A.7})$$

where  $\bar{\tau}_k(\mathbf{w}) = \sum_{l=1}^k \tau(w_l)$  (see [Figure A.2](#) for an example). We can now adapt the renewal distribution to extend it to  $\mathbf{W}$ . Since  $(\theta, \nu)$  satisfies Assumption (h), there exist Markov kernels  $\bar{\nu}_j : \bar{E} \times \mathcal{B}(\bar{E}) \rightarrow \mathbb{R}$ , almost surely distinct stopping times  $\theta^j$  and  $\theta^\dagger = \min_j \theta^j$  such that

$$\nu(w, \cdot) = \sum_{j=1}^m \mathbf{1}_{\{\theta^\dagger(w) = \theta^j(w)\}} \bar{\nu}_j(\mathcal{Y}_{\theta^j}(w), \cdot).$$

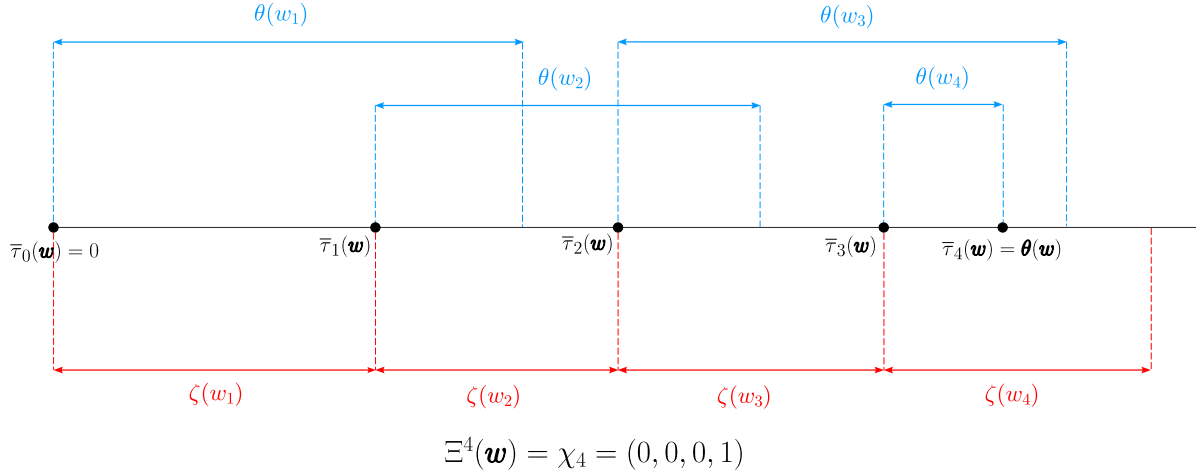


Figure A.2: A realization of the extension  $\theta$  of  $\theta$  on  $\mathbf{W}$

We now define  $\theta^\dagger$  through [\(A.7\)](#), replacing  $\theta$  by  $\theta^\dagger$  in the definitions of  $\Xi, \tau$  and the  $\bar{\tau}_k$ , and proceed similarly with  $\theta^j$  to define  $\theta^j$ . Then we consider the renewal distribution  $\bar{\nu}$  on  $\mathbf{W}$  defined by

$$\nu(\mathbf{w}, \cdot) = \sum_{j=1}^m \mathbf{1}_{\{\theta^\dagger(\mathbf{w}) = \theta^j(\mathbf{w})\}} \bar{\nu}_j(\mathcal{Y}_{\theta^j}(\mathbf{w}), \cdot).$$

We now check that the assumptions for the pasting procedure are verified. We have already seen (see [Proposition A.1](#)) that Assumptions (a) to (f) are valid; Assumption (g) is also verified for  $\theta^\dagger$ : first, the event  $\{\Xi^1 = 0\}$  is negligible with respect to  $\mathbf{P}_\Delta$  (it has probability  $\mathbb{P}_\Delta(\Xi = 0) = 0$  by Assumption (g), since  $\zeta = \theta = 0$  almost surely), so that we can integrate only on its complementary event on which  $\theta^\dagger(\mathbf{w}) = \theta^\dagger(w_1)$ , and obtain

$$\mathbf{P}_\Delta(\theta^\dagger = 0) = \mathbb{P}_\Delta(\theta^\dagger = 0) = 1.$$

Then we can check that the compatibility equation

$$\theta^\dagger(\psi_t \mathbf{w}) = \theta^\dagger(\mathbf{w}) - t$$

holds for  $t, \mathbf{w}$  such that  $\theta^\dagger(\mathbf{w}) \geq t$ , by performing the computation and using the definition (2.11) in [\[NW68\]](#) of the shift operators on  $\mathbf{W}$  and the corresponding compatibility equations for  $\theta^\dagger$  and  $\zeta^\dagger$ . Note that in fact, this compatibility equation is key to providing a meaningful definition of what should be the composition of the successive  $\theta$  random times on the space  $\mathbf{W}$ , and to use it as a new compatible time.

Finally, thanks to the definition of  $\nu$ , Assumptions (h) and (i) directly come from the same assumptions for  $\nu$ .

We can now define the pasting of the pasted process: first, we extend the notation of [Section A.1.1](#): we write  $\mathfrak{P}^{(2)}W := \mathfrak{P}[\mathfrak{P}W] = \mathfrak{P}W$  and denote by  $\mathfrak{P}\mathbf{w} = ((\mathbf{w}_1, y_1), (\mathbf{w}_2, y_2), \dots)$  its generic elements. Now the new pasted process is  $\mathfrak{P}^{\zeta, \theta} \mathcal{Y} := \mathfrak{P}^\theta \mathcal{Y}$ , thus given by

$$\mathfrak{P}^{\zeta, \theta} \mathcal{Y}_t(\mathfrak{P}\mathbf{w}) = \begin{cases} \mathcal{Y}_{t - \bar{\theta}_{i-1}(\mathfrak{P}\mathbf{w})}(w_i) & \text{if } \bar{\theta}_{i-1}(\mathfrak{P}\mathbf{w}) \leq t < \bar{\theta}_i(\mathfrak{P}\mathbf{w}), i \geq 1, \\ \Delta & \text{otherwise,} \end{cases} \quad (\text{A.8})$$

with  $\bar{\theta}_i(\mathfrak{P}\mathbf{w}) = \sum_{k=1}^i \theta(w_k)$ . The family of probability measures that we put on  $\mathfrak{P}^{(2)}W$  is the pasting of the measures  $\mathbf{P}_y$  according to  $\nu$ , namely the probability measures  $\mathfrak{P}^{\zeta, \theta} \mathbb{P}_y^{\lambda, \nu} := \mathfrak{P}^\theta \mathbf{P}_y^\nu$

that are the concatenation on  $\mathfrak{P}^{(2)}W = \prod_{i=1}^{+\infty} W \times \bar{E}$  of the measures  $\mathbf{Q}_y^\nu$  on  $W \times \bar{E}$  defined by extending the following formula, valid for  $A$  in  $\mathcal{W}$  and  $B$  a borel set in  $\bar{E}$ :

$$\mathbf{Q}_y^\nu(A \times B) = \int_A \nu(w, B) \mathbf{P}_y(dw).$$

In the end, if  $Z_i$  is a bounded real valued random variable on  $\Omega^i$ , we can extend it to a random variable  $Z$  on  $\mathfrak{P}^{(2)}W$  depending only on the first  $i$  coordinates  $\omega_1, \dots, \omega_i$ , with  $\omega_k = (w_k, y_k) \in \Omega := W \times \bar{E}$ , and we have, denoting  $\mathfrak{P}^{\zeta, \theta} \mathbb{E}_y^{\lambda, \nu}$  the expectation under  $\mathfrak{P}^{\zeta, \theta} \mathbb{P}_y^{\lambda, \nu}$ ,

$$\mathfrak{P}^{\zeta, \theta} \mathbb{E}_y^{\lambda, \nu}(Z) = \int_{\Omega^i} Z_i(\omega_1, \dots, \omega_i) \prod_{k=1}^i \mathbf{Q}_{y_{k-1}}^\nu(d\omega_k),$$

using the convention  $y_0 = y$ .

### A.1.3.2 Commutativity of the construction

Now we have constructed a process  $\mathfrak{P}^{\zeta, \theta} \mathcal{Y}$  by pasting first according to  $\zeta$  and  $\lambda$ , and then according to  $\theta$  and  $\nu$ . The same procedure allows us to consider the process  $\mathfrak{P}^{\theta, \zeta} \mathcal{Y}$  (with associated probability measures on the same probability space  $\mathfrak{P}^{(2)}W$  denoted  $\mathfrak{P}^{\theta, \zeta} \mathbb{P}_y^{\nu, \lambda}$ ) constructed by pasting first according to  $\theta$  and  $\nu$  and then according to  $\zeta$  and  $\lambda$ . The natural question that arises is whether those processes have the same distribution.

We will compare these two processes with the process obtained by pasting copies of  $\mathcal{Y}$  with pasting time  $\tau = \zeta \wedge \theta$  and renewal distribution

$$\xi(w, \cdot) := M^{\zeta, \theta}(\lambda, \nu)(w, \cdot) = \mathbf{1}_{\{\Xi(w)=0\}} \lambda(w, \cdot) + \mathbf{1}_{\{\Xi(w)=1\}} \nu(w, \cdot),$$

which defines again an instantaneous distribution satisfying (h) and (i). Then we can consider the corresponding probability measures  $\mathfrak{P}^\tau \mathbb{P}_y^\xi$  on  $\mathfrak{P}W$  and the pasted process  $\mathfrak{P}^\tau \mathcal{Y}$ .

**Proposition A.2.** *The processes  $\mathfrak{P}^{\zeta, \theta} \mathcal{Y}$ ,  $\mathfrak{P}^{\theta, \zeta} \mathcal{Y}$  and  $\mathfrak{P}^\tau \mathcal{Y}$  induce the same distribution on the space  $\mathcal{D}(\mathbb{R}_+, E)$  (with respective reference probability measures  $\mathfrak{P}^{\zeta, \theta} \mathbb{P}_y^{\lambda, \nu}$ ,  $\mathfrak{P}^{\theta, \zeta} \mathbb{P}_y^{\nu, \lambda}$  and  $\mathfrak{P}^\tau \mathbb{P}_y^\xi$ ).*

**Example A.1.** *As we mentioned, the construction of the b-BMP (Definition A.3) can be reversed: we can start from independent copies of  $X$ , then paste them to carry out the selection at the boundary first, and then use the exponential times to paste again and carry out the branching.*

*Proof.* We give the proof in the case where  $\mathcal{Y}$  has an infinite lifetime. Recall in addition that  $\theta$  is finite almost surely.

Our setting being symmetric with respect to  $(\zeta, \lambda)$  and  $(\theta, \nu)$ , we only need to show that  $\mathfrak{P}^{\zeta, \theta} \mathcal{Y}$  and  $\mathfrak{P}^{\zeta \wedge \theta} \mathcal{Y}$  have the same distribution. Recall that these are time-homogeneous Markov processes, so that it is enough to prove that the random variables  $\mathfrak{P}^{\zeta, \theta} \mathcal{Y}_t$  and  $\mathfrak{P}^{\zeta \wedge \theta} \mathcal{Y}_t$  with values in  $\bar{E}$  have the same distribution, for all time  $t \geq 0$  and all starting point  $y \in \bar{E}$ . We will thus prove that for any  $t \geq 0$ ,  $y \in E$  and  $f \in \mathcal{C}_0(\bar{E})$ , we have (remember that  $\tau = \zeta \wedge \theta$ )

$$\mathfrak{P}^{\zeta, \theta} \mathbb{E}_y^{\lambda, \nu}(f(\mathfrak{P}^{\zeta, \theta} \mathcal{Y}_t)) = \mathfrak{P}^\tau \mathbb{E}_y^\xi(f(\mathfrak{P}^\tau \mathcal{Y}_t)). \quad (\text{A.9})$$

Since  $f(\Delta) = 0$ , the definition (A.6) of  $\mathfrak{P}^\tau \mathcal{Y}$  ensures that  $f(\mathfrak{P}^\tau \mathcal{Y}_t)$  is zero for  $t \geq \sum_{k=1}^{+\infty} \tau(w_k)$ , so we can write

$$f(\mathfrak{P}^\tau \mathcal{Y}_t(\mathfrak{P}w)) = \sum_{i=1}^{+\infty} \mathbf{1}_{[\bar{\tau}_{i-1}, \bar{\tau}_i)}(\mathfrak{P}w) f(\mathcal{Y}_{t-\bar{\tau}_{i-1}}(\mathfrak{P}w)(w_i)).$$

Step 1. Decomposing both expectations depending on the successive values of  $\Xi$ .

Recall that  $\Xi$  is the random variable on  $W$  defined by

$$\Xi = \mathbf{1}_{\{\tau=\theta\}} = \mathbf{1}_{\{\theta \leq \zeta\}}.$$

First, we write the r.h.s. in (A.9) with sums depending on the number of trajectories before  $t$  for which the random variable  $\Xi$  is equal to 1, and on the number of trajectories for which the random variable  $\Xi$  is equal to 0 between each of them: for  $k, k_1, \dots, k_l \geq 1$  and  $1 \leq r_1 < \dots < r_i$ , write  $\Xi^k(w_1, \dots, w_k) := (\Xi(w_1), \dots, \Xi(w_k))$ ,  $\bar{k}_l := k_1 + \dots + k_l$  and  $\chi_{r_1, \dots, r_i} := (\mathbf{1}_{\{r_1, \dots, r_i\}}(j))_{1 \leq j \leq r_i}$ , so that the event  $\{\Xi^{r_i} = \chi_{r_1, \dots, r_i}\}$  is the event where for all  $1 \leq j \leq i$ , the  $j$ -th ' $\theta$ -type' renewal time is the  $r_j$ -th renewal time  $\bar{\tau}_{r_j}$ . Then we have

$$\mathfrak{P}^\tau \mathbb{E}_y^\xi (f(\mathfrak{P}^\tau \mathcal{Y}_t)) = \sum_{i=1}^{+\infty} \sum_{k_1, \dots, k_i=1}^{+\infty} \mathfrak{P}^\tau \mathbb{E}_y^\xi \left( \mathbf{1}_{\{\Xi^{\bar{k}_i} = \chi_{\bar{k}_1, \dots, \bar{k}_i}\}} \mathbf{1}_{[\bar{\tau}_{\bar{k}_i-1}, \bar{\tau}_{\bar{k}_i})}(t) f(\mathfrak{P}^\tau \mathcal{Y}_t) \right), \quad (\text{A.10})$$

where the random variable in the expectation depends only on the first  $\bar{k}_i$  coordinates of  $\mathfrak{P}w$ , denoted by  $w^{\bar{k}_i} = (w_1, \dots, w_{\bar{k}_i})$ .

The rest of the proof consists in showing that the l.h.s. in (A.9) has the same expression. To this end, we use the lighter notation of Section A.1.3.1. From the definition of  $\mathfrak{P}^\theta \mathbf{P}_y^\nu$ , we obtain

$$\mathfrak{P}^\theta \mathbf{E}_y^\nu (f(\mathfrak{P}^\theta \mathcal{Y}_t)) = \sum_{i=1}^{+\infty} \int_{\Omega^i} \mathbf{1}_{[\bar{\theta}_{i-1}(w^{i-1}), \bar{\theta}_i(w^i))}(t) f(\mathcal{Y}_{t-\bar{\theta}_{i-1}(w^{i-1})}(w^i)) \prod_{l=1}^i \mathbf{Q}_{y_{l-1}}^\nu(d\omega_l), \quad (\text{A.11})$$

where  $w^i = (w_1, \dots, w_i) \in \mathbf{W}^i$ .

Defining for  $i \geq 1$  and  $k^i = (k_1, \dots, k_i)$

$$\tilde{\tau}_{k^i}(w^i) = \sum_{l=1}^i \bar{\tau}_{k_l}(w_l),$$

as well as

$$\mathcal{I}_{k^i} = \int_{\Omega^i} \mathbf{1}_{[\tilde{\tau}_{k^{i-1}}(w^{i-1}), \tilde{\tau}_{k^i}(w^i))}(t) f(\mathcal{Y}_{t-\tilde{\tau}_{k^{i-1}}(w^{i-1})}(w^i)) \prod_{l=1}^i \mathbf{1}_{\{\Xi^{k_l}(w_l) = \chi_{k_l}\}} \mathbf{Q}_{y_{l-1}}^\nu(d\omega_l),$$

still with convention  $y_0 = y$ , we claim that we have

$$\mathfrak{P}^\theta \mathbf{E}_y^\nu (f(\mathfrak{P}^\theta \mathcal{Y}_t)) = \sum_{i=1}^{+\infty} \sum_{k_1, \dots, k_i=1}^{+\infty} \mathcal{I}_{k^i}. \quad (\text{A.12})$$

Indeed, for each  $\omega_l = (w_l, y_l) \in \mathbf{W} \times \bar{E}$ , we rewrite  $\theta(w_l)$  depending on the number of  $\Xi$  random variables equal to 0 before the first 1. Recall the assumption that we made at the start of the proof:  $\mathcal{Y}$  has an infinite lifetime and  $\theta$  is finite a.s. Thus we have almost surely

$$\sum_{k=1}^{+\infty} \mathbf{1}_{\{\Xi^k = \chi_k\}} = 1, \quad (\text{A.13})$$

and

$$\sum_{k_1, \dots, k_i=1}^{+\infty} \mathbf{1}_{\{\Xi^{\bar{k}_i} = \chi_{\bar{k}_i}\}} = 1. \quad (\text{A.14})$$

Now, on the event  $\{\Xi^{\bar{k}_i} = \chi_{k_i}\}$ , we have for all  $l \leq i$

$$\bar{\theta}(\mathbf{w}^l) = \tilde{\tau}_{k^l}(\mathbf{w}^l) \quad (\text{A.15})$$

Hence we can split the integral on  $\Omega^i$  in (A.11) using (A.14), and then use (A.15) to obtain (A.12) as claimed.

We now temporarily fix  $i \geq 1$  and  $k_1, \dots, k_i \geq 1$  and define recursively for  $r = i, i-1, \dots, 0$  the random variables  $Z_r$  on  $\Omega^r$ , starting with  $Z_i$ , defined for  $\omega^i = (\omega_1, \dots, \omega_i) \in \Omega^i$  by

$$Z_i(\omega^i) = \mathbf{1}_{\{\Xi^{k_i}(\omega_i) = \chi_{k_i}\}} \mathbf{1}_{[\tilde{\tau}_{k^{i-1}}(\omega^{i-1}), \tilde{\tau}_{k^i}(\omega^i)]}(t) f(\mathcal{Y}_{t-\tilde{\tau}_{k^{i-1}}(\omega^{i-1})}(\omega_i)), \quad (\text{A.16})$$

then setting for  $i-1 \geq r \geq 1$

$$Z_r(\omega^r) = \mathbf{1}_{\{\Xi^{k_r}(\omega_r) = \chi_{k_r}\}} \int_{\Omega} Z_{r+1}(\omega^{r+1}) \mathbf{Q}_{y_r}^{\nu}(d\omega_{r+1}), \quad (\text{A.17})$$

and finally

$$Z_0 = \int_{\Omega} Z_1(\omega_1) \mathbf{Q}_y^{\nu}(d\omega_1). \quad (\text{A.18})$$

We have by definition, for all  $1 \leq r \leq i$ , that

$$\mathcal{I}_{k^i} = \int_{\Omega^r} Z_r(\omega^r) \prod_{l=1}^{r-1} \mathbf{1}_{\{\Xi^{k_l}(\omega_l) = \chi_{k_l}\}} \mathbf{Q}_{y_{l-1}}^{\nu}(d\omega_l), \quad (\text{A.19})$$

and, for  $r = 0$ ,

$$\mathcal{I}_{k^i} = Z_0. \quad (\text{A.20})$$

*Step 2. Rewriting  $Z_{r-1}$  for  $1 \leq r \leq i$ .*

We will show the following by induction: for  $1 \leq r \leq i$ ,  $Z_r$  depends on  $\omega_r = (\omega_r, y_r) \in \Omega$  only through  $w_r^{k_r} = (w_{r,1}, \dots, w_{r,k_r})$  and  $y_r$ , and we have

$$Z_{r-1}(\omega^{r-1}) = \mathbf{1}_{\{\Xi^{k_{r-1}}(\omega_{r-1}) = \chi_{k_{r-1}}\}} \int_{\Omega^{k_r}} Z_r(\omega^{r-1}, w_r^{k_r}, y_{r,k_r}) \prod_{l=1}^{k_r} \mathbf{Q}_{y_{r,l-1}}^{\xi}(d\omega_{r,l}). \quad (\text{A.21})$$

with the convention  $y_{r,0} = y_{r-1}$  and  $y_0 = y$ . To do so, we use the following lemma repeatedly.

**Lemma A.1.** *Let  $k \geq 1$  and  $Z : W^k \times \bar{E} \rightarrow \mathbb{R}$  be a bounded random variable, which can be extended to a random variable  $\tilde{Z}$  on  $\Omega = \mathbf{W} \times \bar{E}$  though the formula*

$$\tilde{Z}(\omega, y') = Z(w^k, y')$$

*Then we have*

$$\mathbf{E}_y^{\nu}(\tilde{Z} \mathbf{1}_{\{\Xi^k = \chi_k\}}) = \int_{\Omega^k} Z(w^k, y_k) \mathbf{1}_{\{\Xi^k(w^k) = \chi_k\}} \prod_{l=1}^k \mathbf{Q}_{y_{l-1}}^{\xi}(d\omega_l). \quad (\text{A.22})$$

*Proof.* Thanks to the definition of  $\mathbf{Q}_y^{\nu}$ , we have

$$\mathbf{E}_y^{\nu}(\tilde{Z} \mathbf{1}_{\{\Xi^k = \chi_k\}}) = \int_{\Omega} Z(w^k, y') \mathbf{1}_{\{\Xi^k(w^k) = \chi_k\}} \nu(\mathbf{w}, dy') \mathbf{P}_y(d\mathbf{w}),$$

and now thanks to the definition of  $\nu$ , we obtain

$$\mathbf{E}_y^{\nu}(\tilde{Z} \mathbf{1}_{\{\Xi^k = \chi_k\}}) = \int_{\Omega} Z(w^k, y') \mathbf{1}_{\{\Xi^k(w^k) = \chi_k\}} \sum_{j=1}^m \prod_{j' \neq j} \mathbf{1}_{\{\theta^{j'}(\mathbf{w}) \geq \theta^j(\mathbf{w})\}} \bar{\nu}^j(\mathcal{Y}_{\theta^j(\mathbf{w})}(\mathbf{w}), dy') \mathbf{P}_y(d\mathbf{w}).$$



Now for  $\mathbf{w} \in \{\Xi^k = \chi_k\}$ , we have for all  $1 \leq j \leq m$  by definition

$$\theta^j(\mathbf{w}) = \bar{\zeta}_{k-1}(w^{k-1}) + \theta^j(w_k),$$

so that

$$\mathbf{1}_{\{\theta^{j'}(\mathbf{w}) \geq \theta^j(\mathbf{w})\}} = \mathbf{1}_{\{\theta^{j'}(w_k) \geq \theta^j(w_k)\}}$$

and

$$\mathcal{Y}_{\theta^j}(\mathbf{w}) = \mathcal{Y}_{\theta^j}(w_k).$$

Thus we have

$$\mathbf{E}_y^\nu(\tilde{Z}\mathbf{1}_{\{\Xi^k = \chi_k\}}) = \int_{\Omega} Z(w^k, y') \mathbf{1}_{\{\Xi^k(w^k) = \chi_k\}} \sum_{j=1}^m \prod_{j' \neq j} \mathbf{1}_{\{\theta^{j'}(w_k) \geq \theta^j(w_k)\}} \bar{\nu}(\mathcal{Y}_{\theta^j}(w_k), dy') \mathbf{P}_y(d\mathbf{w}),$$

and we recognize  $\nu(w_k, \cdot)$ , so that we have

$$\mathbf{E}_y^\nu(\tilde{Z}\mathbf{1}_{\{\Xi^k = \chi_k\}}) = \int_{\Omega} Z(w^k, y') \mathbf{1}_{\{\Xi^k(w^k) = \chi_k\}} \nu(w_k, dy') \mathbf{P}_y(dw).$$

Now the integrand depends on  $\mathbf{w}$  only through  $w^k$ , so we have by definition of  $\mathbf{P}_y$

$$\mathbf{E}_y^\nu(\tilde{Z}\mathbf{1}_{\{\Xi^k = \chi_k\}}) = \int_{\Omega^k \times \bar{E}} Z(w^k, y') \mathbf{1}_{\{\Xi^k(w^k) = \chi_k\}} \nu(w_k, dy') \prod_{l=1}^k \mathbb{Q}_{y_{l-1}}^\lambda(dw_l),$$

with the convention  $y_0 = y$ . Recalling the definition of  $\mathbb{Q}_y^\lambda$ , we have

$$\mathbf{E}_y^\nu(\tilde{Z}\mathbf{1}_{\{\Xi^k = \chi_k\}}) = \int_{\Omega^k \times \bar{E}} Z(w^k, y') \mathbf{1}_{\{\Xi^k(w^k) = \chi_k\}} \nu(w_k, dy') \prod_{l=1}^k \lambda(w_l, dy_l) \mathbb{P}_{y_{l-1}}(dw_l),$$

and the integrand does not depend on  $y_k$ . Thus we can delete the integral on  $y_k$  in  $\bar{E}$  and obtain

$$\mathbf{E}_y^\nu(\tilde{Z}\mathbf{1}_{\{\Xi^k = \chi_k\}}) = \int_{\Omega^{k-1} \times W \times \bar{E}} Z(w^k, y') \mathbf{1}_{\{\Xi^k(w^k) = \chi_k\}} \nu(w_k, dy') \mathbb{P}_{y_{k-1}}(dw_k) \prod_{l=1}^{k-1} \lambda(w_l, dy_l) \mathbb{P}_{y_{l-1}}(dw_l).$$

Now we rename  $y_k$  the variable  $y'$ , thus obtaining

$$\mathbf{E}_y^\nu(\tilde{Z}\mathbf{1}_{\{\Xi^k = \chi_k\}}) = \int_{\Omega^k} Z(w^k, y_k) \mathbf{1}_{\{\Xi^k(w^k) = \chi_k\}} \nu(w_k, dy_k) \mathbb{P}_{y_{k-1}}(dw_k) \prod_{l=1}^{k-1} \lambda(w_l, dy_l) \mathbb{P}_{y_{l-1}}(dw_l).$$

Now, since  $\lambda(w, \cdot)$  (resp.  $\nu(w, \cdot)$ ) and  $\xi(w, \cdot)$  agree on  $\{\Xi = 0\}$  (resp.  $\{\Xi = 1\}$ ), we can replace both  $\lambda$  and  $\nu$  by  $\xi$  in the previous equation without changing the value, and obtain

$$\mathbf{E}_y^\nu(\tilde{Z}\mathbf{1}_{\{\Xi^k = \chi_k\}}) = \int_{\Omega^k} Z(w^k, y_k) \mathbf{1}_{\{\Xi^k(w^k) = \chi_k\}} \prod_{l=1}^k \xi(w_l, dy_l) \mathbb{P}_{y_{l-1}}(dw_l),$$

which gives (A.22) by definition of  $\mathbb{Q}_y^\xi$ . □

Now let us start with the case  $r = i$ : looking at the definition (A.16) of  $Z_i$ , its only dependence in  $\omega_i$  is through  $\tilde{\tau}_{k^i}$ , so through  $w_i^{k^i}$ , and through  $\mathcal{Y}_{t-\tilde{\tau}_{k^i-1}}(w^{i-1})(w_i)$ , which on the event

$\{t - \tilde{\tau}_{k^{i-1}}(\boldsymbol{w}^{i-1}) \in [0, \bar{\tau}_{k^i}]\}$  also depends only on  $w_i^{k^i}$ . Thus we can apply [Lemma A.1](#) for fixed  $\boldsymbol{w}^{i-1}$  to  $Z_i(\boldsymbol{w}^{i-1}, \cdot)$ , seen as a random variable on  $W^k \times \bar{E}$ , with  $k = k_i$  to obtain<sup>16</sup>

$$\mathbf{E}_y^\nu(Z_i(\boldsymbol{w}^{i-1}, \cdot) \mathbf{1}_{\{\Xi^{k_i} = \chi_{k_i}\}}) = \int_{\Omega^{k_i}} Z_i(\boldsymbol{w}^{i-1}, w_i^{k_i}, y_{i,k_i}) \mathbf{1}_{\{\Xi^{k_i}(w_i^{k_i}) = \chi_{k_i}\}} \prod_{l=1}^{k_i} \mathbb{Q}_{y_{i,l-1}}^\xi(dw_{i,l}). \quad (\text{A.23})$$

Thanks to the recursive definition [\(A.17\)](#) of  $Z_r$ , the l.h.s. of [\(A.23\)](#) equals  $Z_{i-1}$ , so that we have exactly obtained [\(A.21\)](#) for  $r = i$ .

