



# Weak error analysis for time and particle discretizations of some stochastic differential equations non linear in the sense of McKean

Oumaima Bencheikh

## ► To cite this version:

Oumaima Bencheikh. Weak error analysis for time and particle discretizations of some stochastic differential equations non linear in the sense of McKean. Mathematics [math]. Université Paris-Est Marne la vallée, 2020. English. NNT: . tel-03129956

**HAL Id: tel-03129956**

**<https://pastel.hal.science/tel-03129956>**

Submitted on 3 Feb 2021

**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.

# Analyse de l'erreur faible de discrétisation en temps et en particules d'équations différentielles stochastiques non linéaires au sens de McKean

Oumaima Bencheikh

## ► To cite this version:

Oumaima Bencheikh. Analyse de l'erreur faible de discrétisation en temps et en particules d'équations différentielles stochastiques non linéaires au sens de McKean. Analyse numérique [math.NA]. Université Paris-Est, 2020. Français. NNT : 2020PESC1030 . tel-03129074

**HAL Id: tel-03129074**

**<https://pastel.archives-ouvertes.fr/tel-03129074>**

Submitted on 2 Feb 2021

**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.



ÉCOLE DOCTORALE : MATHÉMATIQUES ET SCIENCES  
ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

**THÈSE DE DOCTORAT**  
SPÉCIALITÉ : MATHÉMATIQUES APPLIQUÉES

présentée par  
**Oumaima BENCHEIKH**

---

**Analyse de l'erreur faible de discrétisation en temps et  
en particules d'EDS non linéaires au sens de McKean**

---

**Thèse dirigée par Benjamin JOURDAIN**  
**et préparée au CERMICS, École des Ponts ParisTech**

Soutenue le 22 octobre 2020 devant le Jury composé de :

|                                   |                           |                    |
|-----------------------------------|---------------------------|--------------------|
| <b>M. Benjamin JOURDAIN</b>       | École des Ponts ParisTech | Directeur de thèse |
| <b>M. Noufel FRIKHA</b>           | LPSM, Université de Paris | Rapporteur         |
| <b>M. Lukasz SZPRUCH</b>          | University of Edinburgh   | Rapporteur         |
| <b>Mme. Mireille BOSSY</b>        | INRIA Sophia Antipolis    | Examinatrice       |
| <b>M. J.-François CHASSAGNEUX</b> | LPSM, Université de Paris | Examineur          |
| <b>M. Bernard LAPEYRE</b>         | École des Ponts ParisTech | Examineur          |
| <b>M. Stéphane MENOZZI</b>        | Université d'Évry         | Examineur          |



À la mémoire de mon très cher et regretté Tonton Kebir.



## REMERCIEMENTS

Mes premiers remerciements s'adressent bien évidemment à mon directeur de thèse, Benjamin Jourdain. Je lui suis très reconnaissante pour sa confiance et sa patience à mon égard sans parler de sa constante disponibilité. Je tiens à souligner sa grande pédagogie, sa rigueur scientifique ainsi que son intégrité intellectuelle. Pendant ces trois années de thèse, j'ai énormément appris à ses côtés à la fois de son intuition mathématique, de son exigence de rigueur et de clarté dans les raisonnements mais aussi de l'étendue de ses connaissances. J'espère avoir été à la hauteur de ses attentes.

Je tiens à exprimer mes sincères remerciements à Noufel Frikha et Lukasz Szpruch pour l'intérêt qu'ils ont porté à mes travaux en acceptant d'être les rapporteurs de ma thèse. Je suis également très honorée de la présence de Mireille Bossy, Jean-François Chassagneux, Bernard Lapeyre et Stéphane Menozzi dans mon jury de thèse. Leurs travaux à tous ont beaucoup influencé mes recherches.

Je suis aussi très reconnaissante envers l'Université Mohammed VI Polytechnique pour avoir financé ma thèse.

Merci à mes anciens managers Simon Moreau et Fakher Ben Atig ainsi que tous mes collègues de l'équipe FIRST Credit de BNP Paribas Londres de m'avoir accompagnée dans mes premiers pas dans l'industrie financière, de m'avoir montré les problématiques passionnantes qu'offrent les mathématiques dans ce domaine et de m'avoir encouragée à poursuivre en thèse.

Mes remerciements vont à ceux qui m'ont enseigné les mathématiques au fur et à mesure de ma scolarité avec une pensée particulière à Monsieur Klilou en terminale ainsi qu'à Monsieur Taïbi en classes préparatoires. Je remercie aussi tous mes anciens professeurs pour avoir éveillé mon sens critique depuis mon jeune âge ainsi que pour tous leurs conseils qui me servent à mûrir et à grandir, aujourd'hui encore.

Enseigner à l'École des Ponts ParisTech a été une expérience plaisante et très formatrice. Je tiens en particulier à exprimer ma gratitude envers Bernard Lapeyre, Jean-Philippe Chancelier et Vincent Lefieux qui m'ont confié des rôles d'enseignements pour leurs cours de décision dans l'incertain et de statistiques.

Un grand merci à Isabelle Simunic, la fée du laboratoire qui a su simplifier toutes les situations compliquées. Sa bienveillance, son efficacité et sa grande sympathie ont grandement contribué au bon déroulement de cette thèse. Ces trois années au CERMICS ont été une expérience aussi profitable qu'agréable et je voudrais remercier l'ensemble des doctorants, post-doctorants et permanents du CERMICS que j'ai eu la chance de côtoyer. Sachez que j'ai énormément apprécié le temps que nous avons partagé tous ensemble au laboratoire et ailleurs, avec une mention spéciale à William et Alexandre pour nos discussions sérieuses et pour nos fou-rires interminables.