Now fix  $m$  s.t.  $0 \leq m \leq i - 1$  and assume that [\(A.21\)](#) holds for every  $r$  s.t.  $m + 1 \leq r \leq i$ . Now assume that  $r = m$ . Using these equalities one after another, we have that

$$\begin{aligned} Z_r(\boldsymbol{\omega}^r) &= \mathbf{1}_{\{\Xi^{k_r}(\boldsymbol{\omega}_r) = \chi_{k_r}\}} \int_{\Omega^{k_{r+1}}} Z_{r+1}(\boldsymbol{\omega}^r, \omega_{r+1}^{k_{r+1}}) \prod_{l=1}^{k_{r+1}} \mathbb{Q}_{y_{r+1,l-1}}^\xi(d\omega_{r+1,l}). \\ &= \dots \\ &= \mathbf{1}_{\{\Xi^{k_r}(\boldsymbol{\omega}_r) = \chi_{k_r}\}} \int_{\Omega^{k_{r+1} + \dots + k_i}} Z_i(\boldsymbol{\omega}^r, \omega_{r+1}^{k_{r+1}}, \dots, \omega_i^{k_i}) \\ &\quad \prod_{p=r+1}^i \mathbf{1}_{\{\Xi^{k_p}(w_p^{k_p}) = \chi_{k_p}\}} \prod_{l_p=1}^{k_p} \mathbb{Q}_{y_{p,l_p-1}}^\xi(d\omega_{p,l_p}) \end{aligned}$$

with  $y_{p,0} = y_{p-1,k_{p-1}}$  for  $p \geq r + 2$  and  $y_{r+1,0} = y_r$ . Now, looking at the definition [\(A.16\)](#) of  $Z_i$ , its only dependence in  $\boldsymbol{\omega}_r$  is through  $\tilde{\tau}_{k^{i-1}}$  and  $\tilde{\tau}_{k^i}$ , so through  $\bar{\tau}_{k_r}(\boldsymbol{\omega}_r)$ , which in turn depends on  $\boldsymbol{\omega}_r$  only through  $w_r^{k_r}$ . Thus the last integral we obtained depends on  $\boldsymbol{\omega}_r$  only through  $w_r^{k_r}$  and  $y_r$ , which allows us to apply [Lemma A.1](#) once again, to the random variable  $Z_r(\boldsymbol{\omega}^{r-1}, \cdot)$  on  $\Omega$ , with  $k = k_r$ . Just like in the initial case, [\(A.22\)](#) gives exactly [\(A.21\)](#) in this context.

*Step 3. Computing  $\mathcal{I}_{k^i}$ .*

Applying [\(A.21\)](#) repeatedly now allows us to compute  $\mathcal{I}_{k^i}$ : we have

$$\begin{aligned} \mathcal{I}_{k^i} &= Z_0 \\ &= \int_{\Omega^{k_1}} \mathbf{1}_{\{\Xi^{k_1}(w_1^{k_1}) = \chi_{k_1}\}} Z_1(w_1^{k_1}, y_{1,k_1}) \prod_{l=1}^{k_1} \mathbb{Q}_{y_{1,l-1}}^\xi(dw_{1,l}) \\ &= \int_{\Omega^{k_1}} \mathbf{1}_{\{\Xi^{k_1}(w_1^{k_1}) = \chi_{k_1}\}} \int_{\Omega^{k_2}} \mathbf{1}_{\{\Xi^{k_2}(w_2^{k_2}) = \chi_{k_2}\}} \\ &\quad Z_2(w_1^{k_1}, y_{1,k_1}, w_2^{k_2}, y_{2,k_2}) \prod_{l=1}^{k_2} \mathbb{Q}_{y_{2,l-1}}^\xi(dw_{2,l}) \prod_{l=1}^{k_1} \mathbb{Q}_{y_{1,l-1}}^\xi(dw_{1,l}) \\ &= \dots \\ \mathcal{I}_{k^i} &= \int_{\Omega^{k_1}} \mathbf{1}_{\{\Xi^{k_1}(w_1^{k_1}) = \chi_{k_1}\}} \int_{\Omega^{k_2}} \mathbf{1}_{\{\Xi^{k_2}(w_2^{k_2}) = \chi_{k_2}\}} \dots \int_{\Omega^{k_i}} \mathbf{1}_{\{\Xi^{k_i}(w_i^{k_i}) = \chi_{k_i}\}} \\ &\quad Z_i((w_r^{k_r}, y_{r,k_r})_{1 \leq r \leq i}) \prod_{r=1}^i \prod_{l_r=1}^{k_r} \mathbb{Q}_{y_{r,l_r-1}}^\xi(dw_{r,l_r}) \end{aligned}$$

with

$$\mathbf{1}_{\{\Xi^{k_i}(w_i^{k_i}) = \chi_{k_i}\}} Z_i((w_r^{k_r}, y_{r,k_r})_{1 \leq r \leq i}) = \sum_{j=1}^{k_i} \mathbf{1}_{[T_{j-1}(\cdot), T_j(\cdot)]}(t) f(\mathcal{Y}_{t-T_{j-1}(\cdot)}(w_i^j))$$

and

$$T_j(\cdot) = \sum_{r=1}^{i-1} \bar{\tau}_{k_r}(w_r^{k_r}) + \bar{\tau}_j(w_i^j).$$

<sup>16</sup>Here, we abuse of notation by still writing  $Z_i$  for  $\tilde{Z}_i$ .

Now we can rename  $\omega_{r,l}$  into  $\omega_{\bar{k}_{r-1}+l}$  and obtain

$$\mathcal{I}_{k^i} = \int_{\Omega^{\bar{k}_i}} \mathbf{1}_{\{\Xi^{\bar{k}_i}(w^{\bar{k}_i}) = \chi_{k_1, \dots, k_i}\}} \sum_{j=\bar{k}_{i-1}+1}^{\bar{k}_i} \mathbf{1}_{[\bar{\tau}_{j-1}(w^{j-1}), \bar{\tau}_j(w^j)]}(t) f(\mathcal{Y}_{t-\bar{\tau}_{j-1}(w^{j-1})}(w_j)) \prod_{l=1}^{\bar{k}_i} \mathbb{Q}_{y_{l-1}}^\xi(d\omega_l).$$

Now for  $w \in \mathbf{W}$  such that  $t \in [\bar{\tau}_{j-1}(w^{j-1}), \bar{\tau}_j(w^j)]$ ,  $\mathcal{Y}_{t-\bar{\tau}_{j-1}(w^{j-1})}(w_j)$  equals  $\mathfrak{P}^\tau \mathcal{Y}_t(w)$ , so that we recognize

$$\mathcal{I}_{k^i} = \mathfrak{P}^\tau \mathbb{E}_y^\xi \left( \mathbf{1}_{\{\Xi^{\bar{k}_i} = \chi_{\bar{k}_1, \dots, \bar{k}_i}\}} \mathbf{1}_{[\bar{\tau}_{\bar{k}_{i-1}}, \bar{\tau}_{\bar{k}_i}]}(t) f(\mathfrak{P}^\tau \mathcal{Y}_t) \right).$$

Now, looking back at [\(A.12\)](#), we have

$$\mathfrak{P}^\theta \mathbf{E}_y^\nu(f(\mathfrak{P}^\theta \mathcal{Y}_t)) = \sum_{i=1}^{+\infty} \sum_{k_1, \dots, k_i=1}^{+\infty} \mathfrak{P}^\tau \mathbb{E}_y^\xi \left( \mathbf{1}_{\{\Xi^{\bar{k}_i} = \chi_{\bar{k}_1, \dots, \bar{k}_i}\}} \mathbf{1}_{[\bar{\tau}_{\bar{k}_{i-1}}, \bar{\tau}_{\bar{k}_i}]}(t) f(\mathfrak{P}^\tau \mathcal{Y}_t) \right),$$

which finally gives thanks to [\(A.10\)](#)

$$\mathfrak{P}^\theta \mathbf{E}_y^\nu(f(\mathfrak{P}^\theta \mathcal{Y}_t)) = \mathfrak{P}^\tau \mathbb{E}_y^\xi(f(\mathfrak{P}^\tau \mathcal{Y}_t)).$$

□

#### A.1.4 Distribution of the coordinate processes

We now take a look at the pasting procedure in the case where the base process  $\mathcal{Y}$  is itself a coupling of several Markov processes.

To that end, consider a finite number<sup>17</sup>  $E_1, \dots, E_n$  of locally compact Hausdorff spaces with countable base and set  $E = \prod_{j=1}^n E_j$ . Add a cemetery point  $\Delta$  to  $E$ , write  $\pi_j : E \rightarrow E_j$  for

the canonical projection and extend it to  $\bar{E}$  with  $\pi_j(\Delta) = \Delta_j$ , with  $\bar{E}_j = E_j \cup \{\Delta_j\}$ . Let  $\mathcal{Y} = (W, \mathcal{W}, (\mathcal{W}_t)_{t \geq 0}, (\mathbb{P}_y)_{y \in \bar{E}}, (\mathcal{Y}_t)_{t \geq 0}, (\psi_t)_{t \geq 0})$  be a strong Markov process on  $E$  such that for each  $1 \leq j \leq n$ ,  $Y^j := \pi_j(\mathcal{Y})$  is also a strong Markov process satisfying Assumptions (a) to (f) with respect to the same filtration (we will refer to them as coordinate processes). Since  $Y^j$  is a Markov process, it has a semigroup  $(P_t^j)_{t \geq 0}$ , and the fact that  $\mathcal{Y}$  is a coupling of these Markov processes means that its semigroup, denoted by  $(\mathcal{P}_t)_{t \geq 0}$  verifies the following: for all bounded measurable function  $f^j : E_j \rightarrow \mathbb{R}$  and  $y = (y^1, \dots, y^n) \in E$ , we have, writing  $\hat{f}^j = f^j \circ \pi_j$ ,

$$P_t^j f^j(y^j) = \mathcal{P}_t \hat{f}^j(y). \quad (\text{A.24})$$

We will make the following assumption on the renewal definition  $\lambda$ :

$$\forall y \in E \quad \mathbb{P}_y(\{w \in W, \lambda(w, \{y \in E, \pi_j(\mathcal{Y}_\zeta(w)) = \pi_j(y)\}) = 1\}) = 1, \quad (\text{A.25})$$

which means that the  $j$ -th coordinate is unaffected by the pasting procedure. We prove that if [\(A.25\)](#) is true, thanks to the Markovian nature of the process  $\mathcal{Y}$ , the  $j$ -th coordinate of the pasted process, which we denote by  $\mathfrak{P}^\zeta Y^j$ , has the same distribution as the  $j$ -th coordinate of the base process  $Y^j$ .

<sup>17</sup>We are interested in taking  $n = 3$ , see [Section A.2](#).

**Proposition A.3.** Assume that  $\mathfrak{P}^\zeta \mathcal{Y}$  has almost surely an infinite lifetime (which means that  $\mathfrak{P}^\zeta \mathbb{P}_y^\lambda(\mathfrak{T} = +\infty) = 1$  except when  $y = \Delta$ ) and that for a certain  $j \in \llbracket 1, n \rrbracket$ , [\[A.25\]](#) holds. Then, writing  $\mathfrak{P}^\zeta Y_t^j := \pi_j(\mathfrak{P}^\zeta \mathcal{Y}_t)$ , we have that  $(\mathfrak{P}W, (\mathfrak{P}\mathcal{W}_t)_{t \geq 0}, (\mathfrak{P}^\zeta \mathbb{P}_y^\lambda)_{y \in E}, (\mathfrak{P}^\zeta Y_t^j)_{t \geq 0}, (\mathfrak{P}^\zeta \psi_t)_{t \geq 0})$  is a strong Markov process<sup>18</sup> with semigroup  $P^j$ .

We first prove the following lemma.

**Lemma A.2.** For all  $y$  in  $\overline{E}$ ,  $s \geq 0$ , and  $f \in \mathcal{C}_0(\overline{E})$ , we have

$$\mathbb{E}_y(\mathbf{1}_{\{s \geq \zeta\}} f(\mathcal{Y}_s)) = \mathbb{E}_y(\mathbf{1}_{\{s \geq \zeta\}} \mathcal{P}_{s-\zeta} f(\mathcal{Y}_\zeta)). \quad (\text{A.26})$$

*Proof.* Note that if  $y = \Delta$ , [\[A.26\]](#) holds since both sides are zero. Let us now fix  $y \in E$ . Define for  $w, w' \in W$

$$\phi(w, w') = \mathbf{1}_{\{s \geq \zeta(w)\}} f(\mathcal{Y}_{s-\zeta(w)}(w')) \quad (\text{A.27})$$

as well as  $Z(w) = \phi(w, \psi_\zeta w)$ . We start by making a sequence of discrete approximations of  $\zeta$ : for integers  $k, n \geq 0$ , write  $s_k^n = \frac{ks}{2^n}$ , and define

$$T^n(w) := \frac{s}{2^n} \left\lfloor \frac{2^n}{s} \zeta(w) \right\rfloor,$$

so that  $T^n$  has values in  $\{s_k^n, k \geq 0\}$  and converges to  $\zeta$  from below as  $n$  goes to infinity. Hence for  $w \in W$  such that  $s \geq \zeta(w)$ ,  $s - T^n$  converges to  $s - \zeta$  from above, and using the fact that  $\mathcal{Y}$  has right-continuous paths, we obtain that  $f(\mathcal{Y}_{s-T^n(w)}(w'))$  converges, for such  $w$  and any  $w'$ , to  $f(\mathcal{Y}_{s-\zeta(w)}(w'))$ . Since  $\mathbf{1}_{\{s \geq T^n(w)\}}$  also converges to  $\mathbf{1}_{\{s \geq \zeta(w)\}}$ , we have that  $\phi^n(w, w') := \mathbf{1}_{\{s \geq T^n(w)\}} f(\mathcal{Y}_{s-T^n(w)}(w'))$  converges to  $\phi(w, w')$ . Furthermore, we have

$$\phi^n(w, w') = \sum_{k=0}^{+\infty} \mathbf{1}_{\{T^n(w)=s_k^n\}} \mathbf{1}_{\{s \geq T^n(w)\}} f(\mathcal{Y}_{s-s_k^n}(w')) = \sum_{k=0}^{2^n} \mathbf{1}_{\{T^n(w)=s_k^n\}} f(\mathcal{Y}_{s-s_k^n}(w')),$$

where  $f(\mathcal{Y}_{s-s_k^n})$  is  $\mathcal{W}_\infty$ -measurable and  $\mathbf{1}_{\{T^n=s_k^n\}}$  is  $\mathcal{W}_\zeta$ -measurable, for  $k \leq 2^n$ . Indeed, for  $t \geq 0$ , we use the fact that  $\zeta$  is a stopping time with respect to the right-continuous<sup>19</sup> filtration  $(\mathcal{W}_t)_{t \geq 0}$  to obtain

$$\{T^n = s_k^n\} \cap \{\zeta \leq t\} = \{s_k^n \leq \zeta < s_{k+1}^n\} \cap \{\zeta \leq t\} = \begin{cases} \emptyset \in \mathcal{W}_t \text{ if } t < s_k^n, \\ \{s_k^n \leq \zeta < s_{k+1}^n\} \in \mathcal{W}_{s_{k+1}^n} \subseteq \mathcal{W}_t \text{ if } t \geq s_{k+1}^n, \\ \{s_k^n \leq \zeta \leq t\} \in \mathcal{W}_t \text{ if } s_k^n \leq t < s_{k+1}^n, \end{cases}$$

which shows  $\{T^n = s_k^n\} \in \mathcal{W}_\zeta$ . Thus, as a limit of  $\mathcal{W}_\zeta \otimes \mathcal{W}_\infty$ -measurable random variables,  $\phi$  is itself  $\mathcal{W}_\zeta \otimes \mathcal{W}_\infty$ -measurable.

Now there only remains to show that<sup>20</sup> if  $\phi$  is any  $\mathcal{W}_\zeta \otimes \mathcal{W}_\infty$ -measurable and real-valued bounded random variable on  $W^2$  and

$$\forall w \in W, \quad Z(w) := \phi(w, \psi_\zeta w),$$

<sup>18</sup>Here, we keep the same filtration as for the process  $\mathfrak{P}^\zeta \mathcal{Y}$ . Furthermore, we still consider starting points in  $E$  rather than  $E_j$ : (c) gives for  $\mathfrak{P}^\zeta Y^j$

$$\forall y \in \overline{E} \quad \mathfrak{P}^\zeta \mathbb{P}_y(\mathfrak{P}^\zeta Y_0^j = \pi_j(y)) = 1.$$

<sup>19</sup>We use this assumption to obtain that  $\{\zeta < t\}$  is  $\mathcal{W}_t$  for all  $t \geq 0$ .

<sup>20</sup>This is a part of Exercise III.3.19 in [\[RY99\]](#). This exercise only mentions positive random variables, but the strategy of proof would be the same, and we decided to adapt it to the context we need here.

we have for  $\mathbb{P}_y$ -almost every  $w$

$$\mathbb{E}_y(Z|\mathcal{W}_\zeta)(w) = \mathbb{E}_{\mathcal{Y}_\zeta(w)}(\phi(w, \cdot)). \quad (\text{A.28})$$

Indeed, in our special case where  $\phi$  is as in (A.27), taking the expectation with respect to  $\mathbb{P}_y$  in (A.28) gives (A.26).

In order to prove (A.28), we first prove it for  $\phi$  of the form  $(w, w') \mapsto g(w)h(w')$ , where  $g$  and  $h$  are respectively  $\mathcal{W}_\zeta$ - and  $\mathcal{W}_\infty$ -measurable and real-valued bounded functions. In that case, we use the fact that  $g$  is  $\mathcal{W}_\zeta$ -measurable and the strong Markov property at time  $\zeta$  to have  $\mathbb{P}_y$ -almost surely

$$\mathbb{E}_y(Z|\mathcal{W}_\zeta)(w) = g(w)\mathbb{E}_y(h \circ \psi_\zeta|\mathcal{W}_\zeta)(w) = g(w)\mathbb{E}_{\mathcal{Y}_\zeta(w)}(h),$$

and then as a constant  $g(w)$  can be put inside the expectation again to obtain  $\mathbb{P}_y$ -almost surely

$$\mathbb{E}_y(Z|\mathcal{W}_\zeta)(w) = \mathbb{E}_{\mathcal{Y}_\zeta(w)}(g(w)h) = \mathbb{E}_{\mathcal{Y}_\zeta(w)}(\phi(w, \cdot)).$$

Then we use the monotone class theorem<sup>21</sup> to deduce that (A.28) is indeed true for any  $\mathcal{W}_\zeta \otimes \mathcal{W}_\infty$ -measurable and real-valued bounded random variable on  $W^2$ .  $\square$

*Proof of Proposition A.3.* We first prove the result for a truncated version of our construction (with a finite number of pastings) and then pass to the limit. For  $k \geq 1$  and  $\mathfrak{P}w \in \mathfrak{P}W$ , define the process  $\mathfrak{P}_k^\zeta \mathcal{Y}$  corresponding to the pasting of the first  $k$  pieces of trajectories by

$$\mathfrak{P}_k^\zeta \mathcal{Y}_t(\mathfrak{P}w) = \begin{cases} \mathcal{Y}_{t-\bar{\zeta}_{i-1}(\mathfrak{P}w)}(w_i) & \text{if } \bar{\zeta}_{i-1}(\mathfrak{P}w) \leq t < \bar{\zeta}_i(\mathfrak{P}w), 1 \leq i \leq k-1, \\ \mathcal{Y}_{t-\bar{\zeta}_{k-1}(\mathfrak{P}w)}(w_k) & \text{if } t \geq \bar{\zeta}_{k-1}(\mathfrak{P}w). \end{cases}$$

Notice that since we take the process  $\mathcal{Y}$  rather than its killed version for the last part,  $\mathfrak{P}_k^\zeta \mathcal{Y}$  still takes its values in  $E$ ,  $\mathfrak{P}^\zeta \mathbb{P}_y^\lambda$ -almost surely for all  $y$  in  $E$ . Now we prove by induction on  $k$  that  $\mathfrak{P}_k^\zeta Y_t^j$  has the same distribution as  $Y_t^j$ . This is true (even for all coordinates) for  $k = 1$  because  $\mathfrak{P}_1^\zeta \mathcal{Y} = \mathcal{Y}$ , no pasting being involved. Seen as a random variable on  $\mathfrak{P}W$ ,  $\mathfrak{P}_1^\zeta \mathcal{Y}_t$  thus depends only on the first coordinate: we then have by (A.5) for  $f \in \mathcal{C}_0(\bar{E})$

$$\mathfrak{P}^\zeta \mathbb{E}_y^\lambda(f(\mathfrak{P}_1^\zeta \mathcal{Y}_t)) = \int_\Omega f(\mathcal{Y}_t(w_1)) \mathbb{Q}_y^\lambda(dw_1) = \int_W f(\mathcal{Y}_t(w_1)) \mathbb{P}_y(dw_1) = \mathbb{E}_y(f(\mathcal{Y}_t)),$$

which proves in particular that  $Y^j$  and  $\mathfrak{P}_1^\zeta Y^j$  share the same distribution.

The induction is based on the fact that we have for  $k \geq 1$

$$\mathfrak{P}_{k+1}^\zeta \mathcal{Y}_t(\mathfrak{P}w) = \begin{cases} \mathfrak{P}_k^\zeta \mathcal{Y}_t(\mathfrak{P}w) & \text{if } t < \bar{\zeta}_k(\mathfrak{P}w), \\ \mathcal{Y}_{t-\bar{\zeta}_k(\mathfrak{P}w)}(w_{k+1}) & \text{if } t \geq \bar{\zeta}_k(\mathfrak{P}w). \end{cases}$$

Let  $f^j \in \mathcal{C}_0(\bar{E}_j)$  and  $k \geq 1$ , and assume that  $\mathfrak{P}_k^\zeta Y_t^j$  has the same distribution as  $Y_t^j$ . By definition, we have

$$\begin{aligned} \mathfrak{P}^\zeta \mathbb{E}_y^\lambda(f^j(\mathfrak{P}_{k+1}^\zeta Y_t^j)) &= \mathfrak{P}^\zeta \mathbb{E}_y^\lambda(\mathbf{1}_{\{t < \bar{\zeta}_k\}} \widehat{f}^j(\mathfrak{P}_k^\zeta \mathcal{Y}_t)) \\ &\quad + \int_{\Omega^{k+1}} \mathbf{1}_{\{t \geq \bar{\zeta}_k(w^k)\}} \widehat{f}^j(\mathcal{Y}_{t-\bar{\zeta}_k(w^k)}(w_{k+1})) \mathbb{Q}_{y_k}^\lambda(dw_{k+1}) \cdots \mathbb{Q}_y^\lambda(dw_1). \end{aligned} \quad (\text{A.29})$$

<sup>21</sup>See Theorem 5.2.2 in [Dur19] for instance, the Assumptions (ii) and (iii) being verified for the class of random variables such that (A.28) is true by linearity of the expectation and the monotone convergence theorem.

Denote by  $\mathcal{I}$  the second term in the r.h.s. of [\(A.29\)](#), and fix  $\omega^{k-1} \in \Omega^{k-1}$  such that  $t \geq \bar{\zeta}_{k-1}(w^{k-1})$  (if  $k = 1$ , we have the convention  $\bar{\zeta}_0 = 0$ , without needing to fix any random realization). Let  $s = t - \bar{\zeta}_{k-1}(w^{k-1})$  and

$$\forall (w_k, w_{k+1}) \in W^2 \quad \phi(w_k, w_{k+1}) := \mathbf{1}_{\{s \geq \zeta(w_k)\}} \widehat{f}^j(\mathcal{Y}_{s-\zeta(w_k)}(w_{k+1})).$$

Recall that we write  $\omega_i = (w_i, y_i)$  the elements of  $\Omega = W \times \bar{E}$ . First, since  $\phi$  does not depend on  $y_{k+1}$ , we have for  $\omega_k \in \Omega$

$$\int_{\Omega} \phi(w_k, w_{k+1}) \mathbb{Q}_{y_k}^\lambda(dw_{k+1}) = \int_W \phi(w_k, w_{k+1}) \mathbb{P}_{y_k}(dw_{k+1}) = \mathbf{1}_{\{s \geq \zeta(w_k)\}} \mathcal{P}_{s-\zeta(w_k)} \widehat{f}^j(y_k).$$

Thanks to [\(A.24\)](#), we then have

$$\int_{\Omega} \phi(w_k, w_{k+1}) \mathbb{Q}_{y_k}^\lambda(dw_{k+1}) = \mathbf{1}_{\{s \geq \zeta(w_k)\}} P_{s-\zeta(w_k)}^j f^j(y_k^j),$$

and now we integrate over  $y_k \in \bar{E}$  w.r.t.  $\lambda(w_k, \cdot)$ : thanks to Assumption [\(A.25\)](#),  $\lambda$  preserves (almost surely) the  $j$ -th coordinate, so that the r.h.s. does not depend on  $y_k$ , but rather is constant:

$$\int_{\Omega \times \bar{E}} \phi(w_k, w_{k+1}) \mathbb{Q}_{y_k}^\lambda(dw_{k+1}) \lambda(w_k, dy_k) = \mathbf{1}_{\{s \geq \zeta(w_k)\}} P_{s-\zeta(w_k)}^j f^j(Y_\zeta^j(w_k)).$$

We use [\(A.24\)](#) in reverse and integrate over  $w_k \in W$  w.r.t.  $\mathbb{P}_{y_{k-1}}$  to obtain

$$\int_{\Omega^2} \phi(w_k, w_{k+1}) \mathbb{Q}_{y_k}^\lambda(dw_{k+1}) \mathbb{Q}_{y_{k-1}}^\lambda(dw_k) = \int_{\Omega} \mathbf{1}_{\{s \geq \zeta(w_k)\}} \mathcal{P}_{s-\zeta(w_k)} \widehat{f}^j(\mathcal{Y}_\zeta(w_k)) \mathbb{Q}_{y_{k-1}}^\lambda(dw_k).$$

Again, nothing depends on  $y_{k+1}$ , so we can get rid of the integral on that variable: we have

$$\int_{\Omega^2} \phi(w_k, w_{k+1}) \mathbb{Q}_{y_k}^\lambda(dw_{k+1}) \mathbb{Q}_{y_{k-1}}^\lambda(dw_k) = \int_W \mathbf{1}_{\{s \geq \zeta(w_k)\}} \mathcal{P}_{s-\zeta(w_k)} \widehat{f}^j(\mathcal{Y}_\zeta(w_k)) \mathbb{P}_{y_{k-1}}(dw_k).$$

Now the r.h.s. is of the form of the r.h.s. in [\(A.26\)](#), so that we have thanks to [Lemma A.2](#)

$$\int_{\Omega^2} \phi(w_k, w_{k+1}) \mathbb{Q}_{y_k}^\lambda(dw_{k+1}) \mathbb{Q}_{y_{k-1}}^\lambda(dw_k) = \int_W \mathbf{1}_{\{s \geq \zeta(w_k)\}} \widehat{f}^j(\mathcal{Y}_s(w_k)) \mathbb{P}_{y_{k-1}}(dw_k).$$

Finally, we integrate over  $\omega_1, \dots, \omega_{k-1}$  to obtain, using the convention  $y_0 = y$ ,

$$\begin{aligned} \mathcal{I} &= \int_{\Omega^{k+1}} \phi(w_k, w_{k+1}) \prod_{l=1}^{k+1} \mathbb{Q}_{y_{l-1}}^\lambda(dw_l) \\ &= \int_{W \times \Omega^{k-1}} \mathbf{1}_{\{t \geq \bar{\zeta}_k(w^k)\}} \widehat{f}^j(\mathcal{Y}_{t-\bar{\zeta}_{k-1}(w^{k-1})}(w_k)) \mathbb{P}_{y_{k-1}}(dw_k) \prod_{l=1}^{k-1} \mathbb{Q}_{y_{l-1}}^\lambda(dw_l), \end{aligned}$$

the product in the last integral being empty when  $k = 1$ . Since, for  $w^k$  such that  $t \geq \bar{\zeta}_{k-1}(w^{k-1})$ ,  $\mathfrak{P}_k^\zeta \mathcal{Y}_t(w^k)$  agrees with  $\mathcal{Y}_{t-\bar{\zeta}_{k-1}(w^{k-1})}(w_k)$ , we obtain  $\mathcal{I} = \mathfrak{P}^\zeta \mathbb{E}_y^\lambda(\mathbf{1}_{\{t \geq \bar{\zeta}_k\}} \widehat{f}^j(\mathfrak{P}_k^\zeta \mathcal{Y}_t))$ . Looking back at [\(A.29\)](#) and using the induction hypothesis, we now have

$$\mathfrak{P}^\zeta \mathbb{E}_y^\lambda(\widehat{f}^j(\mathfrak{P}_{k+1}^\zeta \mathcal{Y}_t)) = \mathfrak{P}^\zeta \mathbb{E}_y^\lambda(\widehat{f}^j(\mathfrak{P}_k^\zeta \mathcal{Y}_t)) = P^j f_t^j(y^j).$$

It remains to take the limit in  $k \rightarrow +\infty$ : since by assumption,  $\mathfrak{T}$  is almost surely infinite,  $\bar{\zeta}_k$  goes to infinity almost surely, so that  $\mathfrak{P}_k^\zeta \mathcal{Y}_t$  goes almost surely to  $\mathfrak{P}^\zeta \mathcal{Y}_t$ . We can then use the dominated convergence theorem to prove

$$\mathfrak{P}^\zeta \mathbb{E}_y^\lambda(f^j(\mathfrak{P}^\zeta Y_t^j)) = \lim_{k \rightarrow +\infty} \mathfrak{P}^\zeta \mathbb{E}_y^\lambda(f^j(\mathfrak{P}_k^\zeta Y_t^j)) = P_t^j f^j(y^j).$$

Furthermore, the Markov property for  $\mathfrak{P}^\zeta \mathcal{Y}$  applied to the function  $\widehat{f}^j$  gives

$$\mathfrak{P}^\zeta \mathbb{E}_y^\lambda(\widehat{f}^j(\mathfrak{P}^\zeta \mathcal{Y}_{t+s}) | \mathfrak{P} \mathcal{W}_s) = \mathfrak{P}^\zeta \mathbb{E}_{\mathfrak{P}^\zeta \mathcal{Y}_s}^\lambda(\widehat{f}^j(\mathfrak{P}^\zeta \mathcal{Y}_t)) = P_t^j f^j(\mathfrak{P}^\zeta Y_s^j), \quad (\text{A.30})$$

and this quantity depends on  $\mathfrak{P}^\zeta \mathcal{Y}_s$  only through  $\mathfrak{P}^\zeta Y_s^j$ . Hence  $\mathfrak{P}^\zeta Y^j$  inherits the Markov property from  $\mathfrak{P}^\zeta \mathcal{Y}$ , so that its distribution is fully characterized by its semigroup, and thus the same as that of  $Y^j$ . Furthermore, (A.30) is also true almost surely on  $\{T < +\infty\}$  when replacing  $s$  by a stopping time  $T$  (thanks to the strong Markov property for  $\mathfrak{P}^\zeta \mathcal{Y}$ ), so  $\mathfrak{P}^\zeta Y^j$  is indeed a strong Markov process.  $\square$

## A.2 Formal definition of the coupling

We can now focus on the more technical aspects of the proof of Proposition I.4: we will use the pasting procedure described in Section A.1 to create a strong Markov process in which we find a coupling of an  $N$ -BMP and a  $b$ -BMP that are stochastically ordered until the first time when the  $b$ -BMP has less particles than the  $N$ -BMP. To ensure this, we will use the definitions of these processes and the properties of the pasting procedure that we just proved.

### A.2.1 Construction of the coupling

As for the constructions of the branching processes described in Section A.1.2, we decompose the coupling into pieces of trajectories during which no branching or selection occurs. Within each piece, the base process consists in pairs of particles (one for the  $N$ -BMP and one for the  $b$ -BMP) evolving with constant ordering (built thanks to Theorem II.1<sup>22</sup>), and lonely particles evolving independently. Branching and selection are then implemented at random times, between which we paste these pieces of trajectories.

#### A.2.1.1 State space and base process

Following the notation of Section A.1.2, we consider  $E_1 = \mathbb{R}^N$ ,  $E_2 = \left( \{\diamond\} \sqcup \bigsqcup_{n \geq 1} \mathbb{R}^n \right) \times \mathbb{R}_+$ , and we set  $E_3 = \{p = (p_1, \dots, p_N) \in \mathbb{N}^{*N}, i \neq j \implies p_i \neq p_j\} \cup \{\partial\}$ , which is a discrete space that will encode which particles of the  $N$ -BMP and the  $b$ -BMP evolve in pairs or not. More precisely, we set  $E = E_1 \times E_2 \times E_3$ , and we will use the pasting procedure to build a strong Markov process on  $E$  such that the  $E_1$ -coordinate is an  $N$ -BMP and the  $E_2$ -coordinate a  $b$ -BMP, with stochastic domination<sup>23</sup> of the  $E_2$ -coordinate over the  $E_1$ -coordinate until the first time when the  $E_2$ -coordinate has less than  $N$  particles. We will generically write  $y = (x, (x^+, s), p) = (x, x^+, s, p)$  an element of  $E$ .

Furthermore, starting from a state  $y$  with the following properties, we will ensure that those remain true at all times:

<sup>22</sup>Namely, we consider for each pair a process on  $\mathbb{R}^2$  evolving according to the semigroup  $(P_t^{(2)})_{t \geq 0}$ .

<sup>23</sup>Recall the definition of the stochastic ordering between tuples of positions given in Section I.3

(P1) if  $p = (p^1, \dots, p^N)$ , then  $p^i \leq n$  for all  $1 \leq i \leq N$  (in particular,  $n \geq N$ ),

(P2) if  $p = (p^1, \dots, p^N)$ , then  $x^i \leq x^{p^i, +}$  for all  $1 \leq i \leq N$ .

The  $E_3$ -coordinate is then interpreted as follows:

- if  $p = (p_1, \dots, p_N)$ , for all  $1 \leq i \leq N$ , the  $i$ -th particle in  $E_1$  and the  $p_i$ -th particle in  $E_2$  evolve in pair in order to preserve the ordering between them; all the other particles evolve independently,
- if  $p = \partial$ , all the particles evolve independently.

Recall Assumptions (i), (ii) and (iii) of [Theorem I.1](#): the semigroup  $(P_t)_{t \geq 0}$  of the underlying Markov process  $X$  is Feller and stochastically monotone. Thanks to [Theorem II.1](#), we also have another Feller semigroup  $(P_t^{(2)})_{t \geq 0}$  on  $\mathbb{R}^2$  driving two copies of  $X$  with the same ordering across time<sup>24</sup>. We can thus define a first Feller process  $\mathcal{Y} = (\mathcal{X}, \mathcal{X}^+, \mathcal{T}, \Pi)$  on  $E$  as follows: starting from  $y = (x, x^+, s, p)$  in  $E$ ,

- $\Pi_t$  is constant, equal to  $p$ ,
- $\mathcal{T}_t$  is deterministic too, and grows linearly with  $t$ , *i.e.*  $\mathcal{T}_t = s + t$ ,
- if  $p = (p^1, \dots, p^N)$ ,  $(X^i, X^{p^i, +})$  evolves according to the semigroup  $(P_t^{(2)})_{t \geq 0}$  for all  $1 \leq i \leq N$ , independently from the others,
- all 'free particles' (those from  $\mathcal{X}^+$  whose index is not in  $p$ ) evolve according to the semigroup  $(P_t)_{t \geq 0}$ , independently from the other particles,
- if  $p = \partial$ , the  $N + n$  particles evolve according to the semigroup  $(P_t)_{t \geq 0}$ , independently from each other.

To sum up,  $\mathcal{Y}$  is associated with the semigroup  $(\mathcal{P}_t)_{t \geq 0}$ , defined on the set of bounded measurable functions on  $E$  and generated by its values on the tensor products  $f_1 \cdots f_N g_1 \cdots g_n h_1 h_2$  of families  $(f_1, \dots, f_N, g_1, \dots, g_n, h_1, h_2)$ , with  $f_i, g_j \in \mathcal{B}_b(\mathbb{R})$ ,  $h_1 \in \mathcal{B}_b(\mathbb{R}_+)$  and  $h_2 \in \mathcal{B}_b(E_3)$ , which is

$$\mathcal{P}_t(f_1 \cdots f_N g_1 \cdots g_n h_1 h_2)(y) = \begin{cases} \prod_{i=1}^N P_t^{(2)}(f_i g_{p_i})(x^i, x^{p^i, +}) \prod_{j \notin p} P_t g_j(x^j) h_1(s+t) h_2(p) & \text{if } p \neq \partial, \\ \prod_{i=1}^N P_t f_i(x^i) \prod_{j=1}^n P_t g_j(x^{j, +}) h_1(s+t) h_2(\partial) & \text{otherwise.} \end{cases}$$

With this definition, we observe that  $\mathcal{X}$  has the distribution of  $N$  independent copies of the process  $X$  starting at  $x = (x^1, \dots, x^N)$ , and  $\mathcal{X}^+$  that of  $n$  independent copies of  $X$  starting at  $x^+ = (x^{1, +}, \dots, x^{n, +})$ . Starting from an element  $y$  satisfying (P1) and (P2), these properties are indeed preserved almost surely by this evolution: (P1) depends only on the  $E_3$ -coordinate, which is constant; and (P2) is preserved because each pair  $(\mathcal{X}^i, \mathcal{X}^{p^i, +})$  evolves according to the semigroup  $(P_t^{(2)})_{t \geq 0}$ , which preserves the ordering.

<sup>24</sup>More precisely,  $(P_t^{(2)})_{t \geq 0}$  is associated with a Markov kernel  $P_t^{(2)}$  on  $\mathbb{R}^2$ , such that for all Borel set  $B$  and all  $(x_1, x_2) \in \mathbb{R}^2$ , we have

$$\begin{cases} P_t^{(2)}((x_1, x_2), B \times \mathbb{R}) = P_t(x_1, B), \\ P_t^{(2)}((x_1, x_2), \mathbb{R} \times B) = P_t(x_2, B), \end{cases}$$

and if  $S_+ = \{(x_1, x_2) \in \mathbb{R}^2, x_1 \leq x_2\}$  and  $S_- = \{(x_1, x_2) \in \mathbb{R}^2, x_1 \geq x_2\}$ ,  $P_t^{(2)}((x_1, x_2), \cdot)$  has support in  $S_+$  in  $(x_1, x_2) \in S_+$  and in  $S_-$  if  $(x_1, x_2) \in S_-$ .



We will thus consider only states verifying these properties, meaning that in the following  $\tilde{E}$  will refer to the subset of  $E_1 \times E_2 \times E_3$  of elements satisfying (P1) and (P2). This will ensure that there is a stochastic ordering between the  $E_1$ -coordinate and the  $E_2$ -space coordinate, since if  $y = (x, x^+, s, p)$  is in  $\tilde{E}$  (meaning satisfying (P1) and (P2)) and  $p \neq \partial$ , we have  $x \preceq x^+$ . Indeed, when  $p \neq \partial$ , for all level  $r$ , if  $x \cap [r, +\infty) = (x^{i_1}, \dots, x^{i_m})$ ,  $x^+ \cap [r, +\infty)$  contains at least  $(x^{p_{i_1}^+}, \dots, x^{p_{i_m}^+})$  and thus has at least as many elements.

### A.2.1.2 Probability space and pasting times

Now we can add the branching and death times thanks to the pasting procedure described in the previous section. There will be three types of stopping times:

- (i) we make the paired particles branch together; thus we will use a first stopping time  $\theta^1$ , with exponential distribution  $\mathcal{E}(N)$  and independent from  $\mathcal{Y}$ ,
- (ii) the free particles (those in the  $E_2$ -coordinate that are not paired) also branch, so there is a second exponential time  $\theta^2$  whose parameter depends on the number of free particles,
- (iii) for the  $E_2$ -coordinate, the selection occurs at the first time  $\zeta^2$  when a particle goes below the curve  $b$ .

To define these random times, we need two independent random components. Since the semigroup  $(\mathcal{P}_t)_{t \geq 0}$  is Feller, we can consider the canonical realization of its associated Feller process on the set of càdlàg trajectories  $\mathcal{D} = \mathcal{D}(\mathbb{R}_+, E)$ . In the end, we build the process  $(\mathcal{Y}_t)_{t \geq 0}$  on the probability space  $W = \mathcal{D}(\mathbb{R}_+, E) \times \mathbb{R}_+^2$  equipped with the product  $(\mathcal{W}_t)_{t \geq 0}$  of the completion<sup>25</sup> of the canonical right-continuous filtration on  $\mathcal{D}$  and of the Borel  $\sigma$ -algebras  $\mathcal{B}_t$  generated by the  $[0, s]$ , for  $s \leq t$ , on each  $\mathbb{R}_+$  component of  $W$ .

On  $\mathcal{D}$ , we consider a family  $(\mathbb{S}_y^1)_{y \in E}$  of probability measures such that the canonical<sup>26</sup> process  $(\mathcal{Y}_t)_{t \geq 0}$  on  $\mathcal{D}$  is Feller with semigroup  $(\mathcal{P}_t)_{t \geq 0}$ , and remember that we denote its coordinates by  $\mathcal{Y}_t = (\mathcal{X}_t, \mathcal{X}_t^+, \mathcal{T}_t, \Pi_t)$ . We denote by  $\theta^1$  the projection on the first  $\mathbb{R}_+$  coordinate of  $W$ , equipped with the probability measure  $\mathbb{S}^2$  associated with the exponential distribution  $\mathcal{E}(N)$ , and  $\theta^2$  the projection on the second  $\mathbb{R}_+$  coordinate of  $W$ . For  $y = (x, x^+, s, p)$  in  $E$ , let  $\mathbb{S}_y^3$  be the probability measure on  $\mathbb{R}_+$  corresponding to the exponential distribution<sup>27</sup>  $\mathcal{E}(n - N)$  if  $p = (p^1, \dots, p^N)$ , and  $\mathcal{E}(n)$  if  $p = \partial$ . Now equip  $W$  with the probability measures  $\mathbb{P}_y = \mathbb{S}_y^1 \otimes \mathbb{S}^2 \otimes \mathbb{S}_y^3$ , under which we have a càdlàg realization of our Feller process with two independent random times  $\theta^1$  and  $\theta^2$  (which are stopping times with respect to  $(\mathcal{W}_t)_{t \geq 0}$ ).

We can now define the death time  $\zeta^2$  of a particle of  $\mathcal{X}^+$ : recall that there exists an upper semi-continuous boundary  $b : \mathbb{R}_+^* \rightarrow \mathbb{R}$  and that the  $b$ -BMP corresponds to a BMP whose particles are killed when they go below that boundary. Let

$$\mathcal{U} = \{y = (x, x^+, s, p) \in E, \exists i \in \llbracket 1, n \rrbracket x^{i,+} < b_s\}$$

and let

$$\zeta^2 = \inf\{t \geq 0, \mathcal{Y}_t \in \mathcal{U}\}.$$

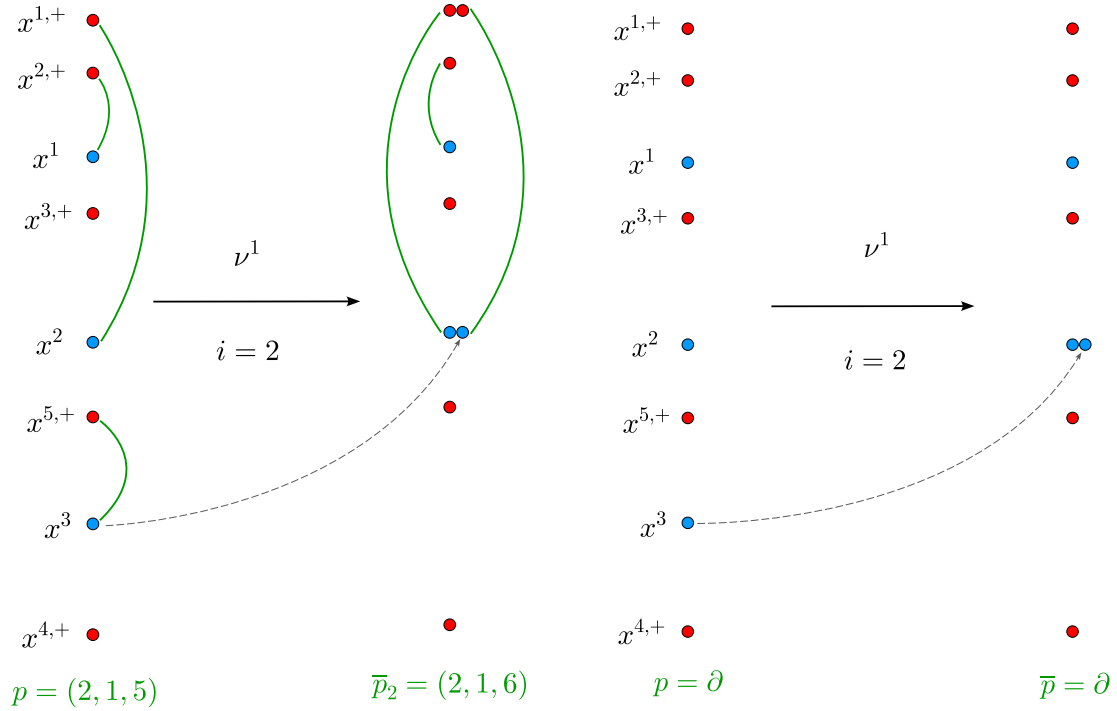
Then  $\zeta^2$  is the first time at which one the particles of  $\mathcal{X}^+$  goes below the boundary  $b$ , and is a stopping time with respect to the filtration  $(\mathcal{W}_t)_{t \geq 0}$ . Indeed, this is the minimum of  $n$  crossing times for each particle, of the form

$$\tau_b^{i,+} = \inf\{s \geq 0, X_s^{i,+} < b_s\},$$

<sup>25</sup>With respect to all probability measures on  $W$ .

<sup>26</sup>The canonical process on  $\mathcal{D}$  is given by  $\mathcal{Y}_t(d) = d_t$ , for  $t \geq 0$  and  $d \in \mathcal{D}$ .

<sup>27</sup>Recall property (P1) that ensures that  $n - N$  is non-negative when  $p \neq \partial$ .



(ia) The coupling between the  $N$ -BMP and the  $b$ -BMP is still active

(ib) The coupling between the  $N$ -BMP and the  $b$ -BMP is no longer active

Figure A.3: Renewal distributions used to build the coupling - Case (i). Red bullet points represent the particles of the  $b$ -BMP, and blue bullet points the particles of the  $N$ -BMP; green bonds show the pairing between particles. The dashed grey arrow is drawn from the killed particle to the newborn particle in the  $N$ -BMP.

as defined in the introduction, which are themselves stopping times:  $\tau_b^{i,+}$  is the first entry time of the set  $\{(s, x) \in \mathbb{R}_+ \times \mathbb{R}, x < b_s\}$  (which is the union of the sets  $\{(s, x) \in \mathbb{R}_+ \times \mathbb{R}, x \leq b_s - \frac{1}{n}\}$ , for  $n \geq 1$ , which are closed by upper semi-continuity of  $b$ ) for the right-continuous process  $(t, X_t)_{t \geq 0}$ , so that we can apply the D ebut theorem (see Theorem IV.50 in [DM75]).

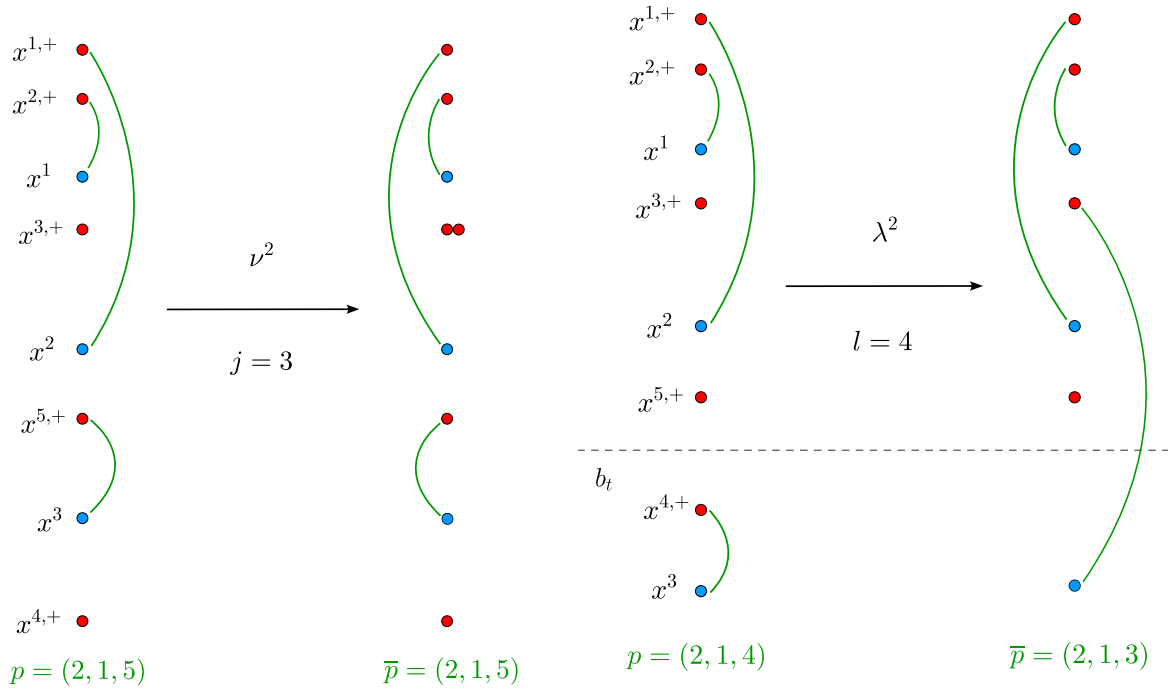
Finally, we define the shift operators  $\psi_t$  on  $W$  in the following way, which ensures that the compatibility equation [A.1] is always true for the different stopping times: on the  $\mathcal{D}$ -coordinate, it is the canonical shift, and on both  $\mathbb{R}_+$ -coordinates, it is the translation, meaning that for  $w = (w^1, w^2, w^3) \in W$ , we have

$$\psi_t(w) = ((w_{s+t}^1)_{s \geq 0}, \mathbf{1}_{\{w^2 \geq t\}}(w^2 - t), \mathbf{1}_{\{w^3 \geq t\}}(w^3 - t)).$$

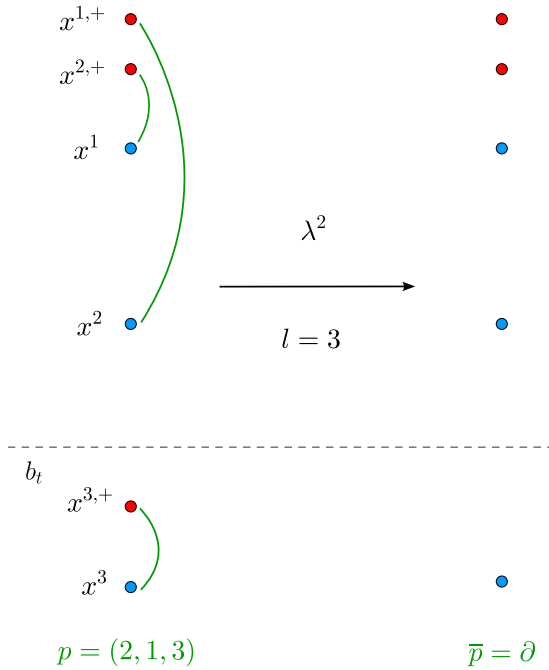
### A.2.1.3 Renewal distribution

Now let us describe how we paste the trajectories together at those stopping times, by defining different renewal distributions, each associated with one type of renewal time. The different cases are presented in Figures [A.3] and [A.4].

- (i) At time  $\theta^1$ , we paste with the distribution  $\nu^1(w, \cdot) = \bar{\nu}^1(\mathcal{Y}_{\theta^1}(w), \cdot)$  constructed below, with two possibilities depending on the  $E_3$ -coordinate. Denote  $y = (x, x^+, s, p) = \mathcal{Y}_{\theta^1}(w)$ .

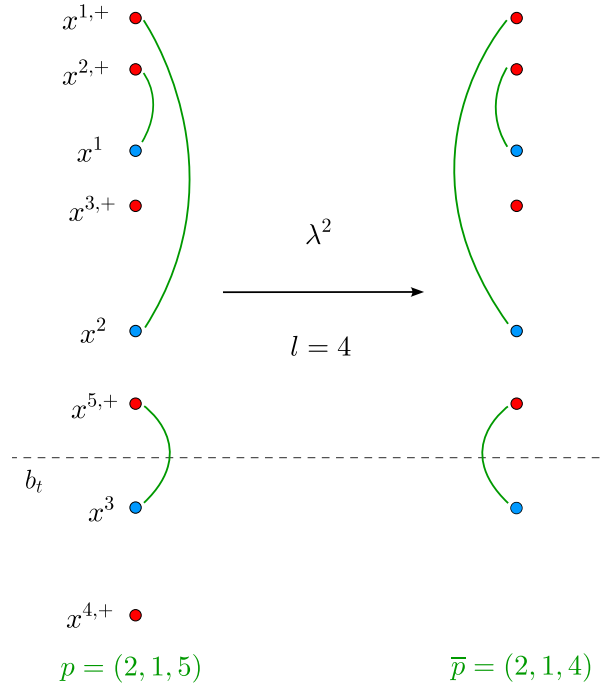


(ii) A new particle is born in the  $b$ -BMP at the position  $x^{3,+}$ .



(iiiia) The  $b$ -BMP particle at the position  $x^{4,+}$  is killed, and no particle is available to recouple the  $N$ -BMP particle at the position  $x^3$ : the processes start evolving independently.

(iiia) The  $b$ -BMP particle at the position  $x^{4,+}$  is killed, the  $N$ -BMP particle at the position  $x^3$  is recoupled.



(iiib) The  $b$ -BMP particle at position  $x^{4,+}$  is killed, and was not coupled.

Figure A.4: Renewal distributions used to build the coupling - Cases (ii) and (iii). Conventions are the same as in [Figure A.3](#); we added a dashed black line at the position of the boundary, below which  $b$ -BMP particles (in red) get killed.

- (a) If  $p = (p_1, \dots, p_N)$ , let  $m$  be the index of the lowest  $E_1$ -particle<sup>28</sup> in  $y$ , and define, for  $1 \leq i \leq N$ ,  $\beta_i(y) := (\bar{x}_i, \bar{x}_{p_i}^+, s, \bar{p}_i)$  with

$$\begin{cases} \bar{x}_i = (x^1, \dots, x^{m-1}, x^i, x^{m+1}, \dots, x^N), \\ \bar{x}_{p_i}^+ = (x^{1,+}, \dots, x^{n,+}, x^{p_i,+}), \\ \bar{p}_i = (p_1, \dots, p_{m-1}, n+1, p_{m+1}, \dots, p_N). \end{cases}$$

We then set  $\bar{\nu}^1(y, \cdot)$  to be the distribution of  $\beta_I(y)$ , where  $I$  is uniformly distributed on  $\llbracket 1, N \rrbracket$ .

In other words, since  $p \neq \partial$ , the  $E_1$ - and  $E_2$ -coordinate are still interacting: we perform the branching of one  $E_1$ -particle, at position  $x^i$ , chosen uniformly at random, and that of its paired  $E_2$ -particle at position  $x^{p_i,+}$ , while killing the lowest  $E_1$ -particle at position  $x^m$  and letting its paired  $E_2$ -particle at position  $x^{p_m,+}$  evolve freely and pairing the newborn particles together.