Aussi, j'adresse une grande pensée à tous mes amis, des études primaires à aujourd'hui. Il est évidemment risqué de se lancer dans une liste exhaustive mais que chaque personne qui lira ces lignes sache combien je suis reconnaissante de notre amitié. Je ne peux néanmoins omettre mes meilleures amies d'en-

fance Maha et Sara, ainsi que mes amis Sophia, Fatine, Rim, Nabil, Anas et Hiba. Je pense aussi à ma team Hoche des classes préparatoires et en particulier à Samia, Cécile et Soufiane. Je n'oublie surtout pas mes acolytes d'école d'ingénieur Nawfal, Samia, Othman, Najib, Amine et Marouane, et enfin ma colocataire Zaineb. Sachez que vous avez tous contribué à cette thèse par votre soutien permanent, vos encouragements et votre bienveillance à mon égard. Je suis heureuse de vous avoir à mes côtés et je pense à vous tous qui m'avez tant apporté.

Enfin, je terminerai par ma chère famille. Ces quelques lignes ne suffiront pas à traduire l'immense gratitude, la reconnaissance et l'affection que j'ai pour vous. Merci à mon Papa d'être mon exemple d'intégrité, de droiture et de courage. Merci à ma Maman d'être mon exemple de bonté, d'affection et d'altruisme. Merci à mes soeurs Asmae et Maria d'être mon exemple de pardon, d'entraide et de gentillesse. Votre amour, votre optimisme et votre générosité représentent pour moi une source constante de motivation et d'inspiration. Je vous dédie donc cette thèse car je vous dois tout. Un énorme merci pour la force que vous me donnez et pour tous les moments futurs que nous partagerons. J'adresse aussi une grande pensée à ma mamie Khiti qui ne m'a jamais oubliée dans ses prières et à tous mes grand-parents partis trop tôt, je pense à vous tous les jours.



---

**Titre :** Analyse de l'erreur faible de discrétisation en temps et en particules d'équations différentielles stochastiques non linéaires au sens de McKean

**Résumé :** Cette thèse est consacrée à l'étude théorique et numérique de l'erreur faible de discrétisation en temps et en particules d'Équations Différentielles Stochastiques non linéaires au sens de McKean. Nous abordons dans la première partie l'analyse de la vitesse faible de convergence de la discrétisation temporelle d'EDS standards. Plus spécifiquement, nous étudions la convergence en variation totale du schéma d'Euler-Maruyama appliqué à des EDS  $d$ -dimensionnelles avec un coefficient de dérive mesurable et un bruit additif. Nous obtenons, en supposant que le coefficient de dérive est borné, un ordre de convergence faible  $1/2$ . En rajoutant plus de régularité sur la dérive, à savoir une divergence spatiale au sens des distributions  $L^p$ -intégrable en espace uniformément en temps pour un certain  $\rho \geq d$ , nous atteignons un ordre de convergence égal à 1 (à un facteur logarithmique près) au temps terminal. En dimension 1, ce résultat est préservé lorsque la dérivée spatiale de la dérive est une mesure en espace avec une masse totale bornée uniformément en temps. Dans la deuxième partie de la thèse, nous analysons l'erreur faible de discrétisation à la fois en temps et en particules de deux classes d'EDS non-linéaires au sens de McKean. La première classe consiste en des EDS multi-dimensionnelles avec des coefficients de dérive et de diffusion réguliers dans lesquels la dépendance en loi intervient au travers de moments. La deuxième classe, quant à elle, consiste en des EDS uni-dimensionnelles avec un coefficient de diffusion constant et un coefficient de dérive singulier où la dépendance en loi intervient au travers de la fonction de répartition. Nous approchons les EDS par les schémas d'Euler-Maruyama des systèmes de particules associés et nous obtenons pour les deux classes un ordre de convergence faible égal à 1 en temps et en particules. Dans la seconde classe, nous prouvons aussi un résultat de propagation du chaos d'ordre optimal  $1/2$  en particules ainsi qu'un ordre fort de convergence égal à 1 en temps et  $1/2$  en particules. Tous nos résultats théoriques sont illustrés par des simulations numériques.

**Mots-clés :** Diffusions non linéaires au sens de McKean, schéma d'Euler-Maruyama, analyse de l'erreur faible, systèmes de particules en interaction, estimation du biais, propagation du chaos, interaction à champ moyen, particules interagissant à travers leur rang.

**Title:** Weak error analysis for time and particle discretizations of some stochastic differential equations non linear in the sense of McKean

**Summary:** This thesis is dedicated to the theoretical and numerical study of the weak error for time and particle discretizations of some Stochastic Differential Equations non linear in the sense of McKean. In the first part, we address the weak error analysis for the time discretization of standard SDEs. More specifically, we study the convergence in total variation of the Euler-Maruyama scheme applied to  $d$ -dimensional SDEs with additive noise and a measurable drift coefficient. We prove weak convergence with order  $1/2$  when assuming boundedness on the drift coefficient. By adding more regularity to the drift, namely the drift has a spatial divergence in the sense of distributions with  $\rho$ -th power integrable with respect to the Lebesgue measure in space uniformly in time for some  $\rho \geq d$ , the order of convergence at the terminal time improves to 1 up to some logarithmic factor. In dimension  $d = 1$ , this result is preserved when the spatial derivative of the drift is a measure in space with total mass bounded uniformly in time. In the second part of the thesis, we analyze the weak error for both time and particle discretizations of two classes of nonlinear SDEs in the sense of McKean. The first class consists in multi-dimensional SDEs with regular drift and diffusion coefficients in which the dependence in law intervenes through moments. The second class consists in one-dimensional SDEs with a constant diffusion coefficient and a singular drift coefficient where the dependence in law intervenes through the cumulative distribution function. We approximate the SDEs by the Euler-Maruyama schemes of the associated particle systems and obtain for both classes a weak order of convergence equal to 1 in time and particles. We also prove, for the second class, a trajectorial propagation of chaos result with optimal order  $1/2$  in particles as well as a strong order of convergence equal to 1 in time and  $1/2$  in particles. All our theoretical results are illustrated by numerical experiments.

**Keywords:** Diffusions non linear in the sense of McKean, Euler-Maruyama scheme, weak error analysis, interacting particle systems, bias estimation, propagation of chaos, mean-field interaction, rank-based model.

|          |                                                                                                                                                        |           |
|----------|--------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>Introduction</b>                                                                                                                                    | <b>1</b>  |
| 1.1      | Équations différentielles stochastiques standards                                                                                                      | 1         |
| 1.1.1    | Discrétisation temporelle                                                                                                                              | 2         |
| 1.1.2    | Analyse de l'erreur forte de discrétisation en temps                                                                                                   | 2         |
| 1.1.3    | Analyse de l'erreur faible de discrétisation en temps                                                                                                  | 3         |
| 1.1.4    | Méthodes de Monte-Carlo multi-pas                                                                                                                      | 4         |
| 1.1.5    | Contribution de la thèse                                                                                                                               | 6         |
| 1.2      | Équations différentielles stochastiques de type McKean-Vlasov                                                                                          | 7         |
| 1.2.1    | Propagation du chaos et analyse de l'erreur forte de discrétisation en particules                                                                      | 8         |
| 1.2.2    | Analyse de l'erreur faible de discrétisation en particules : Biais                                                                                     | 9         |
| 1.2.3    | Échantillonnage antithétique                                                                                                                           | 11        |
| 1.2.4    | Discrétisation temporelle                                                                                                                              | 12        |
| 1.2.5    | Contributions de la thèse                                                                                                                              | 13        |
| <b>I</b> | <b>Standard Stochastic differential Equations</b>                                                                                                      | <b>17</b> |
| <b>2</b> | <b>Convergence in total variation of the Euler-Maruyama scheme applied to diffusion processes with measurable drift coefficient and additive noise</b> | <b>19</b> |
| 2.1      | Introduction                                                                                                                                           | 20        |
| 2.2      | Main results                                                                                                                                           | 23        |
| 2.2.1    | Application to one-dimensional SDEs with non-constant diffusion coefficient                                                                            | 25        |
| 2.2.2    | Existence of densities and mild equations                                                                                                              | 27        |
| 2.3      | Proof of the convergence rate in total variation                                                                                                       | 29        |
| 2.4      | Proof of the convergence rate in total variation when assuming more regularity on $b$ w.r.t. to the space variables                                    | 32        |
| 2.4.1    | Proof of Proposition 2.4.1                                                                                                                             | 33        |
| 2.4.2    | Proof of Theorem 2.2.3                                                                                                                                 | 40        |
| 2.4.3    | Proof of Proposition 2.2.5                                                                                                                             | 43        |
| 2.5      | Numerical Experiments                                                                                                                                  | 45        |
| 2.5.1    | The specific case $\alpha = -\beta = \theta > 0$                                                                                                       | 45        |
| 2.5.2    | General case                                                                                                                                           | 49        |
| 2.5.3    | Conclusion                                                                                                                                             | 52        |
| A.1      | Appendix                                                                                                                                               | 53        |

|           |                                                                                                                                             |           |
|-----------|---------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>II</b> | <b>McKean-Vlasov Stochastic differential Equations</b>                                                                                      | <b>57</b> |
| <b>3</b>  | <b>Bias estimation and antithetic sampling for systems of particles interacting through moments</b>                                         | <b>59</b> |
| 3.1       | Introduction . . . . .                                                                                                                      | 60        |
| 3.2       | Estimation of the bias for systems of particles interacting through moments . . . . .                                                       | 60        |
| 3.3       | Numerical Experiments . . . . .                                                                                                             | 68        |
| 3.3.1     | Generalised Ornstein-Uhlenbeck process . . . . .                                                                                            | 68        |
| 3.3.2     | Plane Rotator . . . . .                                                                                                                     | 71        |
| 3.3.3     | Polynomial Drift . . . . .                                                                                                                  | 72        |
| 3.3.4     | Viscous Burgers equation . . . . .                                                                                                          | 74        |
| <b>4</b>  | <b>Weak and strong error analysis for mean field rank-based particle approximations of one dimensional viscous scalar conservation laws</b> | <b>77</b> |
| 4.1       | Introduction . . . . .                                                                                                                      | 78        |
| 4.2       | Main results . . . . .                                                                                                                      | 81        |
| 4.3       | Dynamics of the reordered particle system and mild formulations . . . . .                                                                   | 84        |
| 4.4       | Proofs of Section 2 . . . . .                                                                                                               | 88        |
| 4.4.1     | Quantitative propagation of chaos result . . . . .                                                                                          | 88        |
| 4.4.2     | Rate of convergence of the strong $L^1$ -error . . . . .                                                                                    | 90        |
| 4.4.3     | Estimation of the bias . . . . .                                                                                                            | 104       |
| 4.5       | Particle initialization . . . . .                                                                                                           | 109       |
| 4.5.1     | Finite Wassertein distance . . . . .                                                                                                        | 109       |
| 4.5.2     | Strong errors of order $N^{-1/2}$ . . . . .                                                                                                 | 112       |
| 4.5.3     | Further properties of the optimal deterministic initialization . . . . .                                                                    | 117       |
| 4.6       | Numerical experiments for the Burgers equation . . . . .                                                                                    | 119       |
| 4.6.1     | Strong $L^1$ -error behaviour . . . . .                                                                                                     | 121       |
| 4.6.2     | Weak $L^1$ -error behaviour . . . . .                                                                                                       | 121       |
| B.2       | Appendix . . . . .                                                                                                                          | 123       |