- (b) If  $p = \partial$ ,  $\bar{\nu}^1(y, \cdot)$  is the distribution of  $\beta_I(y) = (\bar{x}_I, x^+, s, \partial)$ , where  $I$  is uniformly distributed on  $\llbracket 1, N \rrbracket$ : the processes evolve independently, so that we only perform the branching of an  $E_1$ -particle and the killing of the lowest  $E_1$ -particle.
- (ii) At time  $\theta^2$ , we use the renewal distribution  $\nu^2(w, \cdot) = \bar{\nu}^2(\mathcal{Y}_{\theta^2}(w), \cdot)$ , where  $\bar{\nu}^2(y, \cdot)$  is the distribution of  $\beta_J^+(y)$ , where  $\beta_J^+(y) = (x, \bar{x}_J^+, s, p)$  and  $J$  is uniformly distributed on  $\llbracket 1, n \rrbracket \setminus \{p_i, 1 \leq i \leq N\}$ : we perform the branching of a free  $E_2$ -particle, chosen uniformly among all free  $E_2$ -particles.
- (iii) At time  $\zeta^2$ , we paste with the distribution  $\lambda^2(w, \cdot) = \bar{\lambda}^2(\mathcal{Y}_{\zeta^2}(w), \cdot)$ , that performs the killing of the lowest<sup>29</sup>  $E_2$ -particle, which is necessarily below the curve  $b$  (thanks to the definition of  $\zeta^2$ ). There are two possibilities, depending on whether the particle that gets killed is paired or not. Denote  $y = (x, x^+, s, p) = \mathcal{Y}_{\zeta^2}(w)$ .

- (a) If the  $E_2$ -particle which is killed is paired with an  $E_1$ -particle, there is an integer  $i \in \llbracket 1, N \rrbracket$  such that its index is  $p_i$ . Then for  $1 \leq j \leq n$ , write

$$\sigma_{p_i}(l) = \begin{cases} l & \text{if } l < p_i, \\ l - 1 & \text{otherwise,} \end{cases}$$

and define  $x'_{p_i} = (x^{\sigma_{p_i}(l),+})_{l \neq p_i}$ . If  $n = N$ , set  $\bar{\lambda}^2(y, \cdot)$  to be the Dirac mass at  $(x, x'_{p_i}, s, \partial)$ . Otherwise<sup>30</sup>, we have  $n > N$ , and  $\llbracket 1, n \rrbracket \setminus \{p_i, 1 \leq i \leq N\}$  is non-empty. Denote by  $j$  its smallest element<sup>31</sup>, write

$$p' = (\sigma_{p_i}(p_1), \dots, \sigma_{p_i}(p_{i-1}), \sigma_{p_i}(j), \sigma_{p_i}(p_{i+1}), \dots, \sigma_{p_i}(p_N)),$$

and define  $\bar{\lambda}^2(y, \cdot)$  to be the Dirac mass on  $(x, x'_{p_i}, s, p')$ . This amounts to killing  $x^{p_i,+}$  and replacing it by the particle that is at position  $x^{j,+}$  in the pair  $(x^i, x^{p_i,+})$ .

- (b) If the  $E_2$ -particle which is below the curve is not paired, we just have to remove it and to re-index the others:  $\bar{\lambda}^2(y, \cdot)$  is the Dirac mass on  $(x, x_l^+, s, \sigma_l(p))$ , where  $l$  is the index of the killed particle (the lowest one).

<sup>28</sup>Recall that if there are several indices  $m'$  s.t.  $x^{m'} = \min_i x^i$ , we arbitrarily chose  $m$  to be the smallest one.

<sup>29</sup>Just like [Definition A.3](#) of the  $b$ -BMP, this construction takes into account the fact that several particle might go below the curve simultaneously; and again, if several particles are at the lowest position simultaneously, we chose the one with smallest index.

<sup>30</sup>Since the particle is still paired, property (P1) ensures that we have  $n \geq N$ , so that  $n = N$  and  $n > N$  are the only two possibilities.

<sup>31</sup>This an arbitrary choice, we could have chosen it in another way.

Now let  $\tau = \theta^1 \wedge \theta^2 \wedge \zeta^2$  and

$$\xi = \mathbf{1}_{\{\tau=\theta^1\}}\nu^1 + \mathbf{1}_{\{\tau=\theta^2\}}\nu^2 + \mathbf{1}_{\{\tau=\zeta^2\}}\lambda^2 =: M^{\theta^1, \theta^2, \zeta^2}(\nu^1, \nu^2, \lambda^2).$$

Notice that  $\tilde{E}$  is stable under the renewal distribution  $\xi$ :

- $\xi$  preserves (P1): denote by  $y = (x, x^+, s, p)$  the position before applying the renewal, and  $\bar{y} = (\bar{x}, \bar{x}^+, \bar{s}, \bar{p})$  the position afterwards; looking back at the different cases,  $\bar{y}$  satisfies (P1) in each of them:
  - (ia)  $\bar{x}^+$  has  $n + 1$  particles and all coordinates of  $\bar{p}$  are smaller than  $n + 1$  (since  $n + 1$  is, and the other coordinates are smaller than  $n$ , thanks to (P1) for  $y$ ),
  - (ib)  $p = \partial$ ,
  - (ii)  $\bar{x}^+$  has  $n + 1$  particles and all coordinates of  $\bar{p}$  are smaller than  $n$  (thanks to (P1) for  $y$ ),
  - (iiia)  $\bar{x}^+$  has  $n - 1$  particles and the coordinates of  $\bar{p}$  are all smaller than  $\sigma_{p_i}(n) = n - 1$ ,
  - (iiib)  $\bar{x}^+$  has  $n - 1$  particles and the coordinates of  $\bar{p}$  are all smaller than  $\sigma_l(n) = n - 1$ ,
- $\xi$  also preserves (P2); let us check it on the same cases:
  - (ia) writing  $m$  the index of the lowest position in  $x$ , we have  $\bar{x}^j = x^j \leq x^{p_j, +} = \bar{x}^{p_j, +} = \bar{x}^{\bar{p}_j, +}$  for  $j \neq m$ , and  $\bar{x}^m = x^m \leq x^{p_i, +} = \bar{x}^{n+1, +} = \bar{x}^{\bar{p}_m, +}$ ,
  - (ib)  $p = \partial$ ,
  - (ii) neither  $p$  nor the position of paired particles are affected,
  - (iiia) we have chosen  $\sigma_{p_i}$  so that every  $x$  particle, except  $x^i$ , is paired with a particle at the same position as before; and we also have  $\bar{x}^i = x^i \leq x^{p_i, +} \leq x^{j, +} = \bar{x}^{\sigma_{p_i}(j), +} = \bar{x}^{\bar{p}_i, +}$ ,
  - (iiib) we have chosen  $\sigma_l$  so that every  $x$  particle is paired with a particle at the same position as before.

The coupling is now defined as the process  $\mathfrak{P}^\tau \mathcal{Y}$  on  $\mathfrak{P}W$  made by pasting copies of  $\mathcal{Y}$  at successive times  $\tau$  with the renewal distribution  $\xi$ .

### A.2.2 Proof of Proposition I.4

We can now prove Proposition I.4 through the following result on  $\mathfrak{P}^\tau \mathcal{Y}$ . In order to stick with the notation of elements of  $E$ , we write  $\mathfrak{P}^\tau \mathcal{X}$  its  $E_1$ -coordinate,  $(\mathfrak{P}^\tau \mathcal{X}^+, \mathfrak{P}^\tau \mathcal{T})$  its  $E_2$ -coordinate, and  $\mathfrak{P}^\tau \Pi$  its  $E_3$ -coordinate.

**Proposition A.4.** *The coordinate processes of  $\mathfrak{P}^\tau \mathcal{Y}$  have the following distributions:*

- (i)  $\mathfrak{P}^\tau \mathcal{X}$  is an  $N$ -BMP,
- (ii)  $\mathfrak{P}^\tau \mathcal{X}^+$  is a  $b$ -BMP.

Furthermore, if  $T_\partial$  is the hitting time of  $\partial$  for the  $E_3$ -coordinate of  $\mathfrak{P}^\tau \mathcal{Y}$ , then

- (iii)  $T_\partial$  is  $\mathfrak{P}^\tau \mathbb{P}_y^\xi$ -a.s. the first time when  $\mathfrak{P}^\tau \mathcal{X}^+$  has strictly less than  $N$  particles,
- (iv)  $\mathfrak{P}^\tau \mathcal{X}_t \preceq \mathfrak{P}^\tau \mathcal{X}_t^+$  for all  $t < T_\partial$ ,  $\mathfrak{P}^\tau \mathbb{P}_y^\xi$ -a.s.

**Remark A.5.** Looking at the definitions in [Section A.1.2](#), we can deduce that the different branching processes presented in that section have infinite lifetime. We would also have to show here that  $\mathfrak{P}^\tau \mathcal{Y}$  has infinite lifetime, in particular because we use [Proposition A.3](#). We sketch here the main arguments: the branching times add up to infinity since they are exponentially distributed (the argument is the same as the one we could use to prove that a BMP has infinite lifetime, using a sum of independent and exponentially distributed random times), and the death times cannot accumulate since there is always a finite number of particles alive (starting from  $N + n$  particles in the whole process, the  $(N + n + k)$ -th death time cannot happen before the  $k$ -th branching time). In the following, we can thus use [Proposition A.3](#).

*Proof.* [Figure A.5](#) presents the structure of the proof of (i) and (ii). We will use the commutativity of the pasting procedure ([Proposition A.2](#)) to follow the different paths represented by the different arrows and match the definitions of the processes given in [Section A.1.2](#) to the  $E_1$ - and  $E_2$ -coordinates of  $\mathfrak{P}^\tau \mathcal{Y}$ .

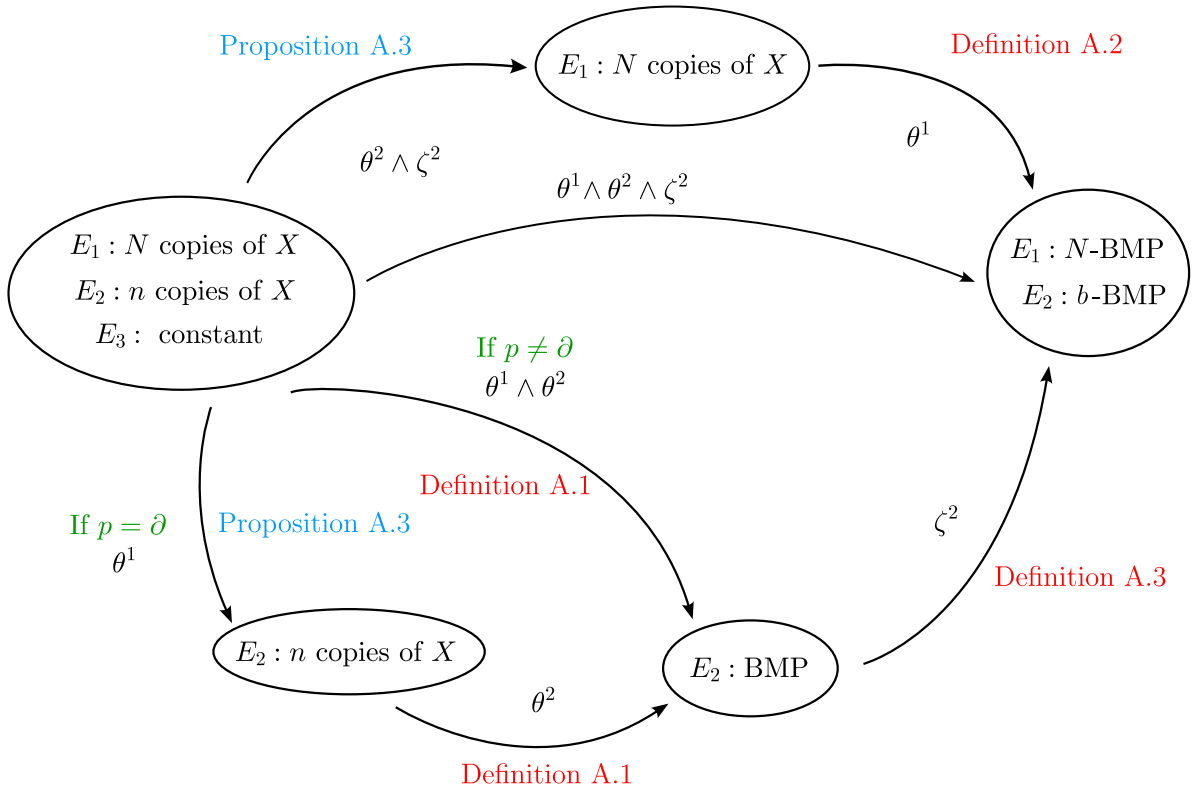


Figure A.5: Sketch of the proof of the distributions of  $\hat{\mathcal{X}}$  and  $\hat{\mathcal{X}}^+$

- (i) In order to prove that  $\mathfrak{P}^\tau \mathcal{X}$  is an  $N$ -BMP, we first use [Proposition A.3](#) with the base process  $\mathcal{Y}$  and the pasting at times  $\theta^2 \wedge \zeta^2$ : we notice that the renewal distributions  $\nu^2$  and  $\lambda^2$  described in [Section A.2.1.3](#)(ii) and (iii) do not affect the  $E_1$ -coordinate; according to [Proposition A.3](#), it means that the pasted  $E_1$ -coordinate process  $\mathfrak{P}^{\theta^2 \wedge \zeta^2} \mathcal{X}$  has under  $\mathfrak{P}^{\theta^2 \wedge \zeta^2} \mathbb{P}_y^{M^{\theta^2, \zeta^2}(\nu^2, \lambda^2)}$  the same distribution as  $\mathcal{X}$  under  $\mathbb{P}_y$ , that is (according to [Section A.2.1.1](#))  $N$  independent copies of  $X$ . We can then use [Definition A.2](#) to prove that, when pasting the pasted process again with the random times  $\mathfrak{P}^{\theta^2 \wedge \zeta^2} \theta^1$  on  $\mathfrak{P}W$ , we obtain an  $N$ -BMP. The key argument to prove that these random times are exponen-

tially distributed and independent from the pasted process are the memorylessness of the exponential distribution and the independence of  $\theta^1$  and  $\mathcal{X}$ .

(ii) Let us now prove that  $\mathfrak{P}^\tau \mathcal{X}^+$  is a  $b$ -BMP. Starting from  $y = (x, x^+, s, p) \in E$ , the proof depends on whether the processes are already in the independent regime or not, *i.e.* on whether  $p = \partial$ .

- If  $p \neq \partial$ , we use the pasting procedure on  $\mathcal{Y}$  at times  $\theta^1 \wedge \theta^2$ : according to their definition,  $\theta^1$  and  $\theta^2$  are independent exponentially distributed times, so that their respective rate add up and  $\theta^1 \wedge \theta^2$  is exponentially distributed with rate  $n$ , and independent from  $\mathcal{X}^+$ . Furthermore,  $\nu := M^{\theta^1, \theta^2}(\nu^1, \nu^2)$  adds a particle at the position of a particle randomly chosen in  $\mathcal{X}_{\theta^1 \wedge \theta^2}^+$ , so that this procedure matches [Definition A.1](#): the pasted  $E_2$ -process  $\mathfrak{P}^{\theta^1 \wedge \theta^2} \mathcal{X}^+$  is a BMP (plus a deterministic time coordinate) under  $\mathfrak{P}^{\theta^1 \wedge \theta^2} \mathbb{P}_y^\nu$ . Then, we paste again copies of this process at the random times  $\mathfrak{P}^{\theta^1 \wedge \theta^2} \zeta^2$ , and since  $\bar{\lambda}^2$  consists in killing the particles below the curve at the pasting time, according to [Definition A.3](#), the final  $E_2$ -process  $\mathfrak{P}^{\theta^1 \wedge \theta^2, \zeta^2} \mathcal{X}^+$  is a  $b$ -BMP, and thanks to [Proposition A.2](#), so is  $\mathfrak{P}^\tau \mathcal{X}^+$ .
- If  $p = \partial$ , we decompose the procedure into three layers. First, we paste at times  $\theta^1$ , and use the fact that the renewal distribution  $\nu^1$  described in [Section A.2.1.3](#)(i) does not affect the  $E_2$ -coordinate to deduce that the pasted  $E_2$ -process  $\mathfrak{P}^{\theta^1} \mathcal{X}^+$  is the same as the initial one, namely  $n$  independent copies of  $X$ . Then we paste copies of  $\mathfrak{P}^{\theta^1} \mathcal{Y}$  at the random times  $\mathfrak{P}^{\theta^1} \theta^2$ ; again thanks to the memorylessness of the exponential distribution and the independence of  $\theta^2$  and  $\mathcal{X}^+$ , these random times are exponentially distributed with same rate  $n$  (since  $p = \partial$ ) and independent from the pasted process, and  $\bar{\nu}^2$  then performs a uniform choice on the branching of one  $E_2$ -particle; this proves thanks to [Definition A.1](#) that the  $E_2$ -process  $\mathfrak{P}^{\theta^1, \theta^2} \mathcal{X}^+$  is a BMP under  $\mathfrak{P}^{\theta^1, \theta^2} \mathbb{P}_y^{\nu^1, \nu^2}$ . Then we finally paste a third time: we use the random times  $\mathfrak{P}^{\theta^1, \theta^2} \zeta^2 := \mathfrak{P}^{\mathfrak{P}^{\theta^1} \theta^2}[\mathfrak{P}^{\theta^1} \zeta^2]$  defined on  $\mathfrak{P}^{(2)}W$  and construct the process  $\mathfrak{P}^{\theta^1, \theta^2, \zeta^2} \mathcal{Y} := \mathfrak{P}^{\mathfrak{P}^{\theta^1, \theta^2} \zeta^2}[\mathfrak{P}^{\theta^1, \theta^2} \mathcal{Y}]$  on  $\mathfrak{P}^{(3)}W = \mathfrak{P}[\mathfrak{P}^{(2)}W]$ . As in the former case, since  $\bar{\lambda}^2$  performs the killing of the particles in the  $E_2$ -coordinate below the curve, the final  $E_2$ -coordinate process  $\mathfrak{P}^{\theta^1, \theta^2, \zeta^2} \mathcal{X}^+$  is a  $b$ -BMP, and applying [Proposition A.2](#) twice, we conclude that so is  $\mathfrak{P}^\tau \mathcal{X}^+$ .

(iii) We now want to show that  $T_\partial$  is the first time when  $\mathfrak{P}^\tau \mathcal{X}^+$  has less than  $N$  particles. For the final two points of the proof, recall properties (P1) and (P2) of the elements  $y$  in  $E'$  (see [Section A.2.1.1](#)), and also recall that  $\mathfrak{P}^\tau \mathcal{Y}$  takes values in  $E'$ .

Thanks to (P1), an element of  $E'$  can have less than  $N$  particles in the  $E_2$ -coordinate only if its  $E_3$ -coordinate is  $\partial$ . Between branching or death times,  $\mathfrak{P}^\tau \mathcal{Y}$  evolves as  $\mathcal{Y}$ , *i.e.* according to the semigroup  $(\mathcal{P}_t)_{t \geq 0}$ , so neither the number of particles nor the fact that its  $E_3$ -coordinate equals  $\partial$  changes. We thus need to prove that the first branching or death time at which  $p$  becomes  $\partial$  is also the first time when  $\mathfrak{P}^\tau \mathcal{X}^+$  has less than  $N$  particles. Looking at the definitions of the renewal distributions (see [Section A.2.1.3](#)), there is only one case with  $p \neq \partial$  that leads to hitting  $\partial$ , case (iiia), in which the  $E_2$ -coordinate has  $N$  particles before the renewal, and  $N - 1$  particles afterwards. Furthermore, all previous renewal times either add particles to  $\mathcal{X}^+$ , or remove some, but only in cases where there are at least  $N + 1$  particles after the selection, so that they do not lead to states with less than  $N$  particles, so indeed  $T_\partial$  is the first time when  $\mathfrak{P}^\tau \mathcal{X}^+$  has less than  $N$  particles.

(iv) As we already mentioned, the fact that we have a stochastic ordering between  $\mathfrak{P}^\tau \mathcal{X}$  and

$\mathfrak{P}^\tau \mathcal{X}^+$  until  $T_\partial$  comes from the fact that for  $y = (x, x^+, s, p) \in E$  such that  $p \neq \partial$ , we have  $x \preceq x^+$ .

□

*Proof of Proposition 1.4 for the upper stochastic barrier.* We can directly prove the result that we are interested in: that there exists a coupling of an  $N$ -BMP and its stochastic upper barrier respectively starting from  $N$  and  $N_\delta^+$  independent positions distributed according to  $\mu_0$ . This coupling is the restriction  $(\mathfrak{P}^\tau \mathcal{X}, \mathfrak{P}^\tau \mathcal{X}^+)$  of the process  $\mathfrak{P}^\tau \mathcal{Y}$  presented in Proposition A.4, and there just remains to define the distribution of the initial configuration.

In order to ensure that the  $N$ -BMP and the  $b$ -BMP start respectively with  $N$  and  $N_\delta^+ = \lceil N(1 + \delta) \rceil > N$  particles at time 0 and are stochastically ordered, we ask that the  $N$  first coordinates of the  $b$ -BMP are the same as those of the  $N$ -BMP, and we can thus couple these particles in ordered pairs. □

Finally, we can use the same kind of construction to define a coupling between the  $N$ -BMP and its stochastic lower barrier. We do not give the full details, since a lot of them are the same as in the construction of the upper barrier. This time we will couple each particle of the  $b$ -BMP until the first time when it has more than  $N$  particles. Hence we have to change the discrete coordinate that describes the links between particles in the state space: define  $E'_3 = \mathbb{N}^N \cup \{\partial\}$ , and  $\tilde{E}'$  the subset of  $E_1 \times E_2 \times E'_3$  of elements  $y = (x, (x^-, s), p) = (x, x^-, s, p)$  such that

(P1') if  $p = (p_1, \dots, p_N)$  and  $I_0 = \{i, p_i = 0\}$ , then  $\#I_0 = N - n$  and  $\{p_i, i \notin I_0\} = \llbracket 1, n \rrbracket$ , where  $n$  is the number of particles in  $x^-$ ,

(P2') if  $p = (p_1, \dots, p_N)$ , then  $x^{p_i, -} \leq x^i$  for any  $i$  such that  $p_i \neq 0$ .

We have again the same interpretation: if  $p_i \in \llbracket 1, n \rrbracket$ , then the  $i$ -th  $E_1$ -particle is paired with the  $p_i$ -th  $E_2$ -particle; if  $p_i = 0$ , then the  $i$ -th  $E_1$ -particle is unpaired; and if  $p = \partial$ , the  $E_1$ - and  $E_2$ -processes evolve independently. Furthermore, (P1') and (P2') ensure that while  $p$  is different from  $\partial$ , each particle of  $\mathcal{X}^-$  is paired with a particle in  $\mathcal{X}$ , and is below it, hence we have  $\mathcal{X}^- \preceq \mathcal{X}$ .

Now we paste copies of a process  $\mathcal{Y}' = (\mathcal{X}', \mathcal{X}'^-, \mathcal{T}', \Pi')$  evolving according to a semigroup built analogously as  $(\mathcal{P}_t)_{t \geq 0}$ , but this time on  $\tilde{E}'$ , and we will use the following times and renewal distribution. Start from a state  $y = (x, x^-, s, p)$  in  $\tilde{E}'$ .

(i) We first need an exponentially distributed time  $\theta'^2$  independent from  $\mathcal{Y}'$  to rule the branching of  $E_2$ -particles.

- (a) If  $p = (p_1, \dots, p_N)$ ,  $\theta'^2$  has distribution  $\mathcal{E}(n)$ , and the renewal distribution does the following: it chooses uniformly one of the coordinates of  $\mathcal{X}'_{\theta'^2}$  with index  $p_i$ , and adds it as an  $(n + 1)$ -th coordinate; it also performs the selection in the  $E_1$ -process by replacing the position of the lowest particle  $\mathcal{X}'_{\theta'^2}{}^m$  by  $\mathcal{X}'_{\theta'^2}{}^i$ , and coupling it with the newborn particle (*i.e.* it replaces  $p_m$  with  $n + 1$ ). If  $p_m = 0$  (*i.e.* the lowest  $E_1$ -particle was unpaired), there is no problem, but if  $p_m \in \llbracket 1, n \rrbracket$ , we also have to find a replacement in the pair involving  $\mathcal{X}'^{p_m, -}$  (to preserve the (P1') property): either  $n < N$ , and there is at least one  $i_0 \neq m$  such that  $p_{i_0} = 0$ , and we can redefine  $p_{i_0} = p_m$  (if there are several such indices, we arbitrarily take the smaller one); then (P2') is also preserved because  $\mathcal{X}'^{p_m, -} \leq \mathcal{X}'^m \leq \mathcal{X}'^{i_0}$ ; or  $N = n$ , there is no available particle and  $p$  becomes  $\partial$  and in the meanwhile  $\mathcal{X}'^-$  gets more than  $N$  particles.
- (b) If  $p = \partial$ ,  $\mathcal{X}'$  and  $\mathcal{X}'^-$  are independent, and  $\theta'^2$  also has distribution  $\mathcal{E}(n)$ , but only induces the branching of an  $E_2$ -particle:  $p$  remains  $\partial$ ,  $\mathcal{X}'_{\theta'^2}$  and  $\mathcal{T}'_{\theta'^2}$  are unchanged, and we add one particle to  $\mathcal{X}'_{\theta'^2}$  at a position chosen uniformly among its coordinates.



- (ii) We also use another exponentially distributed time  $\theta'^1$  for the branching of unpaired  $E_1$ -particles.
  - (a) If  $p = (p_1, \dots, p_N)$ ,  $\theta'^1$  has distribution  $\mathcal{E}(N - n)$ , and the renewal distributions replaces the lowest coordinate of  $\mathcal{X}'_{\theta'^1}$  by one of the coordinates of an unpaired particle in  $\mathcal{X}'_{\theta'^1}$ , chosen uniformly. We do not modify the  $E'_3$  coordinate, which means that if there was a particle paired with the lowest one, it is now paired with the newborn one (and (P2') is preserved).
  - (b) If  $p = \partial$ ,  $\mathcal{X}'$  and  $\mathcal{X}'^-$  are independent,  $\theta'^1$  has distribution  $\mathcal{E}(N)$  and the renewal distribution replaces the lowest coordinate of  $\mathcal{X}'_{\theta'^1}$  by one of its coordinates, chosen uniformly, and nothing else is modified.
- (iii) Finally we also use the first death time  $\zeta'^2$  of an  $E_2$ -particle (with a definition analogous to the former  $\zeta'^2$  used in the coupling with the upper barrier). The renewal distribution then removes the corresponding particle  $\mathcal{X}'^{m,-}$  in  $\mathcal{X}'_{\zeta'^2}$ , and if  $p \neq \partial$ , there was necessarily an index  $i$  such that  $p_i = m$  (thanks to (P1')), and we modify  $p_i$  to be 0 (the  $i$ -th  $E_1$ -particle gets unpaired).

Now we can use the same arguments as in the proof of [Proposition A.4](#) to prove that if we paste copies of  $\mathcal{Y}'$  at the minimum of  $\theta'^1, \theta'^2$  and  $\zeta'^2$ , we indeed build a coupling of an  $N$ -BMP and its lower barrier with stochastic ordering until the first time when the  $b$ -BMP has more than  $N$  particles. Finally, as for the coupling with the upper barrier, we can start from a position with  $N_{\delta}^- = \lfloor N(1 - \delta) \rfloor < N$  paired particles in common for  $\mathcal{X}$  and  $\mathcal{X}'^-$ , independent and distributed according to  $\mu_0$ , and add  $N - N_{\delta}^-$  others to complete  $\mathcal{X}$ .



## Appendix B

# Some results on the geometric distribution

In this last section, we show some bounds on unlikely events regarding geometric distributions that will be used in several places in the proofs.