## 1.1 Équations différentielles stochastiques standards

Nombreux sont les phénomènes physiques qui peuvent être décrits en faisant intervenir des forces dites aléatoires. C'est le cas notamment quand il s'agit de modéliser la cinétique d'une réaction chimique, ou bien la dynamique de particules colloïdales en solution ou encore même les évolutions des marchés financiers. Pour ce faire, nous faisons appel aux Équations Différentielles Stochastiques que l'on considère comme étant des perturbations aléatoires autour d'une Équation Différentielle Ordinaire. Nous disposons donc d'un modèle mathématique décrivant des systèmes faisant intervenir deux types de forces, l'une déterministe et l'autre aléatoire.

On se place sur un espace de probabilité complet que l'on note  $(\Omega, \mathcal{F}, \mathbb{P})$  et l'on se donne  $(W_t)_{t \geq 0}$  un mouvement brownien  $d$ -dimensionnel sur cet espace. On considère également une variable aléatoire  $X_0 \in \mathbb{R}^n$  indépendante du mouvement brownien  $(W_t)_{t \geq 0}$ . Soit  $\mathcal{F}_t = \sigma(X_0, (W_s)_{s \leq t})$  la plus petite filtration contenant la filtration naturelle du mouvement brownien  $(W_t)_{t \geq 0}$  et  $\sigma(X_0)$ . Soit  $T$  un horizon temporel fini et  $b : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  et  $\sigma : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times d}$  deux fonctions mesurables. Nous nous intéressons à l'équation différentielle stochastique  $n$ -dimensionnelle suivante :

$$dX_t = b(t, X_t) dt + \sigma(t, X_t) dW_t, \quad t \in [0, T] \quad (1.1.1)$$

avec valeur initiale  $X_0$ . La fonction  $b$  représente la dérive et  $\sigma$  la diffusion. On appelle solution de l'EDS (1.1.1) un processus  $(X_t)_{t \in [0, T]}$  continu  $\mathcal{F}_t$ -adapté à valeurs dans  $\mathbb{R}^n$  et tel que :

- presque sûrement,  $\int_0^T (|b(s, X_s)| + |\sigma(s, X_s)|^2) ds < +\infty$ .
- presque sûrement,  $\forall t \in [0, T], \quad X_t = X_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s$ .

Il existe plusieurs résultats d'existence et d'unicité de la solution de l'EDS (1.1.1) suivant la régularité des coefficients  $b$  et  $\sigma$ . Le principal résultat prouvé par Itô [39] est l'analogue du théorème de Cauchy-Lipschitz pour les équations différentielles ordinaires, et s'obtient sous des hypothèses similaires sur les coefficients de l'équation :

**Theorem 1.1.1.** *On suppose que :*

$$\exists K > 0, \forall t \in [0, T], \quad \begin{cases} \forall x \in \mathbb{R}^d, |b(t, x)| + |\sigma(t, x)| \leq K(1 + |x|) \\ \forall x, y \in \mathbb{R}^d, |b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq K|x - y| \end{cases}$$

Alors l'EDS (1.1.1) admet une unique solution  $(X_t)_{t \in [0, T]}$ .

En se plaçant dans un cadre beaucoup moins régulier, Zvonkin [91] et Veretennikov [87, 88] prouvent l'existence et l'unicité d'une solution quand  $b$  est bornée et  $\sigma$  lipschitzienne bornée et uniformément non dégénérée par rapport aux variables spatiales. Veretennikov [89] étend ce résultat par la suite en permettant qu'une partie de la diffusion soit dégénérée.

D'une manière générale, cette solution qu'on appelle processus de diffusion peut être vue comme un processus dont le comportement local est celui d'un mouvement brownien dont la variance dépend de la position et du temps. Cette approche descriptive locale est celle adaptée par les algorithmes de résolution numérique dont le plus simple est l'algorithme d'Euler-Maruyama.

### 1.1.1 Discrétisation temporelle

En absence de solution explicite, on construit un schéma d'approximation sur l'intervalle  $[0, T]$ . On se donne une subdivision  $(t_k = kh)_{k \in \llbracket 0, \lfloor \frac{T}{h} \rfloor \rrbracket}$  où  $h$  représente le pas de temps et on approxime notre solution aux points de la subdivision par le schéma d'Euler-Maruyama défini par :

$$X_{t_{k+1}}^h = X_{t_k}^h + b(t_k, X_{t_k}^h) h + \sigma(t_k, X_{t_k}^h) (W_{t_{k+1}} - W_{t_k}). \quad (1.1.2)$$

Ce schéma est bien-sûr analogue au schéma d'Euler pour les équations différentielles ordinaires. Pour passer d'un instant de discrétisation à l'autre, on fige la dérive et la diffusion à leur valeur en  $t_k$  et l'implémentation numérique de ce schéma revient donc à simuler les accroissements browniens qui sont indépendants et identiquement distribués suivant une loi gaussienne multidimensionnelle centrée et de variance  $hI_d$  où  $I_d$  désigne la matrice identité de dimension  $d$ .

Pour effectuer les preuves de convergence, nous introduisons l'interpolation temporelle continue du schéma d'Euler-Maruyama définie par :

$$X_t^h = X_0 + \int_0^t b(\tau_s^h, X_{\tau_s^h}^h) ds + \int_0^t \sigma(\tau_s^h, X_{\tau_s^h}^h) dW_s, \quad t \in [0, T]; \quad (1.1.3)$$

où pour  $s \in (0, t]$ , on définit  $\tau_s^h = \lfloor s/h \rfloor h$  comme étant le dernier instant de discrétisation avant  $s$ .

### 1.1.2 Analyse de l'erreur forte de discrétisation en temps

Les propriétés de convergence du schéma d'Euler-Maruyama ont été longuement étudiées. Quand les coefficients  $b$  et  $\sigma$  sont lipschitziens par rapports aux variables spatiales et  $\alpha$ -höldériens en temps, on obtient une vitesse forte, au plus, en  $\mathcal{O}(\sqrt{h})$  i.e. :

$$\forall p \geq 1, \exists C_p < +\infty, \forall h \in (0, T], \quad \mathbb{E} \left( \sup_{t \leq T} \|X_t - X_t^h\|^{2p} \right)^{1/2p} \leq C_p h^\gamma, \quad \text{où } \gamma = \min(1/2, \alpha).$$

En particulier, lorsque l'EDS est homogène en temps ou bien  $\alpha \geq 1/2$ , la vitesse forte de convergence du schéma d'Euler Maruyama est en  $\sqrt{h}$ . On définit  $\gamma$  comme étant l'ordre de convergence du schéma.

Si les propriétés de convergence forte du schéma d'Euler-Maruyama sont bien comprises pour les EDS à coefficients réguliers, le cas des coefficients irréguliers est toujours un domaine de recherche actif. Le premier résultat remonte à Gyöngy et Krylov [31] qui ont établi la convergence en probabilité (sans vitesse de convergence) lorsque les coefficients sont continus en espace avec unicité trajectorielle de l'EDS. Depuis, une multitude de résultats ont vu le jour et les travaux les plus récents portent sur les EDS dont le coefficient de dérive est lipschitzien par morceaux et le coefficient de diffusion est, quant à lui, globalement lipschitzien satisfaisant une condition de non-dégénérescence sur l'hypersurface de discontinuité de la dérive. Leobacher et Szölgényi [65] prouvent dans ce cas une convergence  $L^2$  d'ordre  $1/4-$  (signifiant  $1/4 - \epsilon$  pour un  $\epsilon > 0$  arbitrairement petit). En dimension  $d = 1$ , Müller-Gronbach et Yaroslavl'tseva [71] récupèrent pour chaque  $p \in [1, +\infty)$  un ordre de convergence  $L^p$  égal à  $1/2$ . En dimension supérieure, cet ordre  $1/2$  (à un facteur logarithmique près) est prouvé par Neuenkirch, Szölgényi et Szpruch [73] pour l'erreur  $L^2$  du schéma d'Euler-Maruyama avec pas de temps adaptatif. Quand le coefficient de diffusion est constant, Dareiotis et Gerencsér [16, 74] obtiennent une convergence  $L^2$  d'ordre  $1/2-$  pour un coefficient de dérive borné, Dini-continu et homogène en temps. Cet ordre est préservé en dimension

1 en assouplissant la Dini-continuité à de la simple mesurabilité.

Pour des coefficients réguliers, il est possible d'améliorer l'ordre de convergence forte du schéma d'Euler-Maruyama en y introduisant des termes supplémentaires. On obtient alors le schéma de Milstein démarrant de  $X_0$  et défini par :

$$\begin{aligned} \tilde{X}_{t_{k+1}}^h &= \tilde{X}_{t_k}^h + b(t_k, \tilde{X}_{t_k}^h) h + \sigma(t_k, \tilde{X}_{t_k}^h) (W_{t_{k+1}} - W_{t_k}) \\ &+ \frac{1}{2} \sum_{j=1}^d \sum_{m=1}^d \partial \sigma_j \sigma_m(t_k, \tilde{X}_{t_k}^h) \left( (W_{t_{k+1}}^j - W_{t_k}^j) (W_{t_{k+1}}^m - W_{t_k}^m) - \mathbf{1}_{\{j=m\}} h + A_{t_{k+1}}^{j,m} \right) \end{aligned} \quad (1.1.4)$$

où  $\partial \sigma_j$  représente la matrice jacobienne du champ de vecteurs  $\sigma_j$  définie par  $\partial \sigma_j = \left( \frac{\partial \sigma_{ij}}{\partial x_k} \right)_{1 \leq i, k \leq d}$  et  $A_{t_{k+1}}^{j,m}$  représente l'aire de Lévy définie par :

$$A_{t_{k+1}}^{j,m} = \int_{t_k}^{t_{k+1}} (W_s^m - W_{t_k}^m) dW_s^j - \int_{t_k}^{t_{k+1}} (W_s^j - W_{t_k}^j) dW_s^m.$$

La vitesse de convergence forte de ce schéma est en  $h$  au lieu de  $\sqrt{h}$  mais ce gain a un coût. En dimension supérieure à 1, le schéma fait intervenir les aires de Lévy qu'on ne sait simuler qu'en dimension 2 uniquement (cf. travaux de Gaines et Lyons [23]).