**Lemma B.1.** *Let  $G_1, \dots, G_n$  be i.i.d. random variables with geometric distribution  $\mathcal{G}(p)$ ,  $p \in (0, 1)$ , and fix  $q = 1 - p$  and  $c = -\ln(q) > 0$ . Then*

(i) *for  $K > 0$ , we have  $\mathbb{P}(G_1 > K) = e^{-cK}$ ,*

(ii) *for  $K > 0$ , we have  $\mathbb{E}(G_1 \mathbf{1}_{\{G_1 > K\}}) \leq (K + 2) \frac{e^{-cK}}{p}$ ,*

(iii) *the Legendre transform of the geometric distribution  $\mathcal{G}(p)$  is given, for  $x \in (1, +\infty)$ , by*

$$\Lambda_p^*(x) = (x - 1) \ln \left( \frac{x - 1}{xq} \right) - \ln(px) \in [0, +\infty), \quad (\text{B.1})$$

*and in particular, we have the following expansion for the Legendre transform  $\Lambda_\theta^*$  when  $p = e^{-\theta}$*

$$\Lambda_{e^{-\theta}}^* \left( \frac{1}{1 - \frac{\theta}{2}} \right) = \frac{1 - \ln(2)}{2} \theta + o(\theta), \quad (\text{B.2})$$

(iv) *for  $x \in (1, \frac{1}{p})$  and  $n \geq 1$*

$$\mathbb{P} \left( \frac{1}{n} \sum_{k=1}^n G_k < x \right) \leq e^{-n\Lambda_p^*(x)}.$$

*Proof.* (i) is classical, and we can compute the exact value for (ii): we have

$$\mathbb{E}(G_1 \mathbf{1}_{\{G_1 > K\}}) = \sum_{k=K+1}^{+\infty} k p q^{k-1}.$$

Set for  $x < 1$

$$g(x) = \sum_{k=K+1}^{+\infty} p q^{k-1} x^k = p q^K \frac{x^{K+1}}{1 - qx},$$

so that

$$\sum_{k=K+1}^{+\infty} k p q^{k-1} x^{k-1} = g'(x) = p q^K \frac{(K+1)x^K(1 - qx) + qx^{K+1}}{(1 - qx)^2}.$$

Evaluating in  $x = 1$ , we obtain

$$\mathbb{E}(G_1 \mathbf{1}_{\{G_1 > K\}}) = \frac{q^K((K+1)p+q)}{p} \leq (K+2) \frac{e^{-cK}}{p}.$$

Let us carry out the computation of the Legendre transform of the geometric distribution: we have for  $a < -\ln(q)$

$$\ln(\mathbb{E}(e^{aG_1})) = \sum_{k=1}^{+\infty} e^{ak} q^{k-1} p = \frac{p e^a}{1 - q e^a},$$

so we can compute for  $x$  in  $(1, +\infty)$  the derivative of  $a \mapsto ax - \ln(\mathbb{E}(e^{aG_1}))$ , which vanishes when  $x - \frac{1}{1 - q e^a}$  is zero, *i.e.* for  $a^* = \ln\left(\frac{x-1}{xq}\right)$ . In the end, we have as claimed

$$\Lambda_p^*(x) = a^* x - \ln(\mathbb{E}(e^{a^* G_1})) = (x-1) \ln\left(\frac{x-1}{xq}\right) - \ln(px).$$

The large deviations estimate (iv) then follows directly from Cramér's theorem (see e.g. [\[DZ10\]](#)).

There only remains to prove the expansion [\[B.2\]](#), which we use in the proof of [Theorem \[I.2\]](#). Take  $x = \frac{1}{1 - \frac{\theta}{2}}$  and  $p = e^{-\theta}$  in [\[B.1\]](#). We have

$$\ln(x e^{-\theta}) = \ln(x) - \theta = -\frac{\theta}{2} + o(\theta).$$

Next  $\frac{x-1}{x} = \frac{\theta}{2}$ , so

$$\ln\left(\frac{x-1}{x(1-e^{-\theta})}\right) = \ln\left(\frac{\theta}{2\theta - \theta^2 + o(\theta^2)}\right) = -\ln(2) + \frac{\theta}{2} + o(\theta),$$

and then

$$(x-1) \ln\left(\frac{x-1}{x(1-e^{-\theta})}\right) = -\frac{\ln(2)}{2} \theta + o(\theta).$$

In the end, we have indeed

$$\Lambda_{e^{-\theta}}^*\left(\frac{1}{1 - \frac{\theta}{2}}\right) = \frac{1 - \ln(2)}{2} \theta + o(\theta). \quad \text{[B.2]}$$

□

## Appendix C

# Additional results related to Chapter II (comonotone flow)

### C.1 About Theorems 4 and 5 in [KKO77]

In this section, we explain why we think an element is missing in the proof of Theorem 4 in [KKO77], and hence of Theorem 5 in [KKO77], which is a corollary.

#### C.1.1 About Theorem 4

Theorem 4 in [KKO77] considers a partially ordered Polish space  $(E, \preceq)$ , two families of  $E$ -valued random variables  $(X_t)_{t \in \mathbb{R}_+}$  and  $(Y_t)_{t \in \mathbb{R}_+}$  such that  $t \mapsto X_t$  and  $t \mapsto Y_t$  are (random) càdlàg paths from  $\mathbb{R}_+$  to  $E$ , and, assumes that, for all  $n \geq 2$ ,  $x_1, \dots, x_{n-1} \in \mathbb{R}$ ,  $y_1, \dots, y_{n-1} \in \mathbb{R}$  such that  $x_i \preceq y_i$  for  $i = 1, \dots, n-1$ , and **ordered time indices**  $0 \leq t_1 < \dots < t_n \in \mathbb{R}$ , one has

$$\mathbb{P}(X_{t_n} \in \cdot | X_{t_1} = x_1, \dots, X_{t_{n-1}} = x_{n-1}) \leq_{\text{st}} \mathbb{P}(Y_{t_n} \in \cdot | Y_{t_1} = y_1, \dots, Y_{t_{n-1}} = y_{n-1}), \quad (\text{C.1})$$

where  $\leq_{\text{st}}$  denotes stochastic domination between probability measures on  $(E, \preceq)$  (see [KKO77]).

Assuming moreover that  $\mathbb{P}(X_0 \in \cdot) \leq_{\text{st}} \mathbb{P}(Y_0 \in \cdot)$ , the conclusion of Theorem 4 is that one can define a family of pairs of  $E$ -valued random variables  $(\tilde{X}_t, \tilde{Y}_t)_{t \geq 0}$  on the same probability space, in such a way that  $t \mapsto \tilde{X}_t$  and  $t \mapsto \tilde{Y}_t$  are (random) càdlàg paths from  $\mathbb{R}_+$  to  $E$ , that almost surely for all  $t$ ,  $\tilde{X}_t \preceq \tilde{Y}_t$ , that  $(\tilde{X}_t)_{t \geq 0}$  has the same (joint) distribution as  $(X_t)_{t \geq 0}$ , and that  $(\tilde{Y}_t)_{t \geq 0}$  has the same (joint) distribution as  $(Y_t)_{t \geq 0}$ .

We now consider the proof of Theorem 4. Given an **increasing sequence**  $0 = t_1 < t_2 < \dots$ , one can use [(C.1)] and Strassen's theorem to define inductively a sequence of pairs of  $E$ -valued random variables  $(\tilde{X}_{t_n}, \tilde{Y}_{t_n})_{n \geq 1}$  in such a way that, almost surely, for all  $n \geq 1$ ,  $\tilde{X}_{t_n} \preceq \tilde{Y}_{t_n}$ , and such that  $(\tilde{X}_{t_1}, \dots, \tilde{X}_{t_n})$  has the same distribution as  $(X_{t_1}, \dots, X_{t_n})$ , and  $(\tilde{Y}_{t_1}, \dots, \tilde{Y}_{t_n})$  has the same distribution as  $(Y_{t_1}, \dots, Y_{t_n})$ . (This mirrors the argument used to prove Theorem 2 in [KKO77].)

The problem is the claim (without an explanation), in the proof of Theorem 4, that this approach can be used to produce a sequence of random variables  $(\tilde{X}_{t_n}, \tilde{Y}_{t_n})_{n \geq 1}$  with the above properties, where  $(t_n)_{n \geq 1}$  is an enumeration of a countable dense subset of  $\mathbb{R}_+$ , since clearly such an enumeration **cannot be done using an increasing sequence**. Thus, to make the proof of Theorem 4 work, it seems that the following strengthening of Assumption [(C.1)] would be needed, allowing for unordered sequences of time indices: for every  $n \geq 2$ ,  $x_1, \dots, x_{n-1} \in \mathbb{R}$ ,  $y_1, \dots, y_{n-1} \in \mathbb{R}$  such that  $x_i \preceq y_i$  for  $i = 1, \dots, n-1$ , and **pairwise distinct time indices**

$t_1, \dots, t_n \in \mathbb{R}_+$ , one has

$$\mathbb{P}(X_{t_n} \in \cdot | X_{t_1} = x_1, \dots, X_{t_{n-1}} = x_{n-1}) \leq_{\text{st}} \mathbb{P}(Y_{t_n} \in \cdot | Y_{t_1} = y_1, \dots, Y_{t_{n-1}} = y_{n-1}). \quad (\text{C.2})$$

### C.1.2 About Theorem 5

Theorem 5 in [KKO77] assumes that  $(X_t)_{t \in \mathbb{R}_+}$  and  $(Y_t)_{t \in \mathbb{R}_+}$  are (possibly time-inhomogeneous) Markov processes, and does not directly assume (C.1), but instead the following stochastic monotonicity condition: for all  $s, t \geq 0$ , and all  $x, y \in E$  such that  $x \preceq y$ ,

$$\mathbb{P}(X_{t+s} \in \cdot | X_t = x) \leq_{\text{st}} \mathbb{P}(Y_{t+s} \in \cdot | Y_t = y), \quad (\text{C.3})$$

with the same conclusion as Theorem 4. The proof of Theorem 5 consists in observing that, using the Markov property, (C.3) implies (C.1), then invoking Theorem 4. However, in view of our previous remark, this approach does not seem to lead to a complete proof of Theorem 5 unless one can prove that (C.3) implies (C.2). Unfortunately, such an implication is not true in general when  $(t_k)_{1 \leq k \leq n}$  is not increasing, as we show in the following counterexample.

Take  $n = 3, t_1 = 0, t_2 = 2, t_3 = 1$  and  $E = \{a, b, c\}$ , where  $a \preceq b \preceq c$ , and where  $X$  and  $Y$  are two versions of the same continuous-time Markov chain on  $E$  with distinct starting points. The infinitesimal generator of the chain is prescribed by the  $Q$ -matrix (see e.g. [And91])

$$L = (L_{xy})_{x \in E, y \in E} = \begin{matrix} & \begin{matrix} a & b & c \end{matrix} \\ \begin{matrix} a \\ b \\ c \end{matrix} & \begin{pmatrix} -2.5 & 1.75 & 0.75 \\ 1.5 & -2.5 & 1 \\ 0.5 & 0 & -0.5 \end{pmatrix} \end{matrix}$$

and defines a Markov semigroup on  $E$ , given, for all  $t \geq 0$ , by the transition matrix  $P_t = \exp(tL)$ . To check that  $L$  indeed defines a stochastically monotone Markov semigroup on  $E$ , we check the condition stated in [And91] (Theorem 3.4 page 249, attributed to [Kir76]), which in our setting reduces to the two conditions  $L_{ac} \leq L_{bc}$  and  $L_{ba} \geq L_{ca}$ , both visibly satisfied. We now check numerically that

$$\mathbb{P}(X_1 \in \cdot | X_0 = a, X_2 = a) \leq_{\text{st}} \mathbb{P}(Y_1 \in \cdot | Y_0 = b, Y_2 = b).$$

Denoting  $P = \exp(L)$ , we have that

$$\mathbb{P}(X_1 \preceq a | X_0 = a, X_2 = a) = \mathbb{P}(X_1 = a | X_0 = a, X_2 = a) = \frac{P_{aa}P_{aa}}{P_{aa}P_{aa} + P_{ab}P_{ba} + P_{ac}P_{ca}},$$

and

$$\mathbb{P}(Y_1 \preceq a | Y_0 = b, Y_2 = b) = \mathbb{P}(Y_1 = a | Y_0 = b, Y_2 = b) = \frac{P_{ba}P_{ab}}{P_{ba}P_{ab} + P_{bb}P_{bb} + P_{bc}P_{cb}}.$$

Numerically computing  $P$  with the SciPy open-source software<sup>1</sup>, we get, rounding to 3 decimal places, that  $\mathbb{P}(X_1 \preceq a | X_0 = a, X_2 = a) \approx 0.362$  and  $\mathbb{P}(Y_1 \preceq a | Y_0 = b, Y_2 = b) \approx 0.374$ , while the stochastic domination  $\mathbb{P}(X_1 \in \cdot | X_0 = a, X_2 = a) \leq_{\text{st}} \mathbb{P}(Y_1 \in \cdot | Y_0 = b, Y_2 = b)$  would imply the inequality  $\mathbb{P}(Y_1 \preceq a | Y_0 = b, Y_2 = b) \leq \mathbb{P}(X_1 \preceq a | X_0 = a, X_2 = a)$ .

To sum up, an element seems to be missing in the proof of Theorem 5. We observe that, in the specific case where  $E = \mathbb{R}$  and where  $(X_t)_{t \geq 0}$  and  $(Y_t)_{t \geq 0}$  are two versions of the same Feller process with distinct starting points, [Theorem III.1] in the present paper can be used to deduce the conclusion of Theorem 5.

<sup>1</sup>Specifically, we used `scipy.linalg.expm` to compute the matrix exponential.

## C.2 Miscellaneous lemmas

**Lemma C.1.** *If  $K^{(n)}$  and  $L^{(n)}$  are order-preserving kernels on  $\mathbb{R}^n$ , so is their composition  $K^{(n)}L^{(n)}$ .*

*Proof.* By definition, for all  $\mathbf{x} \in \mathbb{R}^n$ , we have that  $K^{(n)}L^{(n)}(\mathbf{x}, \mathbb{R}_{\mathbf{x}}^n) = \int_{\mathbb{R}^n} L^{(n)}(\mathbf{y}, \mathbb{R}_{\mathbf{x}}^n) dK^{(n)}(\mathbf{x}, \mathbf{y})$ . Since  $K^{(n)}$  is order-preserving, we have that  $K^{(n)}(\mathbf{x}, \mathbb{R}_{\mathbf{x}}^n) = 1$ , so the previous integral can be rewritten as  $\int_{\mathbb{R}_{\mathbf{x}}^n} L^{(n)}(\mathbf{y}, \mathbb{R}_{\mathbf{x}}^n) dK^{(n)}(\mathbf{x}, \mathbf{y})$ . For all  $\mathbf{y} \in \mathbb{R}_{\mathbf{x}}^n$ , the definition shows that  $\mathbb{R}_{\mathbf{y}}^n \subset \mathbb{R}_{\mathbf{x}}^n$ , so that, since  $L^{(n)}$  is order-preserving, we have  $L^{(n)}(\mathbf{y}, \mathbb{R}_{\mathbf{x}}^n) = 1$ , and we have proved that  $K^{(n)}L^{(n)}(\mathbf{x}, \mathbb{R}_{\mathbf{x}}^n) = 1$ .  $\square$

**Lemma C.2.** *If  $(K^{(n)})_{n \geq 2}$  and  $(L^{(n)})_{n \geq 2}$  are consistent extensions, respectively of  $K$  and  $L$ , then  $(K^{(n)}L^{(n)})_{n \geq 2}$  is a consistent extension of  $KL$ .*

*Proof.* Let  $K^{(1)} = K$  and  $L^{(1)} = L$ . Given integers  $1 \leq k \leq n$ ,  $i_1, \dots, i_k \in \llbracket 1, n \rrbracket$ ,  $\mathbf{x} \in S^n$  and measurable subset  $B$  of  $S^k$ , we have by definition

$$K^{(n)}L^{(n)}(\mathbf{x}, (\pi_{i_1, \dots, i_k}^n)^{-1}(B)) = \int_{S^n} L^{(n)}(\mathbf{y}, (\pi_{i_1, \dots, i_k}^n)^{-1}(B)) dK^{(n)}(\mathbf{x}, \mathbf{y}).$$

Since  $(L^{(n)})_{n \geq 2}$  is a consistent extension of  $L$ , we have that

$$L^{(n)}(\mathbf{y}, (\pi_{i_1, \dots, i_k}^n)^{-1}(B)) = L^{(k)}(\pi_{i_1, \dots, i_k}^n(\mathbf{y}), B),$$

so that

$$K^{(n)}L^{(n)}(\mathbf{x}, (\pi_{i_1, \dots, i_k}^n)^{-1}(B)) = \int_{S^n} L^{(k)}(\pi_{i_1, \dots, i_k}^n(\mathbf{y}), B) dK^{(n)}(\mathbf{x}, \mathbf{y}). \quad (\text{C.4})$$

Denoting by  $(\pi_{i_1, \dots, i_k}^n)_* K^{(n)}(\mathbf{x}, \cdot)$  the image probability measure on  $S^k$  defined by

$$\left[ (\pi_{i_1, \dots, i_k}^n)_* K^{(n)}(\mathbf{x}, \cdot) \right] (A) = K^{(n)}(\mathbf{x}, (\pi_{i_1, \dots, i_k}^n)^{-1}(A)),$$

(C.4) rewrites as

$$K^{(n)}L^{(n)}(\mathbf{x}, (\pi_{i_1, \dots, i_k}^n)^{-1}(B)) = \int_{S^k} L^{(k)}(\mathbf{z}, B) d \left[ (\pi_{i_1, \dots, i_k}^n)_* K^{(n)}(\mathbf{x}, \cdot) \right] (\mathbf{z}).$$

On the other hand, since  $(K^{(n)})_{n \geq 2}$  is a consistent extension of  $K$ , we have that

$$\left[ (\pi_{i_1, \dots, i_k}^n)_* K^{(n)}(\mathbf{x}, \cdot) \right] (A) = K^{(k)}(\pi_{i_1, \dots, i_k}^n(\mathbf{x}), A),$$

so that

$$K^{(n)}L^{(n)}(\mathbf{x}, (\pi_{i_1, \dots, i_k}^n)^{-1}(B)) = \int_{S^k} L^{(k)}(\mathbf{z}, B) dK^{(k)}(\pi_{i_1, \dots, i_k}^n(\mathbf{x}), \mathbf{z}),$$

and thus

$$K^{(n)}L^{(n)}(\mathbf{x}, (\pi_{i_1, \dots, i_k}^n)^{-1}(B)) = (K^{(k)}L^{(k)})(\pi_{i_1, \dots, i_k}^n(\mathbf{x}), B),$$

which proves the conclusion of the Lemma.  $\square$

---

<sup>2</sup>If  $\mathbf{x} = (x_1, \dots, x_n)$ ,  $\mathbf{y} = (y_1, \dots, y_n)$ ,  $\mathbf{z} = (z_1, \dots, z_n)$  are such that  $\mathbf{y} \in \mathbb{R}_{\mathbf{x}}^n$  and  $\mathbf{z} \in \mathbb{R}_{\mathbf{y}}^n$ , we have that, for all  $1 \leq i, j \leq n$ ,  $x_i \leq x_j \Rightarrow y_i \leq y_j$  (since  $\mathbf{y} \in \mathbb{R}_{\mathbf{x}}^n$ ), hence  $x_i \leq x_j \Rightarrow z_i \leq z_j$  (since  $\mathbf{z} \in \mathbb{R}_{\mathbf{y}}^n$ ).

### C.3 Numerical illustrations

In this section, we show some numerical simulations to illustrate our construction, when the underlying process is a square-root diffusion as defined by the following s.d.e.<sup>3</sup>

$$dX_t = a(b - X_t)dt + \sigma\sqrt{X_t}dW_t, \quad (\text{C.5})$$

where  $(W_t)_{t \geq 0}$  is a standard one-dimensional brownian motion. Provided that the condition  $2ab \geq \sigma^2$  is satisfied, (C.5) defines an explicit Feller semigroup on  $]0, +\infty[$ : for  $t > 0$ ,  $P_t(x, \cdot)$  is the distribution of  $Z/(2 * c)$ , where  $c = \frac{2a}{(1 - e^{-at})\sigma^2}$ , and  $Z$  follows a non-central chi-square distribution with  $\frac{4ab}{\sigma^2}$  degrees of freedom and non-centrality parameter  $2cxe^{-at}$ . To stick to the framework of the present paper, that is, Feller processes on  $\mathbb{R}$ , and not on  $]0, +\infty[$ , one may consider the semigroup associated with the process  $(\log(X_t))_{t \geq 0}$  instead, which leads to identical results up to applying a log change of scale. For the sake of simplicity, we keep working with the semigroup associated to  $(X_t)_{t \geq 0}$  in the following illustrations.

We work with the set of parameters  $a = 3$ ,  $b = 2$  and  $\sigma^2 = 8$ . The following figures depict the values of  $N = 5000$  simulated i.i.d. pairs distributed according to  $Q_t^{(2),m}((x_1, x_2), \cdot)$ , for  $x_1 = 0.5$ ,  $x_2 = 2$ ,  $t = 0.5$ , and  $m = 1, \dots, 6$ . Figure C.1 displays each pair  $(x, y)$  as a point with coordinates  $(x, y)$ , while Figure C.2 uses the point with coordinates  $(x, y - x)$  instead. Simulations were done using the R open-source software.

For  $m = 1$ ,  $Q_t^{(2),m}((x_1, x_2), \cdot)$  is just the classical monotonic coupling, which is the reason why the points are lying on a curve. As  $m$  gets larger, the array of points gets more spread out, with barely distinguishable differences between the few last consecutive graphs.

---

<sup>3</sup>This class of processes has been studied by Feller [Fel51], and gained popularity in financial mathematics as a model for the dynamics of interest rates [CIR85].



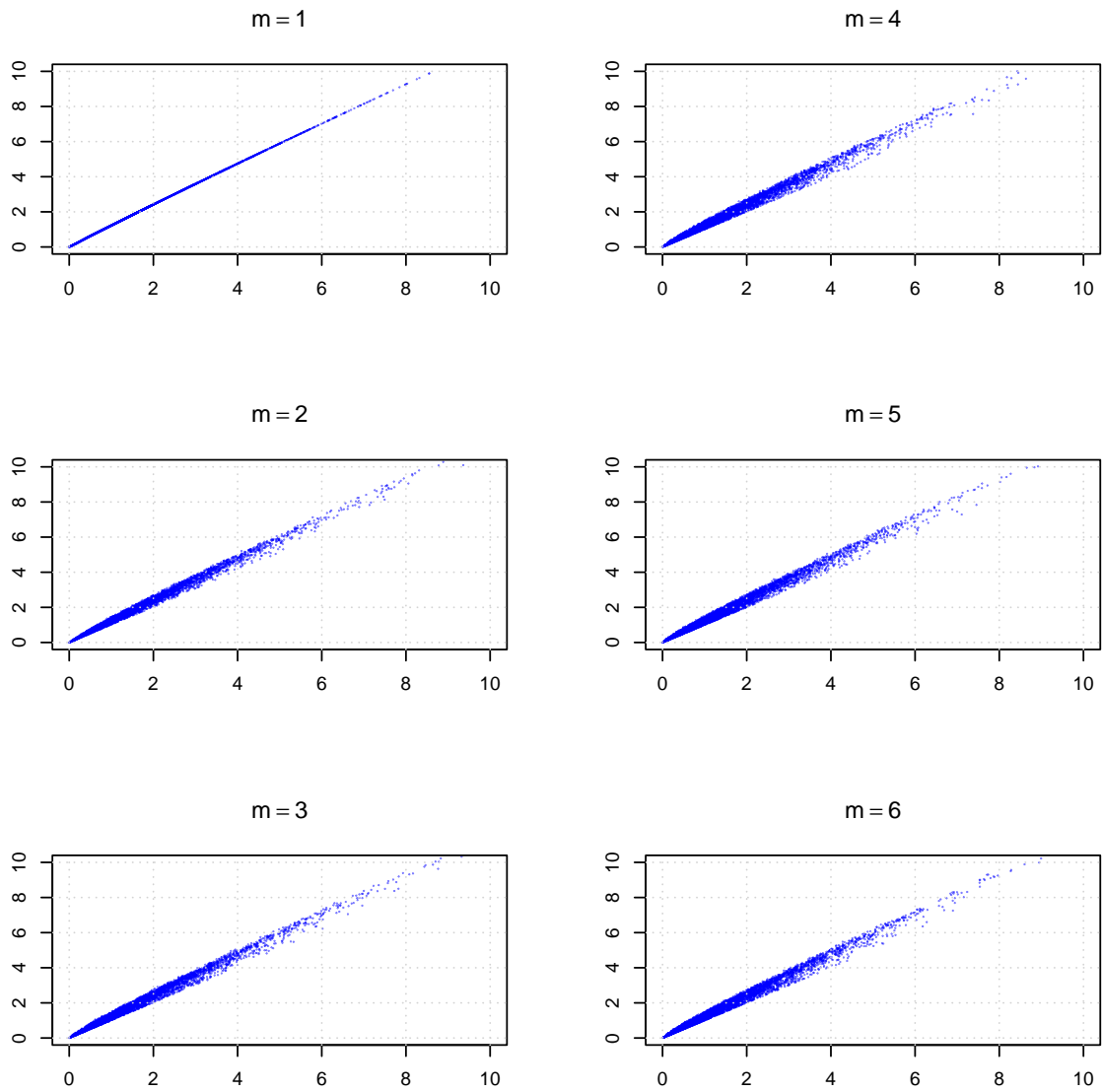


Figure C.1: 5000 simulated pairs following the distribution  $Q_t^{(2),m}((x_1, x_2), \cdot)$ .

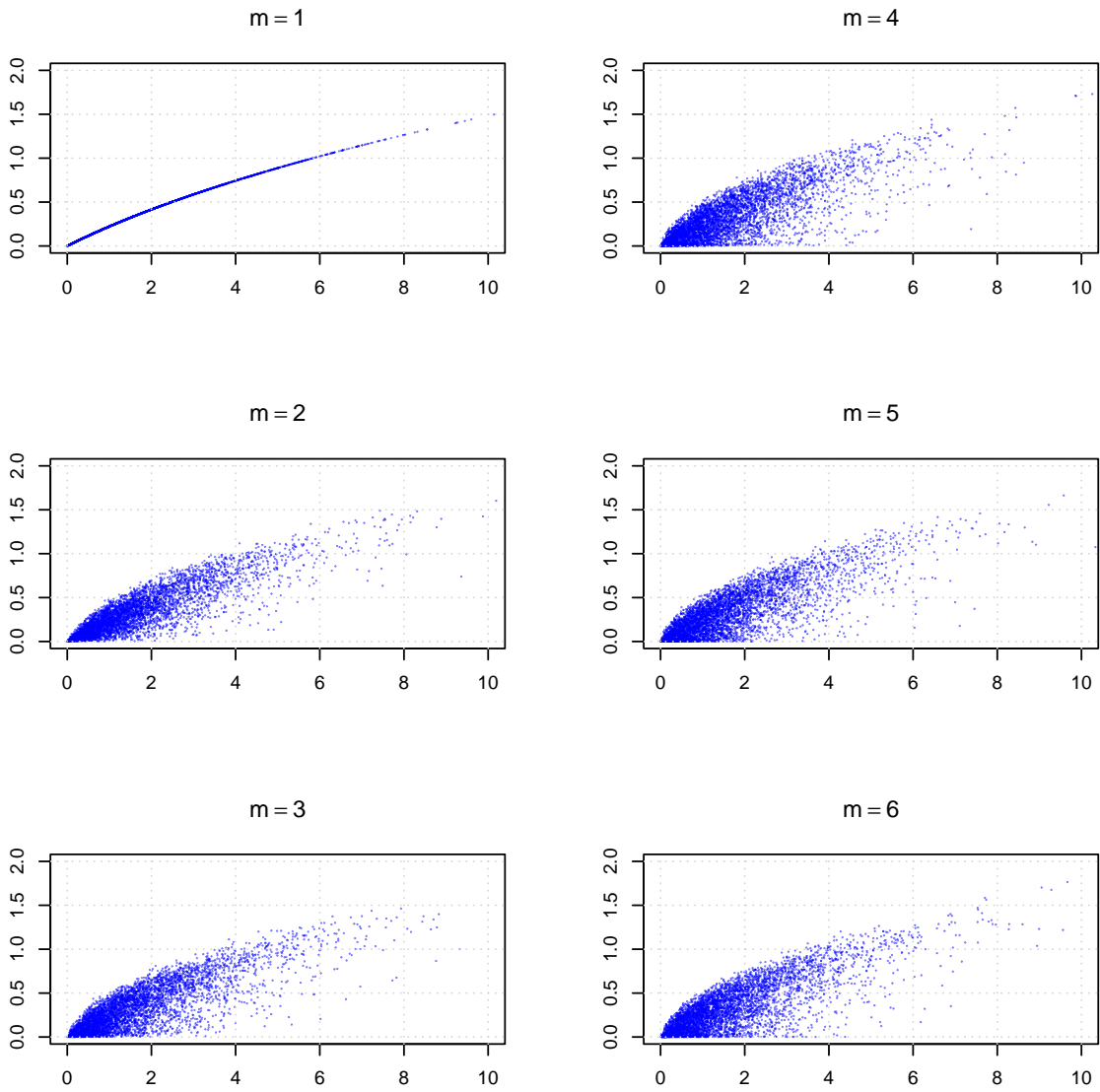


Figure C.2: Transformation  $(x, y) \mapsto (x, y - x)$  applied to the graphs in Figure C.1.



Since  $A$  is constant, we have  $\Phi(t, s) = e^{(t-s)A}$ . Similarly, we define  $\bar{\Phi}(t, s) = e^{(t-s)(A+\bar{A})}$ . We also set

$$\begin{cases} M(t, s) = \int_s^t \Phi(t, \tau) B B^\top \Phi(t, \tau)^\top d\tau, & 0 \leq s \leq t \leq T, \\ \bar{M}(t, s) = \int_s^t \bar{\Phi}(t, \tau) B B^\top \bar{\Phi}(t, \tau)^\top d\tau, & 0 \leq s \leq t \leq T, \end{cases} \quad (\text{D.3})$$

and denote for brevity  $\Phi_{T,0} = \Phi(T, 0)$ ,  $\bar{\Phi}_{T,0} = \bar{\Phi}(T, 0)$ ,  $M_{T,0} = M(T, 0)$ , and  $\bar{M}_{T,0} = \bar{M}(T, 0)$ . Notice that if  $B B^\top$  commutes with  $\Phi$  and  $\bar{\Phi}$ , and if the matrices  $A + A^\top$  and  $A + \bar{A} + A^\top + \bar{A}^\top$  are non-singular, then we have the following closed form expressions for  $M(t, s)$  and  $\bar{M}(t, s)$ , with  $0 \leq s \leq t \leq T$ :

$$\begin{aligned} M(t, s) &= B B^\top \int_s^t e^{A+A^\top}(t-\tau) d\tau = B B^\top [e^{(A+A^\top)t} - e^{(A+A^\top)s}] (A + A^\top)^{-1} \\ \bar{M}(t, s) &= B B^\top \int_s^t e^{A+A^\top+\bar{A}+\bar{A}^\top}(t-\tau) d\tau \\ &= B B^\top [e^{(A+A^\top+\bar{A}+\bar{A}^\top)t} - e^{(A+A^\top+\bar{A}+\bar{A}^\top)s}] (A + A^\top + \bar{A} + \bar{A}^\top)^{-1}. \end{aligned}$$

Now,  $(\Pi_t)_{t \in [0, T]}$  is the solution to the following Riccati ODE, which is independent from the other variables:

$$\begin{cases} \dot{\Pi}_t = -A^\top \Pi_t - \Pi_t A + \Pi_t B B^\top \Pi_t \\ \Pi_0 = \Sigma_0^{-1/2} \left[ \frac{I}{2} + \Sigma_0^{1/2} \Phi_{T,0}^\top M_{T,0}^{-1} \Phi_{T,0} \Sigma_0^{1/2} - \left( \frac{I}{4} + \Sigma_0^{1/2} \Phi_{T,0}^\top M_{T,0}^{-1} \Sigma_T M_{T,0}^{-1} \Phi_{T,0} \Sigma_0^{1/2} \right)^{1/2} \right] \Sigma_0^{-1/2}. \end{cases} \quad (\text{D.4})$$

Under some conditions, Riccati equations admit explicit solutions in dimension 1, see e.g. page 110 in [CD18]. More generally, we can solve (D.4) using a forward time-marching method.

We can then show that

$$n_t = \Pi_t \bar{\Phi}(t, T) \bar{M}(T, t) \bar{M}_{T,0}^{-1} \bar{\Phi}_{T,0} z_0 + \Pi_t \bar{M}(T, 0) \bar{\Phi}(T, t)^\top \bar{M}_{T,0}^{-1} z_T + \bar{\Phi}(T, t)^\top \bar{M}_{T,0}^{-1} (z_T - \bar{\Phi}_{T,0} z_0), \quad (\text{D.5})$$

where  $(z_t)_{t \in [0, T]}$  is defined by

$$z_t = \bar{\Phi}(T, t)^\top \bar{M}_{T,0}^{-1} (\bar{x}_T - \bar{\Phi}_{T,0} \bar{x}_0). \quad (\text{D.6})$$

Now, up to the computation of  $\Pi$ , we have an explicit formula for  $n$  and we can obtain the mean  $\bar{\mu}_t$  from (D.1) and then the covariance matrix  $\Sigma_t$  from (D.2). The optimal mean field distribution at time  $t$  is the Gaussian distribution with mean  $\bar{\mu}_t$  and variance  $\Sigma_t$ .

## D.2 Computation of $\mathcal{G}$

In this section, we discuss some of the issues arising when computing  $\mathcal{G}$  in practice, as well as our solutions. As the reader may notice, the losses defined in (III.25) and (III.28) depend on

the exact form of functionals  $\mathcal{F}$  and  $\mathcal{G}$ . As  $\mathcal{F}$  is already defined in an explicit, easy-to-compute form, we are left with the problem of figuring out a good approach to compute  $\mathcal{G}$ . Recalling the definition of  $\mathcal{G}$ , we have the following observation,

$$\begin{aligned}\mathcal{G}(\mathbf{a}, \mathbf{b}) &= - \inf_{m \geq 0} \int_{\mathcal{Q}_T} m(t, x) \left( \mathbf{a}(t, x) - H(x, m(t, x), \mathbf{b}(t, x)) \right) dx dt \\ &= - \int_{\mathcal{Q}_T} \inf_{m \geq 0} \left[ m(\mathbf{a}(t, x) - H(x, m, \mathbf{b}(t, x))) \right] dx dt\end{aligned}\tag{D.7}$$

$$= - \int_{\mathcal{Q}_T} \mathcal{K}(\mathbf{a}(t, x), \mathbf{b}(t, x)) dx dt,\tag{D.8}$$

where we denoted  $\mathcal{K}(a, b) = \inf_{m \geq 0} \left[ m(\mathbf{a} - H(x, m, \mathbf{b})) \right]$ .

Therefore, the form of  $\mathcal{G}$  depends on  $\mathcal{K}$ . In general, we do not know any closed form of  $\mathcal{K}$  in terms of  $\mathbf{a}$  and  $\mathbf{b}$ . However, for several Hamiltonian functions  $H(x, m, p)$  of interest, we can derive such closed-form solution. Here we demonstrate some of the calculations to deliver a general idea.

**Example 4** (Mean-field Aversion). *Consider the running cost  $f(m, v) = \frac{1}{4}\|v\|^2 + m$  and drift  $b(x, m, v) = v$ . The corresponding Hamiltonian is  $H(x, m, p) = \|p\|^2 - m$ . Then*

$$\mathcal{K}(\mathbf{a}, \mathbf{b}) = \inf_{m \geq 0} m^2 + (\mathbf{a} - \|\mathbf{b}\|^2)m = \begin{cases} -\frac{1}{4}(\mathbf{a} - \|\mathbf{b}\|^2)^2 & \text{if } \mathbf{a} - \|\mathbf{b}\|^2 \geq 0 \\ 0 & \text{otherwise.} \end{cases}\tag{D.9}$$

**Example 5** (Mean-field Maximum Entropy). *Consider the running cost  $f(m, v) = \frac{1}{4}\|v\|^2 + \log(m)$  and drift  $b(x, m, v) = v$ , the corresponding Hamiltonian is  $H(x, m, p) = \frac{1}{2}\|p\|^2 - \log(m)$ , then*

$$\mathcal{K}(\mathbf{a}, \mathbf{b}) = \inf_{m \geq 0} m \log m + (\mathbf{a} - \frac{1}{2}\|\mathbf{b}\|^2)m = -\exp\left(\frac{1}{2}\|\mathbf{b}\|^2 - \mathbf{a} - 1\right).\tag{D.10}$$

**Example 6** (Continuous Optimal Transport). *Consider the running cost  $f(m, v) = \frac{1}{2}\|v\|^2$  and drift  $b(x, m, v) = v$ , the corresponding Hamiltonian is  $H(x, m, p) = \frac{1}{2}\|p\|^2$ , then*

$$\mathcal{K}(\mathbf{a}, \mathbf{b}) = \inf_{m \geq 0} m(\mathbf{a} - \frac{1}{2}\|\mathbf{b}\|^2) = \begin{cases} 0 & \text{if } \mathbf{a} - \frac{1}{2}\|\mathbf{b}\|^2 \geq 0 \\ -\infty & \text{otherwise.} \end{cases}\tag{D.11}$$

**Example 7** (Mean-field Congestion). *Consider the running cost  $f(m, v) = \frac{1}{4}m\|v\|^2$  and drift  $b(x, m, v) = v$ , the corresponding Hamiltonian is  $H(x, m, p) = \frac{\|p\|^2}{m}$ , then*

$$\mathcal{K}(\mathbf{a}, \mathbf{b}) = \inf_{m \geq 0} m\left(\mathbf{a} - \frac{\|\mathbf{b}\|^2}{m}\right) = \begin{cases} -\|\mathbf{b}\|^2 & \text{if } \mathbf{a} \geq 0 \\ -\infty & \text{otherwise.} \end{cases}\tag{D.12}$$

From these examples, it can be seen that  $\mathcal{K}$  has a closed form for many smooth Hamiltonian. However, the existence of closed form expressions of  $\mathcal{K}$  alone is not enough for making the

training loss tractable. In the example of (D.11) and (D.12),  $\mathcal{K}$  takes values  $-\infty$ , which makes  $\mathcal{G}$  singular and computationally intractable at some points. Moreover, gradient-based training cannot be carried out successfully in the presence of infinite values as well. The presence of  $-\infty$  in  $\mathcal{K}$  is due to the degeneracy of  $H$  in terms of order in  $m$ . For cases with singular  $\mathcal{K}$  and  $\mathcal{G}$ , we need an additional trick to tackle this issue.

Recall that in Algorithm 1, the update for function  $q$  is given by,

$$\begin{aligned} q^{(k)} &= \arg \min_{q: \mathcal{Q}_T \rightarrow \mathbb{R}^{d+1}} \mathcal{G}(q) + \langle \lambda^{(k-1)}, q \rangle + \frac{r}{2} \|\Lambda u^{(k)} - q\|^2 \\ &= \arg \min_{q: \mathcal{Q}_T \rightarrow \mathbb{R}^{d+1}} \int_{\mathcal{Q}_T} \left( -\mathcal{K}(q(t, x)) + \langle \lambda^{(k-1)}(t, x), q(t, x) \rangle + \frac{r}{2} \|\Lambda u^{(k)}(t, x) - q(t, x)\|^2 \right) dx dt \end{aligned} \quad (\text{D.13})$$

Since we don't have any additional constraint on the value of  $q$ , the function  $q^{(k)}$  that minimizes the integral in (D.13) should minimize the integrand point wisely. Therefore, it holds that:

$$\begin{aligned} q^{(k)}(t, x) &= \arg \min_{q \in \mathbb{R}^{d+1}} \left( -\mathcal{K}(q) + \langle \lambda^{(k-1)}(t, x), q \rangle + \frac{r}{2} \|\Lambda u^{(k)}(t, x) - q\|^2 \right) \\ &= \arg \min_{q \in \mathbb{R}^{d+1}} \sup_{m \geq 0} m(H(x, m, q_2) - q_1) + \langle \lambda^{(k-1)}(t, x), q \rangle + \frac{r}{2} \|\Lambda u^{(k)}(t, x) - q\|^2, \end{aligned} \quad (\text{D.14})$$

where we denote  $q = (q_1, q_2)$ , with  $q_1 \in \mathbb{R}$  and  $q_2 \in \mathbb{R}^d$ . For fixed  $\lambda^{(k-1)}$  and  $u^{(k)}$ , let us define  $\mathcal{L}(q, m) = m(H(x, m, q_2) - q_1) + \langle \lambda^{(k-1)}(t, x), q \rangle + \frac{r}{2} \|\Lambda u^{(k)}(t, x) - q\|^2$ . Under certain conditions, the following minimax equality holds for the right-hand side of (D.14).

$$\inf_{q \in \mathbb{R}^{d+1}} \sup_{m \geq 0} \mathcal{L}(q, m) = \sup_{m \geq 0} \inf_{q \in \mathbb{R}^{d+1}} \mathcal{L}(q, m) \quad (\text{D.15})$$

Therefore, we can exploit (D.15) to address the issues of singular  $\mathcal{K}$ . We notice that for the cases of congestion and general linear quadratic,  $\mathcal{L}(q, m)$  is quadratic in  $q$  for fixed  $m$ . This means that we can solve explicitly in a closed form for  $\inf_q \mathcal{L}(q, m)$  for fixed  $m$ , then we solve for the maximization problem over  $m \geq 0$ . In this way, we can effectively avoid the issues generated by infinity values in  $\mathcal{K}$ .

Moreover, using this trick, we directly obtain the value for  $q^{(k)}$  at any  $(t, x)$  based on the value of  $u^{(k)}(t, x)$  and  $\lambda^{(k-1)}(t, x)$ . Therefore, we can skip the neural network training in Algorithm 2 for function  $q$  in each DeepADMM iteration. Instead, we compute the values for  $q^{(k)}(t, x)$  directly following the above procedure whenever the values are needed to compute  $\mathcal{L}^{(u)}$  and  $\mathcal{L}^{(\lambda)}$ .

### D.3 Loss plots

In this section, we provide plots for the training losses of the proposed algorithms on each of the examples discussed in the text. For each test case and each of the three methods, we ran the algorithm 10 times and computed the average loss (or average component of the loss, for

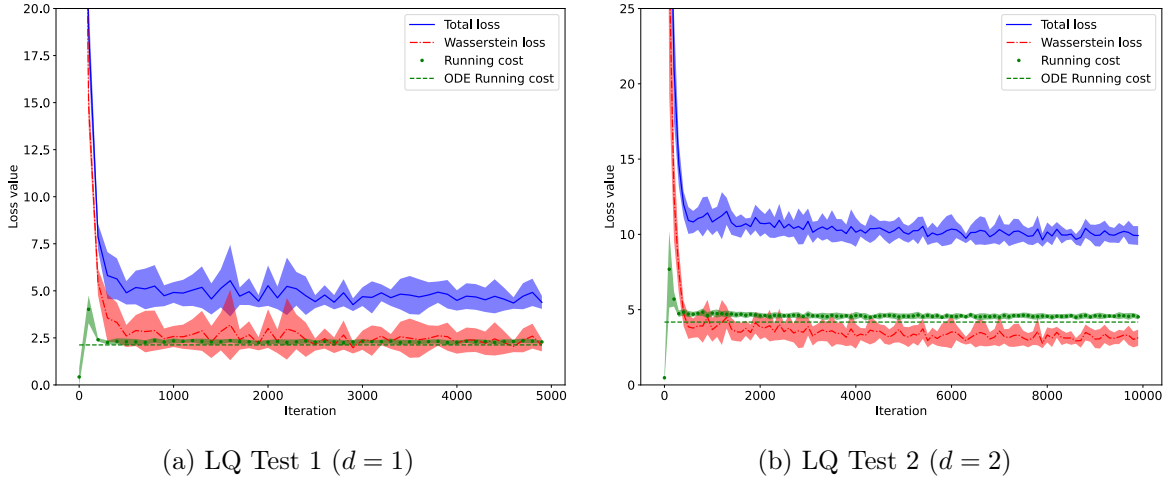


Figure D.1: Evolution of the loss in the linear-quadratic case for method 1.

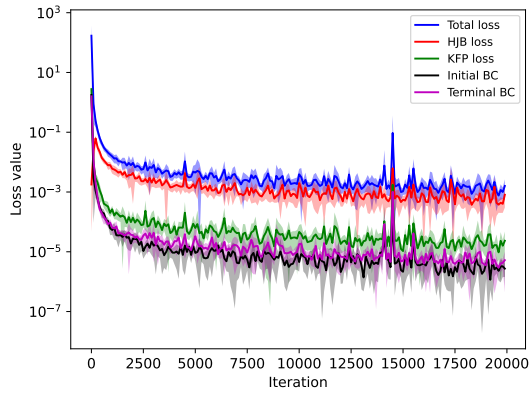
each component). We also computed the standard deviation over these 10 runs. In each plot, we display the average loss by a line and we represent by a shaded area the region contained between the average minus the standard deviation and the average plus the standard deviation.

We start by presenting the loss plots for the two LQ test cases in [Figure D.1](#) for Method 1, [Figure D.2](#) for Method 2, and [Figure D.3](#) for Method 3.

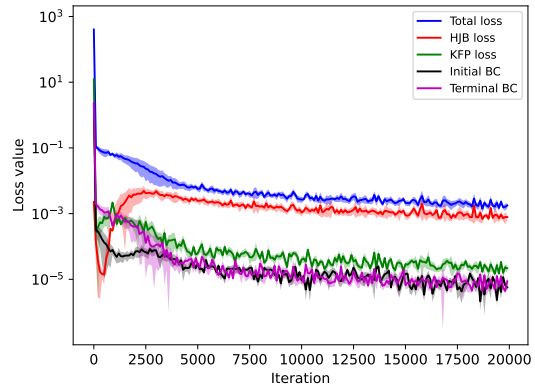
We then present the losses for the congestion test cases. Contrary to the LQ case, we do not have any benchmark solution to compare our numerical results to. We already explained why the methods gave consistent results, namely that we see the density spreading before reforming the terminal distribution, in contrast with the LQ case where the shape of the population remains the same along the trajectory (see [III.4.2.2](#)). The following comments can be made about the plots of the losses:

- [Figures D.4](#) and [D.5](#) show the losses for Method 1. The total loss is the sum of the “running cost” loss and the “terminal penalization” loss multiplied by  $C_W$ . For instance in dimension 1 we took  $C_W = 10$ , meaning that the penalization is ten times the Wasserstein distance between the effective terminal distribution of agents and the desired one. In all three cases, we observe that for the first iterations, the distance is high and the running cost almost zero, which comes from the fact that the first try is not to move. Then the algorithm makes the penalization decrease by paying a trade-off in the form of the running cost. In the case with no congestion ([Figure D.4a](#)), it seems to reach a plateau which approximately corresponds to the theoretical optimal running cost<sup>1</sup>, while in the cases with congestion (in dimension 1 and 5), the running cost keeps on decaying slowly as the algorithm tries to spread as much as possible before moving.

<sup>1</sup>This test case with no congestion is in fact another instance of the LQ test in which we only want to move from one Gaussian distribution to another one with the same variance. Hence the optimal cost is the distance between the two means, which in this case is 2.



(a) LQ Test 1 ( $d = 1$ )

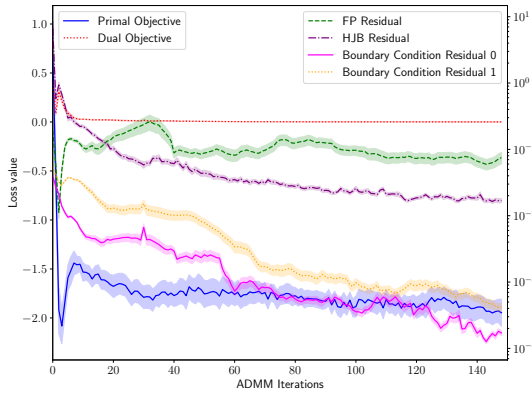


(b) LQ Test 2 ( $d = 2$ )

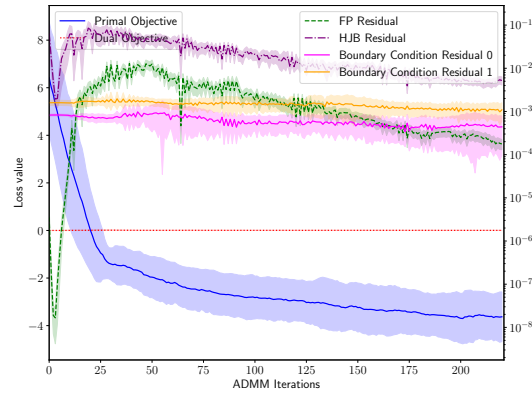
Figure D.2: Evolution of the loss in the linear-quadratic case for method 2.

- Figures [D.6](#) and [D.7](#) show the losses for Method 2. The “HJB loss” is the squared  $L^2$  residual for the HJB equation. The “KFP loss” is the squared  $L^2$  residual for the KFP equation. The “initial BC” loss and the “terminal BC” loss respectively correspond to the squared  $L^2$  error on the initial and terminal distributions. The “total loss” is the sum of the other losses, up to multiplicative weights.
- Figures [D.8](#) and [D.9](#) show, for Method 3, the squared  $L^2$  residuals for the HJB and KFP equations, as well as the squared  $L^2$  loss for the initial and terminal conditions. Note that these losses are not directly minimized during the algorithm of Method 3, but they are minimized as a by-product of the iterations.



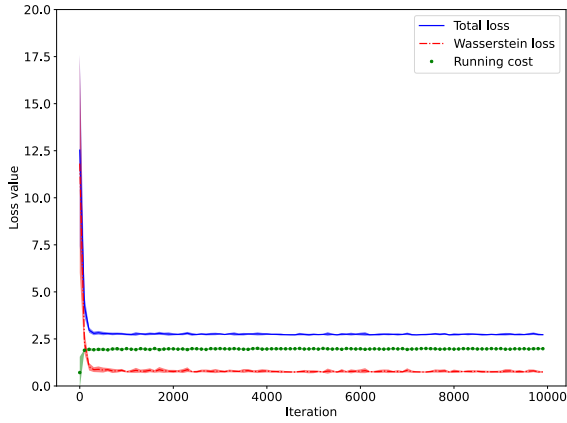


(a) LQ Test 1 ( $d = 1$ )

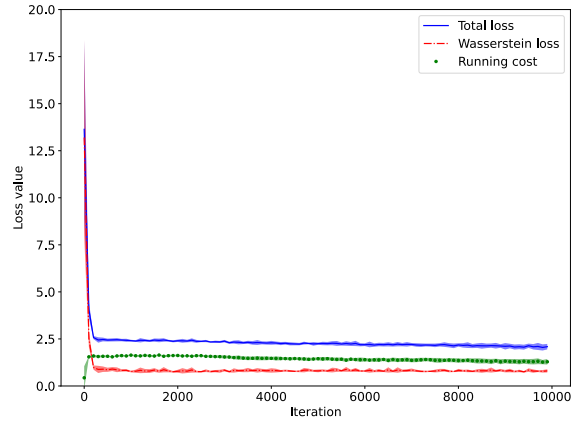


(b) LQ Test 2 ( $d = 2$ )

Figure D.3: Evolution of the loss in the linear-quadratic case for method 3.



(a) Case 1, No congestion ( $d = 1$ )



(b) Case 2, Congestion ( $d = 1$ )

Figure D.4: Evolution of the loss in the congestion cases 1 and 2 for method 1.

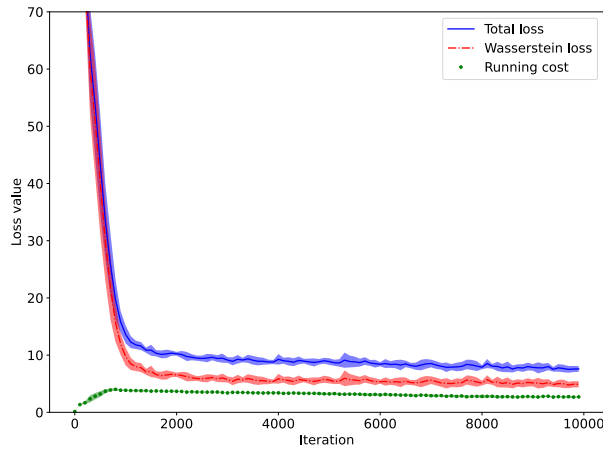
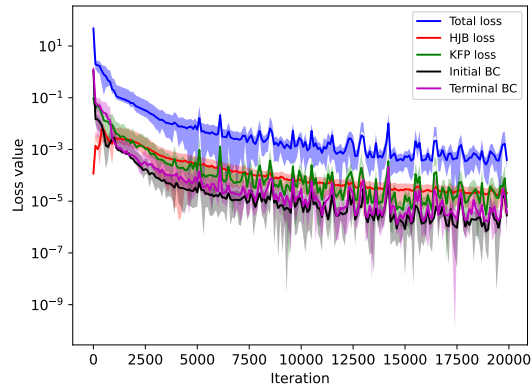
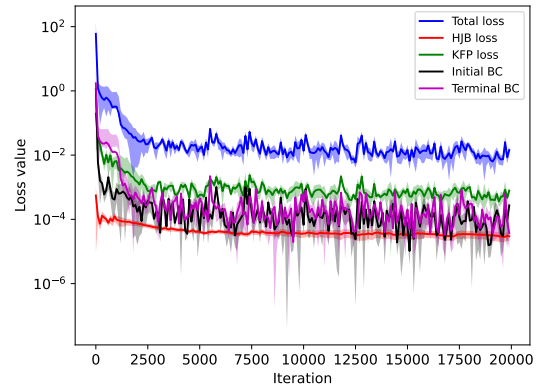


Figure D.5: Evolution of the loss in the congestion case in dimension 5 for method 1.



(a) Case 1, No congestion ( $d = 1$ )



(b) Case 2, Congestion ( $d = 1$ )

Figure D.6: Evolution of the loss in the congestion cases 1 and 2 for method 2.

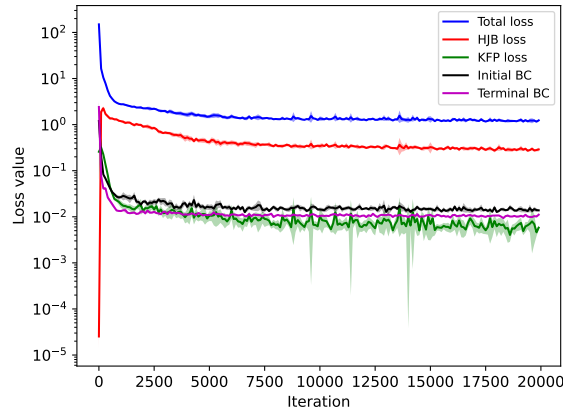
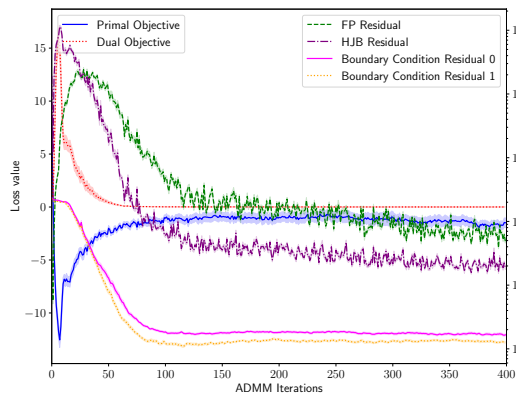
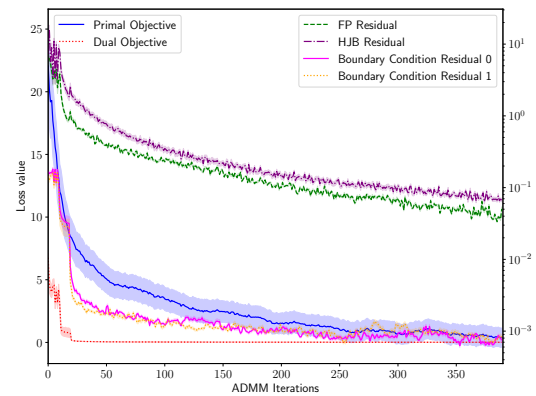


Figure D.7: Evolution of the loss in the congestion case in dimension 5 for method 2.