### 1.1.3 Analyse de l'erreur faible de discrétisation en temps

Une quantité utile à étudier, qui apparaît dans de nombreux domaines d'application allant de la physique à la finance, est l'erreur dite faible qui, pour une fonction de test appropriée  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  s'écrit :

$$\mathbb{E}[f(X_t)] - \mathbb{E}[f(X_t^h)].$$

Il existe une littérature très riche concernant la vitesse faible de convergence du schéma d'Euler-Maruyama pour des coefficients réguliers et/ou non-dégénérés. Le premier résultat apparaît dans l'article fondateur de Talay et Tubaro [85]:

**Theorem 1.1.2.** *On suppose que les coefficients  $b$  et  $\sigma$  sont  $C^\infty$  sur  $[0, T] \times \mathbb{R}^n$  avec des dérivées à tous ordres bornées et que  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  est  $C^\infty$  avec des dérivées à croissance polynomiale. Alors il existe une suite de réels  $(C_\ell)_{\ell \geq 1}$  telle que pour tout  $L \in \mathbb{N}^*$ , on obtient un développement de l'erreur de la forme :*

$$\mathbb{E}[f(X_t)] - \mathbb{E}[f(X_t^h)] = C_1 h + C_2 h^2 + \dots + C_L h^L + \mathcal{O}(h^{L+1}).$$

On en déduit que :

$$\exists C < +\infty, \forall h \in (0, T], \quad |\mathbb{E}[f(X_t)] - \mathbb{E}[f(X_t^h)]| \leq Ch.$$

Ce dernier résultat reste vrai en supposant simplement que les coefficients  $b$  et  $\sigma$  sont  $C^3$  avec des dérivées bornées et  $f$  une fonction  $C^3$  dont les dérivées sont à croissance polynomiale. Dans le cadre non-dégénéré (sous certaines conditions d'ellipticité ou d'hypoellipticité uniformes), Guyon [30] montre qu'il est même possible de choisir  $f$  comme une distribution tempérée où  $\mathbb{E}[f(X_t)]$  (respectivement  $\mathbb{E}[f(X_t^h)]$ ) s'interprète alors comme l'action de la distribution sur la densité de  $X_t$  (respectivement  $X_t^h$ ), et de conserver une vitesse faible de convergence en  $\mathcal{O}(h)$ . Quoi qu'il en soit, il existe une abondance de résultats dans le cadre régulier.

Le cas des coefficients non-réguliers, quant à lui, a rarement été envisagé. Dans le cadre de coefficients bornés non-dégénérés, Mikulevicius et Platen [68] prouvent que, sous ellipticité uniforme, lorsque les coefficients sont  $\alpha$ -höldériens par rapport aux variables spatiales et  $\alpha/2$ -höldériens par rapport à la variable temporelle avec  $\alpha \in (0, 1)$ ,  $\mathbb{E}[f(X_T)]$  est approximée avec un ordre  $\frac{\alpha}{2}$  par  $\mathbb{E}[f(X_T^h)]$  quand la fonction test  $f$  est  $C^2$  avec des dérivées secondes  $\alpha$ -höldériennes. Dans le cas de bruit additif (i.e. le coefficient de diffusion est constant), Kohatsu-Higa, Lejay et Yasuda [58], prouvent que pour  $f$  de classe  $C^3$  avec des dérivées à croissance polynomiale, la convergence se maintient avec un ordre  $1/2$ — lorsque  $n = d \geq 2$  (resp.  $1/3$ — quand  $n = d = 1$ ) avec une dérive homogène en temps, bornée et lipschitzienne

sauf sur un ensemble  $G$  tel que  $\varepsilon^{-d}$  fois la mesure de Lebesgue de  $\{x \in \mathbb{R}^d : \inf_{y \in G} |x - y| \leq \varepsilon\}$  est bornée.

Une autre quantité qui traduit également l'erreur faible et qui est tout aussi utile à étudier est la différence entre les densités de  $X_t$  et  $X_t^h$  par rapport à la mesure de Lebesgue qu'on désigne respectivement par  $p(t, \cdot)$  et  $p^h(t, \cdot)$  lorsque celles-ci existent :

$$|p(t, x) - p^h(t, x)|.$$

Cette erreur correspond à  $(\mathbb{E}[f(X_t)] - \mathbb{E}[f(X_t^h)])$  en choisissant  $f$  comme une masse de Dirac. Dans le cadre de coefficients très réguliers, Gobet et Labart [27] prouvent que la valeur absolue de la différence des densités est bornée par une densité gaussienne multipliée par un facteur d'ordre 1. Ils supposent que les coefficients  $b$  et  $\sigma$  sont continûment différentiables et bornés avec des dérivées temporelles à l'ordre 1 uniformément bornées et des dérivées spatiales jusqu'à l'ordre 3 uniformément bornées. De plus, la diffusion  $\sigma$  est uniformément elliptique et sa dérivée temporelle à l'ordre 1 est aussi une fonction continûment différentiable bornée avec sa dérivée spatiale à l'ordre 1 uniformément bornée.

Dans un cadre moins régulier, Konakov et Menozzi [60] prouvent que la valeur absolue de la différence des densités est bornée par une densité gaussienne multipliée par un facteur d'ordre  $\frac{\alpha}{2}$  — lorsque les coefficients sont uniformément elliptiques, bornés et  $\frac{\alpha}{2}$ -höldériens en temps et  $\alpha$ -höldériens en espace. Frikha [22] traite des EDS unidimensionnelles homogènes en temps, impliquant éventuellement un terme de temps local en plus d'un coefficient de dérive borné mesurable et d'un coefficient de diffusion borné uniformément elliptique et  $\alpha$ -höldérien. Il prouve que la valeur absolue de la différence des densités est inférieure à une densité gaussienne multipliée par un facteur d'ordre  $\frac{\alpha}{2}$ .

Les derniers travaux dans ce domaine sont ceux de Suo, Yuan et Zhang [83]. Ils étudient dans un cadre multidimensionnel des EDS avec bruit additif et un coefficient de dérive homogène en temps avec au plus une croissance linéaire et satisfaisant une condition de régularité de type  $\alpha$ -höldérienne intégrée contre une certaine mesure gaussienne. Lorsque ce coefficient a une croissance sous-linéaire (et sous certaines restrictions sur l'horizon temporel lorsqu'il a une croissance linéaire), ils prouvent une convergence de la norme  $L^1$  de la différence des densités, qui représente une convergence en variation totale, d'ordre  $\frac{\alpha}{2}$ .

#### 1.1.4 Méthodes de Monte-Carlo multi-pas

Les simulations Monte-Carlo sont largement utilisées afin de déterminer la valeur de  $\mathbb{E}[f(X_T)]$ . L'estimateur le plus simple de cette quantité reste la moyenne des valeurs  $f(X_T^h)$  à partir de la simulation de  $X_T^{r,h}$  du schéma de  $R$  trajectoires indépendantes. En notant  $\hat{Y}$  l'estimateur, on a  $\hat{Y} = \frac{1}{R} \sum_{r=1}^R f(X_T^{r,h})$ . Sous réserve de régularité sur les coefficients de dérive  $b$  et de diffusion  $\sigma$  ainsi que la fonction  $f$ , l'erreur quadratique moyenne (MSE) de l'estimateur  $\hat{Y}$  qui est égale à  $\mathbb{E} \left[ \left( \hat{Y} - \mathbb{E}[f(X_T)] \right)^2 \right]$  est de la forme asymptotique suivante :

$$\text{MSE} \simeq c_1 R^{-1} + c_2 h^2$$

où  $c_1$  et  $c_2$  sont des constantes positives. Le premier terme correspond à la variance de  $\hat{Y}$  issue de l'échantillonnage Monte-Carlo et le second terme correspond au carré de l'erreur faible introduite par la discrétisation du schéma d'Euler-Maruyama. Ainsi pour une précision  $\epsilon$ , une erreur quadratique moyenne en  $\mathcal{O}(\epsilon^2)$  nécessite que  $R = \mathcal{O}(\epsilon^{-2})$  et  $h = \mathcal{O}(\epsilon)$ . Nous obtenons donc un coût en  $\mathcal{O}(\epsilon^{-3})$ .

Giles [24] propose une méthode de Monte-Carlo multi-pas (MLMC) qui permet de réduire le coût à  $\mathcal{O}(\epsilon^{-2} \log(\epsilon)^2)$  voire  $\mathcal{O}(\epsilon^{-2})$ . Dans sa forme la plus générale, la méthode de Monte-Carlo multi-pas utilise un nombre de niveaux de discrétisation  $\ell = 0, 1, \dots, L$  avec  $\ell = 0$  représentant le niveau le plus grossier et  $\ell = L$  le niveau le plus fin de discrétisation. On choisit une suite géométrique de différents pas de temps  $(h_\ell = 2^{-\ell} T)_{0 \leq \ell \leq L}$  avec  $h_L$  le plus petit pas de temps correspondant au pas de temps  $h$  initial, qui détermine la taille du biais issu de la discrétisation d'Euler-Maruyama. La construction de l'estimateur MLMC repose alors sur l'égalité suivante :

$$\mathbb{E} \left[ f(X_T^{h_L}) \right] = \mathbb{E} \left[ f(X_T^{h_0}) \right] + \sum_{\ell=1}^L \mathbb{E} \left[ f(X_T^{h_\ell}) - f(X_T^{h_{\ell-1}}) \right].$$



Il s'agit donc d'estimer indépendamment chacune des espérances des différences de manière à minimiser la variance totale pour un coût de calcul donné. On peut alors utiliser l'estimateur suivant :

$$\hat{Y} = \frac{1}{R_0} \sum_{r=1}^{R_0} f(X_T^{r,h_0}) + \sum_{\ell=1}^L \left\{ \frac{1}{R_\ell} \sum_{r=1}^{R_\ell} \left( f(X_T^{\ell,r,h_\ell}) - f(X_T^{\ell,r,h_{\ell-1}}) \right) \right\}$$

où  $R_\ell \in \mathbb{N}^*$  correspond au nombre de tirages pour le niveau de discrétisation  $\ell \in \{0, \dots, L\}$  et  $X_T^{\ell,r,h_\ell}$  représente la  $r$ -ème approximation au niveau  $\ell$  de  $X_T$  parmi  $R_\ell$  tirages. L'essence de cette approximation est que la quantité  $f(X_T^{\ell,r,h_\ell}) - f(X_T^{\ell,r,h_{\ell-1}})$  provient de deux approximations discrètes avec différents pas de temps mais à partir de la même trajectoire brownienne. On implémente donc facilement cette méthode en commençant par construire les accroissements browniens pour la simulation de la trajectoire discrète permettant d'évaluer  $f(X_T^{\ell,r,h_\ell})$  et en les additionnant ensuite par groupes de 2 afin d'obtenir les accroissements browniens discrets permettant d'évaluer  $f(X_T^{\ell,r,h_{\ell-1}})$ . On rappelle ici le Théorème de Giles [24] qui donne la complexité de l'estimation MLMC :