(a) Case 1, No Congestion ( $d = 1$ )



(b) Case 2, Congestion ( $d = 1$ )

Figure D.8: Evolution of the loss in the congestion cases 1 and 2 for method 3.

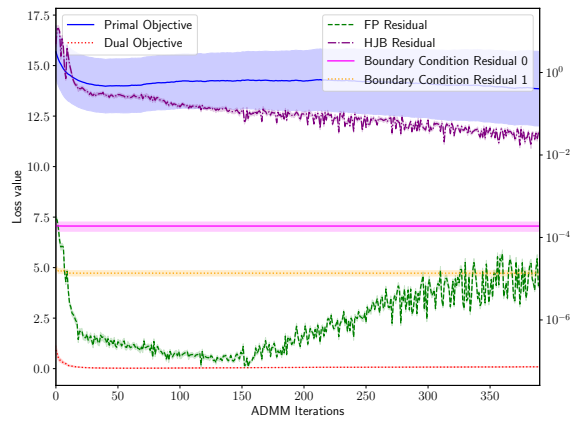


Figure D.9: Evolution of the loss in the congestion case in dimension 5 for method 3.



# Index of notations for Chapters I and II

- $(K_t)_{t \in \mathbb{R}_+}$ : Markov semigroup on  $S$ , [61](#)  
 $(P_t)_{t \geq 0}$ : semigroup of  $X$ , [35](#)  
 $(S, \mathcal{S})$ : lcsm space, [61](#)  
 $(\mathcal{F}_t)_{t \geq 0}$ : complete and right-continuous filtration on  $(\Omega, \mathcal{F})$ , [36](#)  
 $(\mathbb{Q}_x)_{x \in \mathbb{R}}$ : family of probability measures on  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0})$ , [6](#), [35](#)  
 $(\check{P}_t^{(n)})_{t \in \mathbb{R}_+}$ : Markov semigroup constructed in the proof of [Theorem II.1](#), [23](#), [63](#)  
 $A, A^*$ : generator of the semigroup  $(P_t)_{t \geq 0}$  and its adjoint, [17](#), [28](#), [38](#)  
 $D_+$ : set of positive rational dyadic numbers, [71](#)  
 $F$ : hydrodynamic limit of the  $N$ -BMP, [7](#)  
 $F^N$ : empirical c.d.f. of the  $N$ -BMP, [3](#), [36](#)  
 $F_{x,t}^{[-1]}$ : quantile function of the distribution  $P_t(x, \cdot)$ , [70](#)  
 $G$ : complement of the c.d.f. of  $X$  conditioned on not crossing  $b$ , [40](#)  
 $G^n$ : complement of the empirical c.d.f. of  $n$  BMPs, [41](#)  
 $G^{\pm, \delta, N}$ : complement of the empirical c.d.f. of  $\mathcal{X}^\pm$ , [45](#)  
 $K, L$ : Markov kernels on  $(S, \mathcal{S})$ , [61](#)  
 $N_\delta^+$ :  $\lceil N(1 + \delta) \rceil$ , initial size of  $\mathcal{X}^+$ , [45](#)  
 $N_\delta^-$ :  $\lfloor N(1 - \delta) \rfloor$ , initial size of  $\mathcal{X}^-$ , [45](#)  
 $N_t, N_t^i$ : size at time  $t$  of the  $b$ -BMP made from  $\mathcal{X}, \mathfrak{x}^i$ , [41](#), [47](#)  
 $U(t, \cdot)$ : distribution of  $X$  on  $\{\tau_b > t\}$ , [36](#)  
 $W_1$ : Wasserstein-1 metric on  $\mathcal{P}(S)$ , [67](#)  
 $X, (X_t)_{t \geq 0}$ : underlying Markov process, [6](#), [35](#)  
 $\mathcal{B}_b$ : set of bounded real-valued Borel functions on  $\mathbb{R}$ , [35](#)  
 $\mathcal{B}_b(S)$ : set of bounded real-valued measurable functions on  $S$ , [61](#)  
 $\mathcal{B}_b^\nearrow(\mathbb{R})$ : set of bounded non-decreasing real-valued functions on  $\mathbb{R}$ , [63](#)  
 $\mathcal{C}_0$ : set of continuous functions on  $\mathbb{R}$  vanishing at  $\pm\infty$ , [35](#)  
 $\mathcal{C}_0(S)$ : set of continuous real-valued functions on  $S$  vanishing at infinity, [62](#)  
 $\mathcal{C}_b$ : set of continuous bounded functions on  $\mathbb{R}$ , [35](#)  
 $\mathcal{C}_b(S)$ : set of bounded continuous real-valued functions on  $S$ , [62](#)  
 $\mathcal{D}(I, E)$ : set of càdlàg function from an interval  $I$  of  $\mathbb{R}$  with values in  $E$ , [36](#)  
 $\Gamma_t^\pm$ : favorable on which we can build a coupling  $\mathcal{X} \preceq \mathcal{X}^+$  (resp.  $\mathcal{X}^- \preceq \mathcal{X}$ ), [46](#)  
 $\text{Lip}(S)$ : set of real-valued Lipschitz functions on  $S$ , [67](#)  
 $\mathcal{N}^{\mathfrak{x}}, \mathcal{N}^i$ : size of the BMP, [41](#)  
 $\mathfrak{N}_{x,y}$ : number of particles that cross downwards the interval  $[y, x]$ , [55](#)  
 $\mathbb{P}$ : distribution on  $\Omega$  for an  $N$ -BMP starting with  $N$  particles with i.i.d. positions given by  $\mu_0$ , [36](#)  
 $\mathcal{P}(S)$ : set of Borel probability measures on  $S$ , [66](#)  
 $\mathbb{Q}_{\mu_0}$ : probability measure on  $\Omega$  under which  $X$  starts with initial distribution  $\mu_0$ , [35](#)  
 $\mathcal{SM}_b(\mathbb{R}^n)$ : set of real-valued bounded Borel super-modular functions on  $\mathbb{R}^n$ , [63](#)  
 $\mathcal{X}$ :  $N$ -BMP, [39](#), [49](#)  
 $\mathcal{X}^\pm$ : stochastic barriers, [39](#), [45](#)  
 $\Xi_{x,y}$ : event on which no particle of  $\mathcal{X}$  crosses downwards the interval  $[y, x]$ , [55](#)  
 $\chi(\mathfrak{X}, r)$ : number of particles above  $r$  for the  $b$ -BMP made from  $\mathfrak{X}$ , [40](#)  
 $\delta_x$ : Dirac mass at  $x$ , [3](#)  
 $b$ : boundary (upper semi-continuous curve), [6](#), [35](#)  
 $\mathfrak{M}_n$ : set of Feller semigroups on  $\mathbb{R}^n$  that are extensions of  $(P_t)_{t \geq 0}$ , [63](#)  
 $\mathfrak{x}, \mathfrak{x}^i$ : BMP starting from one particle, [40](#)  
 $\mathcal{X}$ : BMP starting from  $n$  particles, [41](#), [47](#)  
 $\mu_0$ : initial distribution of the particles, [35](#)  
 $\|\cdot\|_\infty$ : supremum norm on  $\mathcal{B}_b$ , [37](#)  
 $\|\cdot\|_{\text{Lip}}$ : norm on  $\text{Lip}(S)$ , [67](#)

$\psi$ : limite hydrodynamique du  $N$ -BBM, [3]  
 $\sigma_{i\ell}$ :  $\ell$ -th instant of death in the  $i$ -th family of  $\mathcal{X}$ , [47]  
 $\leq_{\text{sm}}$ : super-modular partial order on the set of probability measures on  $\mathbb{R}^n$ , [23, 63]  
 $\leq_{\text{st}}$ : stochastic partial order on the set of probability measures on  $\mathbb{R}$ , [63]  
 $\tau_b$ : first downwards crossing time of the boundary  $b$  for the underlying process  $X$ , [6, 36]  
 $\tau_y$ : downwards crossing time of  $y$  for  $X$ , [37]  
 $\mathbf{1}$ : constant function equal to 1 on  $S$ , [61]  
 $d_S$ : metric on  $S$ , [67]  
 $m_t^N$ : lowest particle position of the  $N$ -BMP at time  $t$ , [37]  
 $u^\alpha$ : travelling wave solution for the free boundary problem [(2), 30]  
 $u_0^\alpha$ : p.d.f. of the quasi-stationary distribution for the Brownian motion with drift  $-\alpha$  killed at 0, [16]

# Index of notations for Chapter III and Appendix D

- $(X_t^{i,v_\theta})_{t \geq 0}$ : position of agents, [81](#)  
 $A, \bar{A}, B, R, \bar{x}_0, \bar{x}_T, \Sigma_0, \Sigma_T$ : parameters for the LQ problem, [92](#), [141](#)  
 $C_W, C^{(KFP)}, C_0^{(KFP)}, C_T^{(KFP)}, C^{(HJB)}$ : penalization constants, [83](#), [86](#)  
 $H$ : Hamiltonian of the control problem for one agent, [84](#)  
 $J^{MFC}$ : MFC cost, [79](#)  
 $J^{MFOT}$ : MFOT cost, [79](#)  
 $J^{N,NT}$ : approximation of  $J^{MFOT}$  used in Method 1, [82](#)  
 $L$ : Lagrangian of the control problem for one agent, [84](#)  
 $R, \gamma, c, \bar{x}_0, \bar{x}_T, \Sigma_0, \Sigma_T$ : parameters for the congestion problem, [98](#)  
 $X^v, (X_t^v)_{0 \leq t \leq T}$ : random position of agents driven by [\(III.2\)](#), [79](#)  
 $\mathcal{B}$ : loss function for the primal problem, [88](#)  
 $\mathcal{L}(\theta; S)$ : random loss used in the deep learning methods, [82](#)  
 $\mathcal{P}_2(\mathcal{Q})$ : set of square-integrable probability measures on  $\mathcal{Q}$ , [79](#)  
 $\mathcal{W}_2$ : Wasserstein-2 metric on  $\mathcal{P}_2(\mathcal{Q})$ , [81](#)  
 $\mathcal{W}_p$ : Wasserstein- $p$  metric on empirical distributions, [83](#)  
 $\bar{\mu}^{N,v_\theta}(t)$ : empirical distribution of agents at time  $t$ , [81](#)  
 $\hat{v}$ : learnt control, [92](#)  
 $\mathbf{K}$ : state space for the primal problem, [88](#)  
 $\mathcal{A}, \mathcal{F}, \mathcal{G}$ : loss functions for the dual problem, [89](#), [143](#)  
 $\mathcal{Q}$ : spatial domain, [79](#)  
 $\mathcal{Q}_T$ : time-space domain, [79](#)  
 $\mu^v(t)$ : distribution of agents at time  $t$ , [79](#)  
 $\rho_0$ : initial distribution of agents, [79](#)  
 $\rho_T$ : target distribution of agents, [79](#)  
 $\theta, \omega, \psi$ : parameters of neural networks, [81](#), [86](#)
- $b, \sigma$ : coefficients of the SDE driving the agents, [79](#)  
 $f$ : running cost function, [79](#)  
 $g$ : terminal cost function, [79](#)  
 $m(t, \cdot)$ : density of  $\mu_t^v$ , [84](#)  
 $m_\theta, u_\omega$ : neural network approximations for solutions of [\(III.12\)](#), [86](#)  
 $u$ : adjoint state of the PDE system [\(III.11\)](#), [85](#)  
 $v$ : control function, [79](#)  
 $v^*$ : optimal, [92](#)  
 $v_\theta$ : approximation of optimal control by a neural network, [81](#)





# Index of notations for Appendices A and B

- $(P_t^j)_{t \geq 0}$ : semigroup of the coordinate process  $Y^j$ , [117](#)  
 $(P_t^{(2)})_{t \geq 0}$ : Feller order-preserving semigroup for the evolution of paired particles, [122](#)  
 $(P_t)_{t \geq 0}$ : semigroup of the underlying Markov process  $(X_t)_{t \geq 0}$ , [122](#)  
 $(\mathcal{W}_t)_{t \geq 0}$ : complete right-continuous filtration, [104](#)  
 $(\psi_t)_{t \geq 0}$ : shift operator on  $W$ , [104](#)  
 $(\mathcal{P}_t)_{t \geq 0}$ : semigroup of the base process  $\mathcal{Y}$ , [117](#)  
 $(\mathcal{P}_t)_{t \geq 0}$ : semigroup of the initial evolution of the coupling, [122](#)  
 $E, E_i$ : locally compact Hausdorff space with countable basis, [104](#)  
 $E_1$ : state space  $\mathbb{R}^N$  of the  $N$ -BMP, [109](#)  
 $E_2$ : state space  $\left( \{\diamond\} \sqcup \bigsqcup_{n \geq 1} \mathbb{R}^n \right) \times \mathbb{R}_+$  of the  $b$ -BMP, [109](#)  
 $E_3$ : discrete space describing the links between particles in the coupling, [121](#)  
 $M^{\zeta, \theta}(\lambda, \nu)$ : mixed renewal distribution, [105](#), [112](#)  
 $T_\partial$ : hitting time of  $\partial$  for the  $E_3$ -coordinate of the coupling, [127](#)  
 $W$ : initial probability space, [104](#)  
 $X, (X_t)_{t \geq 0}$ : underlying Markov process, [122](#)  
 $\mathcal{C}_0(\bar{E})$ : set of continuous functions on  $\bar{E}$  vanishing at  $\Delta$ , [104](#)  
 $\Delta$ : isolated cemetery point in  $E$ , [104](#)  
 $\mathbb{P}_y$ : initial probability measure on  $W$ , [104](#)  
 $\mathfrak{P}W, \mathbf{W}$ : probability space  $\prod_{i=1}^{+\infty} (W \times \bar{E})$  for the pasted process, [106](#)  
 $\mathfrak{P}_y^{\zeta, \theta}, \mathbf{P}_y$ : probability measure on  $\mathfrak{P}W$ , [106](#)  
 $\mathfrak{P}^{\zeta} \mathcal{Y}, \mathcal{Y}$ : pasted process, [106](#)  
 $\mathfrak{P}_k^{\zeta} \mathcal{Y}, \mathfrak{P}_k^{\zeta} Y^j$ : truncated version of the pasted process, [119](#)  
 $\mathfrak{P}^{(2)}W$ : probability space for the pasting of pasted process, [111](#)  
 $\mathfrak{P}^{\zeta, \theta} \mathbb{P}_y^{\lambda, \nu}, \mathfrak{P}^{\theta} \mathbf{P}_y^{\nu}$ : probability measures on  $\mathfrak{P}^{(2)}W$  for the pasting of pasted process, [112](#)  
 $\mathfrak{P}^{\zeta, \theta} \mathcal{Y}, \mathfrak{P}^{\theta} \mathcal{Y}$ : pasting of pasted process, [111](#)  
 $\mathbb{Q}_y^{\lambda}$ : probability measure on  $\Omega$ , [106](#)  
 $\mathcal{Y}, (\mathcal{Y}_t)_{t \geq 0}$ : base process (strong Markov process on  $E$ ), [104](#)  
 $\bar{E}$ : compactification  $E \cup \{\Delta\}$  of the state space  $E$ , [104](#)  
 $\bar{\mathcal{Y}}$ : killed version of the base process, [105](#)  
 $\bar{\zeta}_i$ :  $i$ -th cumulative renewal time, [106](#)  
 $\nu$ : extension of  $\nu$  on  $W$ , [111](#)  
 $\theta$ : extension of  $\theta$  on  $W$ , [110](#)  
 $\partial$ : state in  $E_3$  describing independence between the  $N$ -BMP and the  $b$ -BMP, [121](#)  
 $\lambda, \nu, \xi$ : renewal distribution, [105](#), [110](#)  
 $\lambda^2$ : renewal distribution performing the killing of a particle in the  $b$ -BMP, [126](#)  
 $\mathfrak{T}$ : lifetime of the pasted process  $\mathfrak{P}\mathcal{Y}$ , [106](#)  
 $\nu^1$ : renewal distribution performing the branching and selection in the  $N$ -BMP, and the branching of paired particles in the  $b$ -BMP, [124](#)  
 $\pi_j$ : canonical projection from  $E = \prod_{i=1}^n E_i$  to  $E_j$ , [117](#)  
 $\theta^1$ : branching time for the  $N$ -BMP (and paired particles), [123](#)  
 $\theta^2$ : branching time for free particles in the  $b$ -BMP, [123](#)  
 $\tau$ : minimum of  $\zeta$  and  $\theta$ , [110](#)  
 $\zeta, \theta, \tau$ : renewal time, [105](#), [110](#)  
 $\zeta^2$ : first death time in the  $b$ -BMP, [123](#)  
 $\zeta^\dagger$ : compatible stopping time on  $W$ , [104](#)  
 $\zeta_i$ :  $i$ -th lifetime interval, [106](#)



# Références

- [AACN<sup>+</sup>18] Ali Al-Aradi, Adolfo Correia, Danilo Naiff, Gabriel Jardim, and Yuri Saporito. Solving nonlinear and high-dimensional partial differential equations via deep learning. preprint, arXiv:1811.08782, 2018.
- [ABBP19] Louigi Addario-Berry, Julien Berestycki, and Sarah Penington. Branching Brownian motion with decay of mass and the nonlocal Fisher-KPP equation. *Comm. Pure Appl. Math.*, 72(12):2487–2577, 2019.
- [ABBS13] Elie Aïdékon, Julien Berestycki, Eric Brunet, and Zhan Shi. Branching Brownian motion seen from its tip. *Probab. Theory Related Fields*, 157(1-2):405–451, 2013.
- [Abu15] Mario Abundo. An overview on inverse first-passage-time problems for one-dimensional diffusion processes. In *Recent advances in probability and statistics*, volume 12 of *Lect. Notes Semin. Interdiscip. Mat.*, pages 1–44. Semin. Interdiscip. Mat. (S.I.M.), Potenza, 2015.
- [ACCD12] Yves Achdou, Fabio Camilli, and Italo Capuzzo-Dolcetta. Mean field games: numerical methods for the planning problem. *SIAM Journal on Control and Optimization*, 50(1):77–109, 2012.
- [AFG11] Amine Asselah, Pablo A. Ferrari, and Pablo Groisman. Quasistationary distributions and Fleming-Viot processes in finite spaces. *J. Appl. Probab.*, 48(2):322–332, 2011.
- [AFGJ16] Amine Asselah, Pablo A. Ferrari, Pablo Groisman, and Matthieu Jonckheere. Fleming-Viot selects the minimal quasi-stationary distribution: the Galton-Watson case. *Ann. Inst. Henri Poincaré Probab. Stat.*, 52(2):647–668, 2016.
- [AGS05] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2005.
- [AL15] Yves Achdou and Mathieu Laurière. On the system of partial differential equations arising in mean field type control. *Discrete and Continuous Dynamical Systems*, 35(9):3879–3900, 2015.
- [AL16a] Yves Achdou and Mathieu Laurière. Mean field type control with congestion. *Applied Mathematics & Optimization*, 73(3):393–418, 2016.
- [AL16b] Yves Achdou and Mathieu Laurière. Mean field type control with congestion (II): An augmented Lagrangian method. *Applied Mathematics & Optimization*, 74(3):535–578, 2016.

- [AL19] Yves Achdou and Jean-Michel Lasry. Mean field games for modeling crowd motion. *Contributions to partial differential equations and applications*, pages 17–42, 2019.
- [And91] William J. Anderson. *Continuous-time Markov chains*. Springer Series in Statistics: Probability and its Applications. Springer-Verlag, New York, 1991. An applications-oriented approach.
- [And17] Roman Andreev. Preconditioning the augmented Lagrangian method for stationary mean field games with diffusion. *SIAM Journal on Scientific Computing*, 39(6):A2763–A2783, 2017.
- [Anu80] Svetlana V. Anulova. Markov times with given distribution for a Wiener process. *Teor. Veroyatnost. i Primenen.*, 25(2):366–369, 1980.
- [AP18] Yves Achdou and Alessio Porretta. Mean field games with congestion. In *Annales de l’Institut Henri Poincaré C, Analyse non linéaire*, volume 35, pages 443–480. Elsevier, 2018.
- [Ata23] Rami Atar. Hydrodynamics of particle systems with selection via uniqueness for free boundary problems. preprint, arXiv:2011.07535, 2023.
- [Ata24] Rami Atar. A Durrett-Remenik particle system in  $\mathbb{R}^d$ . preprint, 2024.
- [AZ01] Marco Avellaneda and Jingyi Zhu. Distance to default. *RISK*, 2001.
- [BB00] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.
- [BBD18] Julien Berestycki, Éric Brunet, and Bernard Derrida. Exact solution and precise asymptotics of a Fisher-KPP type front. *J. Phys. A*, 51(3):035204, 21, 2018.
- [BBNP21] Julien Berestycki, Éric Brunet, James Nolen, and Sarah Penington. A free boundary problem arising from branching Brownian motion with selection. *Trans. Amer. Math. Soc.*, 374(9):6269–6329, 2021.
- [BBNP22] Julien Berestycki, Éric Brunet, James Nolen, and Sarah Penington. Brownian bees in the infinite swarm limit. *Ann. Probab.*, 50(6):2133–2177, 2022.
- [BBP19] Julien Berestycki, Éric Brunet, and Sarah Penington. Global existence for a free boundary problem of Fisher-KPP type. *Nonlinearity*, 32(10):3912–3939, 2019.
- [BBS15] Julien Berestycki, Nathanaël Berestycki, and Jason Schweinsberg. Critical branching Brownian motion with absorption: particle configurations. *Ann. Inst. Henri Poincaré Probab. Stat.*, 51(4):1215–1250, 2015.
- [BC15] Jean-David Benamou and Guillaume Carlier. Augmented Lagrangian methods for transport optimization, mean field games and degenerate elliptic equations. *Journal of Optimization Theory and Applications*, 167(1):1–26, 2015.
- [BCDMN19] Jean-David Benamou, Guillaume Carlier, Simone Di Marino, and Luca Nenna. An entropy minimization approach to second-order variational mean-field games. *Mathematical Models and Methods in Applied Sciences*, 29(08):1553–1583, 2019.

- [BCGL20] Julio Backhoff, Giovanni Conforti, Ivan Gentil, and Christian Léonard. The mean field Schrödinger problem: ergodic behavior, entropy estimates and functional inequalities. *Probability Theory and Related Fields*, 178(1):475–530, 2020.
- [BCL16] Jean-David Benamou, Guillaume Carlier, and Maxime Laborde. An augmented Lagrangian approach to Wasserstein gradient flows and applications. *ESAIM: Proceedings and surveys*, 54:1–17, 2016.
- [BCR22] Lucian Beznea, Iulian Cîmpean, and Michael Röckner. Strong Feller semigroups and Markov processes: A counter example. preprint, arXiv:2211.14789, 2022.
- [BD97] Eric Brunet and Bernard Derrida. Shift in the velocity of a front due to a cutoff. *Phys. Rev. E (3)*, 56(3):2597–2604, 1997.
- [BDMM06] Eric Brunet, Bernard Derrida, Alfred H. Mueller, and Stéphane Munier. Noisy traveling waves: effect of selection on genealogies. *Europhys. Lett.*, 76(1):1–7, 2006.
- [Bec19] Erin Beckman. *Asymptotic Behavior of Certain Branching Processes*. PhD thesis, Duke University, 2019.
- [BEES18] Mathias Beiglböck, Manu Eder, Christiane Elgert, and Uwe Schmock. Geometry of distribution-constrained optimal stopping problems. *Probab. Theory Related Fields*, 172(1-2):71–101, 2018.
- [Ber96] Jean Bertoin. *Lévy processes*, volume 121 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1996.
- [BEV10] Nicholas H. Barton, Alison M. Etheridge, and Amandine Véber. A new model for evolution in a spatial continuum. *Electron. J. Probab.*, 15:no. 7, 162–216, 2010.
- [BF23a] Jean Bérard and Briec Frénais. Hydrodynamic limit of N-branching Markov processes. preprint, arXiv:2311.12453, 2023.
- [BF23b] Jean Bérard and Briec Frénais. Monotone coupling of Feller Markov processes on the real line. preprint, arXiv:2311.18391, 2023.
- [BFL<sup>+</sup>23] Sebastian Baudalet, Briec Frénais, Mathieu Laurière, Amal Machtalay, and Yuchen Zhu. Deep learning for mean field optimal transport. preprint, arXiv:2302.14739, to appear in *ESAIM: ProcS*, 2023.
- [BFY13] Alain Bensoussan, Jens Frehse, and Phillip Yam. *Mean field games and mean field type control theory*, volume 101. Springer, 2013.
- [BFY15] Alain Bensoussan, Jens Frehse, and Sheung Chi Phillip Yam. The master equation in mean field theory. *Journal de Mathématiques Pures et Appliquées*, 103(6):1441–1474, 2015.
- [BG10] Jean Bérard and Jean-Baptiste Gouéré. Brunet-Derrida behavior of branching-selection particle systems on the line. *Communications in Mathematical Physics*, 298(2):323–342, 2010.

- [BHIM96] Krzysztof Burdzy, Robert Holyst, David Ingerman, and Peter March. Configurational transition in a Fleming-Viot-type model and probabilistic interpretation of laplacian eigenfunctions. *Journal of Physics A: Mathematical and General*, 29(11):2633, jun 1996.
- [BHM00] Krzysztof Burdzy, Robert Holyst, and Peter March. A Fleming-Viot particle representation of the Dirichlet Laplacian. *Comm. Math. Phys.*, 214(3):679–703, 2000.
- [Bil95] Patrick Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, third edition, 1995. A Wiley-Interscience Publication.
- [BJ22] Charles Boubel and Nicolas Juillet. The Markov-quantile process attached to a family of marginals. *J. Éc. polytech. Math.*, 9:1–62, 2022.
- [BLL21] Charles Bertucci, Jean-Michel Lasry, and Pierre-Louis Lions. Master equation for the finite state space planning problem. *Archive for Rational Mechanics and Analysis*, 242(1):327–342, 2021.
- [BPC<sup>+</sup>11] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends<sup>®</sup> in Machine learning*, 3(1):1–122, 2011.
- [Bra78] Maury D. Bramson. Maximal displacement of branching Brownian motion. *Comm. Pure Appl. Math.*, 31(5):531–581, 1978.
- [Bra83] Maury D. Bramson. Convergence of solutions of the Kolmogorov equation to travelling waves. *Mem. Amer. Math. Soc.*, 44(285):iv+190, 1983.
- [BT24] Julien Berestycki and Oliver Tough. Selection principle for the  $N$ -BBM. preprint, arXiv:2407.05792, 2024.
- [CCCS06] Lan Cheng, Xinfu Chen, John Chadam, and David Saunders. Analysis of an inverse first passage problem from risk management. *SIAM J. Math. Anal.*, 38(3):845–873, 2006.
- [CCCS11] Xinfu Chen, Lan Cheng, John Chadam, and David Saunders. Existence and uniqueness of solutions to the inverse boundary crossing problem for diffusions. *Ann. Appl. Probab.*, 21(5):1663–1693, 2011.
- [CCN13] Pierre Cardaliaguet, Guillaume Carlier, and Bruno Nazaret. Geodesics for a class of distances in the space of probability measures. *Calculus of Variations and Partial Differential Equations*, 48(3):395–420, 2013.
- [CCS22] Xinfu Chen, John Chadam, and David Saunders. Higher-order regularity of the free boundary in the inverse first-passage problem. *SIAM J. Math. Anal.*, 54(4):4695–4720, 2022.
- [CD18] René Carmona and François Delarue. *Probabilistic theory of mean field games with applications I-II*. Springer, 2018.

- [CDMGP16] Gioia Carinci, Anna De Masi, Cristian Giardinà, and Errico Presutti. *Free boundary problems in PDEs and particle systems*, volume 12 of *SpringerBriefs in Mathematical Physics*. Springer, [Cham], 2016.
- [CG15] Pierre Cardaliaguet and P. Jameson Graber. Mean field games systems of first order. *ESAIM: Control, Optimisation and Calculus of Variations*, 21(3):690–722, 2015.
- [CGL20] Haoyang Cao, Xin Guo, and Mathieu Laurière. Connecting GANs, MFGs, and OT. preprint, arXiv:2002.04112, 2020.
- [CGP15a] Yongxin Chen, Tryphon T. Georgiou, and Michele Pavon. Optimal steering of a linear stochastic system to a final probability distribution, part I. *IEEE Transactions on Automatic Control*, 61(5):1158–1169, 2015.
- [CGP15b] Yongxin Chen, Tryphon T. Georgiou, and Michele Pavon. Optimal steering of a linear stochastic system to a final probability distribution, part II. *IEEE Transactions on Automatic Control*, 61(5):1170–1180, 2015.
- [CGP16] Yongxin Chen, Tryphon T. Georgiou, and Michele Pavon. On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. *Journal of Optimization Theory and Applications*, 169(2):671–691, 2016.
- [CGP18a] Yongxin Chen, Tryphon T. Georgiou, and Michele Pavon. Optimal steering of a linear stochastic system to a final probability distribution, part III. *IEEE Transactions on Automatic Control*, 63(9):3112–3118, 2018.
- [CGP18b] Yongxin Chen, Tryphon T. Georgiou, and Michele Pavon. Steering the distribution of agents in mean-field games system. *Journal of Optimization Theory and Applications*, 179(1):332–357, 2018.
- [Che04] Mu-Fa Chen. *From Markov chains to non-equilibrium particle systems*. World Scientific Publishing Co., Inc., River Edge, NJ, second edition, 2004.
- [CIL92] Michael G. Crandall, Hitoshi Ishii, and Pierre-Louis Lions. User’s guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc. (N.S.)*, 27(1):1–67, 1992.
- [CIR85] John C. Cox, Jonathan E. Ingersoll, Jr., and Stephen A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, 1985.
- [CL19] René Carmona and Mathieu Laurière. Convergence analysis of machine learning algorithms for the numerical solution of mean field control and games: II: The finite horizon case. preprint, arXiv:1908.01613, 2019.
- [CL21a] René Carmona and Mathieu Laurière. Convergence analysis of machine learning algorithms for the numerical solution of mean field control and games I: the ergodic case. *SIAM Journal on Numerical Analysis*, 59(3):1455–1485, 2021.
- [CL21b] René Carmona and Mathieu Laurière. Deep learning for mean field games and mean field control with applications to finance. preprint, arXiv:2107.04568, 2021.
- [CMSM13] Pierre Collet, Servet Martínez, and Jaime San Martín. *Quasi-stationary distributions*. Probability and its Applications (New York). Springer, Heidelberg, 2013. Markov chains, diffusions and dynamical systems.

- [Cut13] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.
- [CV21] Nicolas Champagnat and Denis Villemonais. Convergence of the Fleming-Viot process toward the minimal quasi-stationary distribution. *ALEA Lat. Am. J. Probab. Math. Stat.*, 18(1):1–15, 2021.
- [DG77] Richard M. Dudley and Sam Gutmann. Stopping times with given laws. In *Séminaire de Probabilités, XI (Univ. Strasbourg, Strasbourg, 1975/1976)*, volume Vol. 581 of *Lecture Notes in Math.*, pages 51–58. Springer, Berlin-New York, 1977.
- [DGL96] Luc Devroye, László Györfi, and Gábor Lugosi. *A probabilistic theory of pattern recognition*, volume 31 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1996.
- [Dix84] Jacques Dixmier. *General topology*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1984. Translated from the French by Sterling K. Berberian.
- [DM75] Claude Dellacherie and Paul-André Meyer. *Probabilités et potentiel*, volume No. XV of *Publications de l’Institut de Mathématique de l’Université de Strasbourg*. Hermann, Paris, 1975. Chapitres I à IV, Édition entièrement refondue, Actualités Scientifiques et Industrielles, No. 1372. [Current Scientific and Industrial Topics].
- [DM93] Gianni Dal Maso. *An introduction to  $\Gamma$ -convergence*, volume 8 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser Boston, Inc., Boston, MA, 1993.
- [DMFP15] Anna De Masi, Pablo A. Ferrari, and Errico Presutti. Symmetric simple exclusion process with free boundaries. *Probab. Theory Related Fields*, 161(1-2):155–193, 2015.
- [DMFPSL19a] Anna De Masi, Pablo A. Ferrari, Errico Presutti, and Nahuel Soprano-Loto. Hydrodynamics of the N-BBM process. In Giambattista Giacomin, Stefano Olla, Ellen Saada, Herbert Spohn, and Gabriel Stoltz, editors, *Stochastic Dynamics Out of Equilibrium*, pages 523–549, Cham, 2019. Springer International Publishing.
- [DMFPSL19b] Anna De Masi, Pablo A. Ferrari, Errico Presutti, and Nahuel Soprano-Loto. Non local branching Brownian motions with annihilation and free boundary problems. *Electron. J. Probab.*, 24:Paper No. 63, 30, 2019.
- [DP15] Mark H. A. Davis and Martijn R. Pistorius. Explicit solution of an inverse first-passage time problem for Lévy processes and counterparty credit risk. *Ann. Appl. Probab.*, 25(5):2383–2415, 2015.
- [DR11] Rick Durrett and Daniel Remenik. Brunet-Derrida particle systems, free boundary problems and Wiener-Hopf equations. *Ann. Probab.*, 39(6):2043–2078, 2011.
- [Dud78] Richard M. Dudley. Central Limit Theorems for Empirical Measures. *The Annals of Probability*, 6(6):899 – 929, 1978.



- [Dur19] Rick Durrett. *Probability—theory and examples*, volume 49 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, fifth edition, 2019.
- [DZ10] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.
- [EEH14] Boris Ettinger, Steven N. Evans, and Alexandru Hening. Killed Brownian motion with a prescribed lifetime distribution and models of default. *Ann. Appl. Probab.*, 24(1):1–33, 2014.
- [EHW20] Boris Ettinger, Alexandru Hening, and Tak Kwong Wong. The inverse first passage time problem for killed Brownian motion. *Ann. Appl. Probab.*, 30(3):1251–1275, 2020.
- [EJ16] Erik Ekström and Svante Janson. The inverse first-passage problem and optimal stopping. *Ann. Appl. Probab.*, 26(5):3154–3177, 2016.
- [EMS13] Steven N. Evans, Ben Morris, and Arnab Sen. Coalescing systems of non-Brownian particles. *Probab. Theory Related Fields*, 156(1-2):307–342, 2013.
- [EVY20] Alison M. Etheridge, Amandine Véber, and Feng Yu. Rescaling limits of the spatial lambda-Fleming-Viot process with selection. *Electron. J. Probab.*, 25:Paper No. 120, 89, 2020.
- [Fas05] Antonio Fasano. *Mathematical models of some diffusive processes with free boundaries*, volume 11 of *MAT. Serie A: Conferencias, Seminarios y Trabajos de Matemática [MAT. Series A: Mathematical Conferences, Seminars and Papers]*. Universidad Austral, Facultad de Ciencias Empresariales (FCE-UA), Departamento de Matemática, Rosario, 2005.
- [Fel51] William Feller. Two singular diffusion problems. *Ann. of Math. (2)*, 54:173–182, 1951.
- [Fis37] Ronald A. Fisher. The wave of advance of advantageous genes. *Annals of Eugenics*, 7(4):355–369, 1937.
- [FM01] James Allen Fill and Motoya Machida. Stochastic monotonicity and realizable monotonicity. *Ann. Probab.*, 29(2):938–978, 2001.
- [FM07] Pablo A. Ferrari and Nevena Marić. Quasi stationary distributions and Fleming-Viot processes in countable spaces. *Electron. J. Probab.*, 12:no. 24, 684–702, 2007.
- [FSV<sup>+</sup>19] Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-ichi Amari, Alain Trounev, and Gabriel Peyré. Interpolating between Optimal Transport and MMD using Sinkhorn Divergences. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2681–2690. PMLR, 2019.
- [FV79] Wendell H. Fleming and Michel Viot. Some measure-valued Markov processes in population genetics theory. *Indiana Univ. Math. J.*, 28(5):817–843, 1979.

- [FZ20] Jean-Pierre Fouque and Zhaoyu Zhang. Deep learning methods for mean field control problems with delay. *Frontiers in Applied Mathematics and Statistics*, 6:11, 2020.
- [GH69] Richard J. Griego and Reuben Hersh. Random evolutions, Markov chains, and systems of partial differential equations. *Proc. Nat. Acad. Sci. U.S.A.*, 62:305–308, 1969.
- [GJ18] Pablo Groisman and Matthieu Jonckheere. Front propagation and quasi-stationary distributions for one-dimensional Lévy processes. *Electron. Commun. Probab.*, 23:Paper No. 93, 11, 2018.
- [GJ19] Pablo Groisman and Matthieu Jonckheere. Front propagation and quasi-stationary distributions: two faces of the same coin. In *Sojourns in probability theory and statistical physics. III. Interacting particle systems and random walks, a Festschrift for Charles M. Newman*, volume 300 of *Springer Proc. Math. Stat.*, pages 242–254. Springer, Singapore, [2019] ©2019.
- [GJM20] Pablo Groisman, Matthieu Jonckheere, and Julián Martínez. F-KPP scaling limit and selection principle for a Brunet-Derrida type particle system. *ALEA Lat. Am. J. Probab. Math. Stat.*, 17(1):589–607, 2020.
- [GK04] Ilie Grigorescu and Min Kang. Hydrodynamic limit for a Fleming-Viot type system. *Stochastic Process. Appl.*, 110(1):111–143, 2004.
- [GMST19] P. Jameson Graber, Alpár R. Mészáros, Francisco J. Silva, and Daniela Tonon. The planning problem in mean field games as regularized mass transport. *Calculus of Variations and Partial Differential Equations*, 58(3):1–28, 2019.
- [GMW22] Maximilien Germain, Joseph Mikael, and Xavier Warin. Numerical resolution of McKean-Vlasov FBSDEs using neural networks. *Methodology and Computing in Applied Probability*, pages 1–30, 2022.
- [GPW21] Maximilien Germain, Huyên Pham, and Xavier Warin. Neural networks-based algorithms for stochastic control and PDEs in finance. preprint, arXiv:2101.08068, 2021.
- [GSL21] Pablo Groisman and Nahuel Soprano-Loto. Rank dependent branching-selection particle systems. *Electron. J. Probab.*, 26:Paper No. 158, 27, 2021.
- [HCM07] Minyi Huang, Peter E. Caines, and Roland P. Malhamé. Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and decentralized  $\varepsilon$ -Nash equilibria. *IEEE transactions on automatic control*, 52(9):1560–1571, 2007.
- [HHK06] John W. Harris, Simon C. Harris, and Andreas 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.
- [HHL22] Jiequn Han, Ruimeng Hu, and Jihao Long. Learning high-dimensional McKean-Vlasov forward-backward stochastic differential equations with general distribution dependence. preprint, arXiv:2204.11924, 2022.

- [HJW17] Jiequn Han, Arnulf Jentzen, and E Weinan. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Communications in mathematics and statistics*, 5(4):349–380, 2017.
- [HJW18] Jiequn Han, Arnulf Jentzen, and E Weinan. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, 115(34):8505–8510, 2018.
- [HL22] Ruimeng Hu and Mathieu Laurière. Recent developments in machine learning methods for stochastic control and games. *Preprint. SSRN:4096569*, 2022.
- [HMC06] Minyi Huang, Roland P. Malhamé, and Peter E. Caines. Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. *Communications in Information & Systems*, 6(3):221–252, 2006.
- [HR17] Simon C. Harris and Matthew I. Roberts. The many-to-few lemma and multiple spines. *Ann. Inst. Henri Poincaré Probab. Stat.*, 53(1):226–242, 2017.
- [HT16] S. Herrmann and E. Tanré. The first-passage time of the Brownian motion to a curved boundary: an algorithmic approach. *SIAM J. Sci. Comput.*, 38(1):A196–A215, 2016.
- [HT23] Camilo Hernández and Ludovic Tangpi. Propagation of chaos for mean field Schrödinger problems. preprint, arXiv:2304.09340, 2023.
- [IM12] Kiyosi Itô and Henry P. McKean. *Diffusion Processes and their Sample Paths*. Classics in Mathematics. Springer Berlin Heidelberg, 2012.
- [INW68a] Nobuyuki Ikeda, Masao Nagasawa, and Shinzo Watanabe. Branching Markov processes I. *Journal of Mathematics of Kyoto University*, 8(2):233 – 278, 1968.
- [INW68b] Nobuyuki Ikeda, Masao Nagasawa, and Shinzo Watanabe. Branching Markov processes II. *Journal of Mathematics of Kyoto University*, 8(3):365 – 410, 1968.
- [JKZ09] Ken Jackson, Alexander Kreinin, and Wanhe Zhang. Randomization in the first hitting time problem. *Statist. Probab. Lett.*, 79(23):2422–2428, 2009.
- [Kal02] Olav Kallenberg. *Foundations of Modern Probability*. Springer, 2nd edition, 2002.
- [Kel72] Hans G. Kellerer. Markov-Komposition und eine Anwendung auf Martingale. *Math. Ann.*, 198:99–122, 1972.
- [Kes78] Harry Kesten. Branching Brownian motion with absorption. *Stochastic Process. Appl.*, 7(1):9–47, 1978.
- [Kir76] Bernd-Michael Kirstein. Monotonicity and comparability of time-homogeneous Markov processes with discrete state space. *Math. Operationsforsch. Statist.*, 7(1):151–168, 1976.
- [KK23] Alexander Klump and Martin Kolb. Uniqueness of the inverse first-passage time problem and the shape of the Shiryaev boundary. *Theory Probab. Appl.*, 67(4):570–592, 2023. Reprint of Teor. Veroyatn. Primen. **67** (2022), 717–744.

- [KK24] Alexander Klump and Martin Kolb. An elementary approach to the inverse first-passage-time problem for soft-killed Brownian motion. *J. Appl. Probab.*, 61(1):279–300, 2024.
- [KKO77] Teturo Kamae, Ulrich Krengel, and George L. O’Brien. Stochastic inequalities on partially ordered spaces. *Ann. Probability*, 5(6):899–912, 1977.
- [Klu23] Alexander Klump. The inverse first-passage time problem as hydrodynamic limit of a particle system. *Methodol. Comput. Appl. Probab.*, 25(1):Paper No. 42, 29, 2023.
- [Kol11] Vassili N. Kolokoltsov. Stochastic monotonicity and duality of one-dimensional Markov processes. *Mat. Zametki*, 89(5):694–704, 2011.
- [KP06] Andreas E. Kyprianou and Zbigniew Palmowski. Quasi-stationary distributions for Lévy processes. *Bernoulli*, 12(4):571–581, 2006.
- [KPP37] Andrey Kolmogorov, Ivan Petrovsky, and Nikolai Piskunov. Étude de l’équation de la diffusion avec croissance de la quantité de matière et son application à un problème biologique. *Bulletin de l’Université d’État à Moscou*, 1, 1937.
- [KS23] Alexander Klump and Mladen Savov. Conditions for existence and uniqueness of the inverse first-passage time problem applicable for Lévy processes and diffusions. preprint, arXiv:2305.10967, 2023.
- [Kun97] Hiroshi Kunita. *Stochastic flows and stochastic differential equations*, volume 24 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1997. Reprint of the 1990 original.
- [Kun19] Hiroshi Kunita. *Stochastic flows and jump-diffusions*, volume 92 of *Probability Theory and Stochastic Modelling*. Springer, Singapore, 2019.
- [LCST22] Guan-Horng Liu, Tianrong Chen, Oswin So, and Evangelos Theodorou. Deep generalized Schrödinger bridge. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022.
- [Lee18] Jimyeong Lee. A free boundary problem with non local interaction. *Math. Phys. Anal. Geom.*, 21(3):Paper No. 24, 21, 2018.
- [Lee20] Jimyeong Lee. Free boundary problems and biological systems with selection rules. *Arch. Math. (Basel)*, 114(1):85–95, 2020.
- [Ler86] Hans Rudolf Lerche. *Boundary crossing of Brownian motion*, volume 40 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin, 1986. Its relation to the law of the iterated logarithm and to sequential analysis.
- [LFL<sup>+</sup>20] Alex Tong Lin, Samy Wu Fung, Wuchen Li, Levon Nurbekyan, and Stanley J. Osher. Apac-net: Alternating the population and agent control via two neural networks to solve high-dimensional stochastic mean field games. preprint, arXiv:2002.10113, 2020.
- [Lig05] Thomas M. Liggett. *Interacting particle systems*. Classics in Mathematics. Springer-Verlag, Berlin, 2005. Reprint of the 1985 original.

- [Lin99] Torgny Lindvall. On Strassen’s theorem on stochastic domination. *Electron. Comm. Probab.*, 4:51–59, 1999.
- [LJR04] Yves Le Jan and Olivier Raimond. Flows, coalescence and noise. *Ann. Probab.*, 32(2):1247–1315, 2004.
- [LJR20] Yves Le Jan and Olivier Raimond. Flows, coalescence and noise. A correction. *Ann. Probab.*, 48(3):1592–1595, 2020.
- [LL06a] Jean-Michel Lasry and Pierre-Louis Lions. Jeux à champ moyen. I–le cas stationnaire. *Comptes Rendus Mathématique*, 343(9):619–625, 2006.
- [LL06b] Jean-Michel Lasry and Pierre-Louis Lions. Jeux à champ moyen. II–Horizon fini et contrôle optimal. *Comptes Rendus Mathématique*, 343(10):679–684, 2006.
- [LL07] Jean-Michel Lasry and Pierre-Louis Lions. Mean field games. *Japanese journal of mathematics*, 2(1):229–260, 2007.
- [LMS00] F. Javier López, Servet Martínez, and Gerardo Sanz. Stochastic domination and Markovian couplings. *Adv. in Appl. Probab.*, 32(4):1064–1076, 2000.
- [LP14] Mathieu Laurière and Olivier Pironneau. Dynamic programming for mean-field type control. *Comptes Rendus Mathématique*, 352(9):707–713, 2014.
- [LS92] Steven Lalley and Thomas Sellke. Limit theorems for the frontier of a one-dimensional branching diffusion. *Ann. Probab.*, 20(3):1310–1340, 1992.
- [Mai13] Pascal Maillard. The number of absorbed individuals in branching Brownian motion with a barrier. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(2):428–455, 2013.
- [Mai16] Pascal Maillard. Speed and fluctuations of  $N$ -particle branching Brownian motion with spatial selection. *Probab. Theory Related Fields*, 166(3-4):1061–1173, 2016.
- [McK66] Henry P. McKean. A class of Markov processes associated with nonlinear parabolic equations. *Proc. Nat. Acad. Sci. U.S.A.*, 56:1907–1911, 1966.
- [McK75] Henry P. McKean. Application of Brownian Motion to the Equation of Kolmogorov-Petrovskii-Piskunov. *Communications on Pure and Applied Mathematics*, XXVIII:323–331, 1975.
- [Mey75] Paul-André Meyer. Renaissance, recollements, mélanges, ralentissement de processus de Markov. *Ann. Inst. Fourier (Grenoble)*, 25(3-4):xxiii, 465–497, 1975.
- [Mik21] Toshio Mikami. Stochastic optimal transport revisited. *SN Partial Differential Equations and Applications*, 2(1):5, 2021.
- [MNW19] Gonzalo Mena and Jonathan Niles-Weed. Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. *Advances in neural information processing systems*, 32, 2019.
- [MS02] Alfred Müller and Dietrich Stoyan. *Comparison methods for stochastic models and risks*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2002.

- [MSM94] Servet Martínez and Jaime San Martín. Quasi-stationary distributions for a Brownian motion with drift and associated limit laws. *J. Appl. Probab.*, 31(4):911–920, 1994.
- [MT21] Quentin Merigot and Boris Thibert. Optimal transport: discretization and algorithms. In *Handbook of Numerical Analysis*, volume 22, pages 133–212. Elsevier, 2021.
- [MV12] Sylvie Méléard and Denis Villemonais. Quasi-stationary distributions and population processes. *Probab. Surv.*, 9:340–410, 2012.
- [OPS19] Carlo Orrieri, Alessio Porretta, and Giuseppe Savaré. A variational approach to the mean field planning problem. *Journal of Functional Analysis*, 277(6):1868–1957, 2019.
- [PC19] Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends<sup>®</sup> in Machine Learning*, 11(5-6):355–607, 2019.
- [Pes02] Goran Peskir. On integral equations arising in the first-passage problem for Brownian motion. *J. Integral Equations Appl.*, 14(4):397–423, 2002.
- [Pol84] David Pollard. *Convergence of Stochastic Processes*. Springer, 1984.
- [Por13] Alessio Porretta. On the planning problem for a class of mean field games. *Comptes Rendus Mathématique*, 351(11-12):457–462, 2013.
- [Por14] Alessio Porretta. On the planning problem for the mean field games system. *Dynamic Games and Applications*, 4(2):231–256, 2014.
- [Pot23] Yoann Potiron. Approximation convergence in the inverse first-passage time problem. preprint, arXiv:2106.11573, 2023.
- [PS06] Goran Peskir and Albert Shiryaev. *Optimal stopping and free-boundary problems*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2006.
- [PW01] Klaus Pötzelberger and Liqun Wang. Boundary crossing probability for Brownian motion. *J. Appl. Probab.*, 38(1):152–164, 2001.
- [Ray56] Daniel Ray. Stationary Markov processes with continuous paths. *Trans. Amer. Math. Soc.*, 82:452–493, 1956.
- [Roc70] R. Tyrrell Rockafellar. *Convex analysis*, volume 18. Princeton university press, 1970.
- [ROL<sup>+</sup>20] Lars Ruthotto, Stanley J. Osher, Wuchen Li, Levon Nurbekyan, and Samy Wu Fung. A machine learning framework for solving high-dimensional mean field game and mean field control problems. *Proceedings of the National Academy of Sciences*, 117(17):9183–9193, 2020.
- [RPK19] Maziar Raissi, Paris Perdikaris, and George E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

- [RY99] Daniel Revuz and Marc Yor. *Continuous Martingales and Brownian Motion*. Springer, 3rd edition, 1999.
- [Sal88] Paavo Salminen. On the first hitting time and the last exit time for a Brownian motion to/from a moving boundary. *Adv. in Appl. Probab.*, 20(2):411–426, 1988.
- [San15] Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 55(58-63):94, 2015.
- [Sch15] Erwin Schrödinger. Zur Theorie der Fall-und Steigversuche an Teilchen mit Brownscher Bewegung. *Physikalische Zeitschrift*, 16:289–295, 1915.
- [SK67] Richard Sinkhorn and Paul Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967.
- [SS18] Justin Sirignano and Konstantinos Spiliopoulos. DGM: A deep learning algorithm for solving partial differential equations. *Journal of computational physics*, 375:1339–1364, 2018.
- [Str65] Volker Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965.
- [Str67] Volker Strassen. Almost sure behavior of sums of independent random variables and martingales. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. II: Contributions to Probability Theory, Part 1*, pages 315–343. Univ. California Press, Berkeley, CA, 1967.
- [Szn91] Alain-Sol Sznitman. Topics in propagation of chaos. In *École d’Été de Probabilités de Saint-Flour XIX—1989*, volume 1464 of *Lecture Notes in Math.*, pages 165–251. Springer, Berlin, 1991.
- [TN22] Oliver Tough and James Nolen. The Fleming-Viot process with McKean-Vlasov dynamics. *Electron. J. Probab.*, 27:Paper No. 101, 72, 2022.
- [Tou23] Oliver Tough. Selection principle for the Fleming-Viot process with drift  $-1$ . preprint, arXiv:2306.03585, 2023.
- [TT13] Xiaolu Tan and Nizar Touzi. Optimal transportation under controlled stochastic dynamics. *The Annals of Probability*, 41(5):3201, 2013.
- [vD91] Erik A. van Doorn. Quasi-stationary distributions and convergence to quasi-stationarity of birth-death processes. *Adv. in Appl. Probab.*, 23(4):683–700, 1991.
- [vDP13] Erik A. van Doorn and Philip K. Pollett. Quasi-stationary distributions for discrete-state models. *European J. Oper. Res.*, 230(1):1–14, 2013.
- [Vil09a] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.
- [Vil09b] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.
- [Vil14] Denis Villemonais. General approximation method for the distribution of Markov processes conditioned not to be killed. *ESAIM Probab. Stat.*, 18:441–467, 2014.

- [Vil15] Denis Villemonais. Minimal quasi-stationary distribution approximation for a birth and death process. *Electron. J. Probab.*, 20:no. 30, 18, 2015.
- [Vil21] Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.
- [Wan13] Jie-Ming Wang. Stochastic comparison for Lévy-type processes. *J. Theoret. Probab.*, 26(4):997–1019, 2013.
- [Wer21] Florian Werner. Concatenation and pasting of right processes. *Electron. J. Probab.*, 26:Paper No. 50, 21, 2021.
- [WP97] Liqun Wang and Klaus Pötzelberger. Boundary crossing probability for Brownian motion and general boundaries. *J. Appl. Probab.*, 34(1):54–65, 1997.
- [Yag47] Akiva M. Yaglom. Certain limit theorems of the theory of branching random processes. *Doklady Akad. Nauk SSSR (N.S.)*, 56:795–798, 1947.
- [Yam23] Kosuke Yamato. Existence of quasi-stationary distributions for spectrally positive Lévy processes on the half-line. *ALEA Lat. Am. J. Probab. Math. Stat.*, 20(1):629–643, 2023.
- [Zha00] Shaoyi Zhang. Existence and application of optimal Markovian coupling with respect to non-negative lower semi-continuous functions. *Acta Math. Sin. (Engl. Ser.)*, 16(2):261–270, 2000.
- [ZS09] Cristina Zucca and Laura Sacerdote. On the inverse first-passage-time problem for a Wiener process. *Ann. Appl. Probab.*, 19(4):1319–1346, 2009.



L'objet central de cette thèse est un système de particules appelé  $N$ -processus de Markov branchant qui généralise le  $N$ -mouvement brownien branchant étudié par Maillard en 2016 : on part de  $N$  particules aux positions initiales  $x^1, \dots, x^N$ ; chaque particule se déplace alors sur la droite réelle suivant la loi d'un processus markovien homogène en temps donné (on parle de processus sous-jacent), indépendamment des autres trajectoires. De plus, chaque particule branche à taux 1 : c'est-à-dire qu'après une durée aléatoire de loi exponentielle  $\mathcal{E}(1)$ , celle-ci se duplique pour donner naissance à deux nouvelles particules à la même position. Enfin, simultanément à l'apparition d'une nouvelle particule dans le système à chaque instant de branchement, on retire du système la particule dont la position est la plus basse. Cette étape de sélection permet de conserver à tout instant une taille de population égale à  $N$ .

Nos principaux résultats concernent l'existence d'une limite hydrodynamique pour ce système de particules, c'est-à-dire l'existence d'une limite pour la fonction de répartition empirique associée à ce processus, quand la taille de population  $N$  tend vers l'infini. Sous certaines hypothèses de régularité, nous démontrons que cette fonction de répartition empirique converge vers la fonction de répartition de la loi du processus sous-jacent conditionné à ne pas avoir franchi une certaine frontière, cette dernière étant caractérisée comme solution du problème inverse du premier temps de passage avec loi exponentielle  $\mathcal{E}(1)$  pour le processus sous-jacent. Notre stratégie de preuve utilise un couplage avec des processus auxiliaires dont la construction fait appel à une hypothèse de monotonie stochastique sur le processus sous-jacent. En lien avec cette preuve, nous établissons dans un contexte général l'existence d'un couplage préservant l'ordre, entre plusieurs trajectoires d'un même processus de Markov satisfaisant l'hypothèse de monotonie stochastique et la propriété de Feller.

Enfin, nous nous intéressons par ailleurs à un second problème, celui du transport optimal en champ moyen, avec un point de vue entièrement numérique : nous développons trois méthodes d'apprentissage profond fournissant des solutions approchées de ce problème et les mettons en œuvre dans différents cas tests.



**INSTITUT DE RECHERCHE MATHÉMATIQUE AVANCÉE**  
UMR 7501  
Université de Strasbourg et CNRS  
7 Rue René Descartes  
67 084 STRASBOURG CEDEX

Tél. 03 68 85 01 29  
Fax 03 68 85 03 28  
<https://irma.math.unistra.fr>  
[irma@math.unistra.fr](mailto:irma@math.unistra.fr)

École doctorale

Mathématiques,  
sciences de l'information  
et de l'ingénieur | ED 269  
Université de Strasbourg



Institut de Recherche  
Mathématique Avancée

IRMA 2024/013  
<https://theses.hal.science/tel-04834834>

ISSN 0755-3390