**Theorem 1.1.3.** *Soit  $f(X_T)$  une fonctionnelle de la solution d'une EDS et  $f(X_T^{\ell,h_\ell})$  son approximation au niveau  $\ell$ . On suppose qu'il existe des estimateurs indépendants  $Y_\ell = \frac{1}{R_\ell} \sum_{r=1}^{R_\ell} \left( f(X_T^{\ell,r,h_\ell}) - f(X_T^{\ell,r,h_{\ell-1}}) \right)$  issus de  $R_\ell$  tirages Monte-Carlo et des constantes positives  $\alpha, \beta, \gamma, c_1, c_2, c_3$  telles que  $\alpha \geq \min(\beta, \gamma)$  et :*

1.  $\left| \mathbb{E} \left[ f(X_T^{\ell,h_\ell}) - f(X_T) \right] \right| \leq c_1 h_\ell^\alpha,$
2.  $\mathbb{E}[Y_\ell] = \begin{cases} \mathbb{E} \left[ f(X_T^{h_0}) \right], & \ell = 0 \\ \mathbb{E} \left[ f(X_T^{\ell,h_\ell}) - f(X_T^{\ell,h_{\ell-1}}) \right], & \ell > 0, \end{cases}$
3.  $V_\ell \leq c_2 R_\ell^{-1} h_\ell^\beta$  où  $V_\ell$  fait référence à la variance de l'échantillon  $Y_\ell$ ,
4.  $C_\ell \leq c_3 R_\ell h_\ell^{-\gamma}$  où  $C_\ell$  fait référence à la complexité de  $Y_\ell$ .

Alors il existe une constante positive  $c_4$  telle que pour tout  $\epsilon < e^{-1}$ , il y a des valeurs  $L$  et  $R_\ell$  pour lesquelles l'estimateur Monte Carlo multi-pas  $\hat{Y} = \sum_{\ell=0}^L Y_\ell$  admet une erreur quadratique moyenne bornée strictement par  $\epsilon^2$  et sa complexité  $C$  est bornée par :

$$C \leq \begin{cases} c_4 \epsilon^{-2}, & \beta > \gamma \\ c_4 \epsilon^{-2} \log(\epsilon)^2, & \beta = \gamma \\ c_4 \epsilon^{-2-(\gamma-\beta)/\alpha}, & 0 < \beta < \gamma. \end{cases}$$

En effet, pour assurer un biais inférieur à  $\frac{\epsilon}{2}$ , il suffit de choisir  $L$  tel que :  $L = \frac{\log(\epsilon^{-1})}{\log(2)} + \mathcal{O}(1)$  quand  $\epsilon \rightarrow 0$ . Comme la variance totale de l'estimateur  $\hat{Y}$  est  $\sum_{\ell=0}^L R_\ell^{-1} V_\ell$  alors le coût de calcul est proportionnel à  $\sum_{\ell=0}^L R_\ell h_\ell^{-1}$ . Bien qu'on utilise différents niveaux avec différentes erreurs de discrétisation pour estimer l'espérance de  $f(X_T^h)$ , le biais dépend de la précision du niveau le plus fin  $L$ . Pour un schéma d'ordre fort 1/2, dans le cas d'un schéma d'Euler par exemple où  $\alpha = \gamma = \beta = 1$ , on a une complexité en  $\mathcal{O}(\epsilon^{-2} \log(\epsilon)^2)$ . En revanche, pour un schéma d'ordre fort supérieur strictement à 1/2, on obtient une complexité en  $\mathcal{O}(\epsilon^{-2})$  comme pour une méthode de Monte Carlo sans biais.

Il est maintenant clair que l'efficacité de la méthode MLMC repose sur l'ordre de convergence forte de la discrétisation. Utiliser le MLMC avec la discrétisation de Milstein qui atteint l'ordre fort 1 pour les EDS scalaires permet d'améliorer la complexité en se débarrassant du  $(\log(\epsilon)^2)$ . Toutefois, ce gain en

convergence n'est pas sans contraintes. En effet, le schéma de Milstein nécessite la simulation des aires de Lévy pour lesquelles il n'existe pas de méthode efficace connue sauf en dimension 2. Giles et Szpruch [26] construisent alors un estimateur MLMC antithétique, permettant de négliger les aires de Lévy en utilisant le schéma de Milstein tronqué suivant :

$$\begin{aligned} \hat{X}_{t_{k+1}}^h = & \hat{X}_{t_k}^h + b(t_k, \hat{X}_{t_k}^h)h + \sigma(t_k, \hat{X}_{t_k}^h)(W_{t_{k+1}} - W_{t_k}) \\ & + \frac{1}{2} \sum_{j=1}^d \sum_{m=1}^d \partial \sigma_j \sigma_m(t_k, \hat{X}_{t_k}^h) \left( (W_{t_{k+1}}^j - W_{t_k}^j)(W_{t_{k+1}}^j - W_{t_k}^j) - \mathbf{1}_{\{j=m\}}h \right) \end{aligned} \quad (1.1.5)$$

ainsi que le couplage antithétique suivant :

$$\hat{Y}_\ell = \frac{1}{2} \left( f(\hat{X}_T^{\ell, h_\ell}) + f(\tilde{X}_T^{\ell, h_\ell}) \right) - f(\hat{X}_T^{\ell, h_{\ell-1}})$$

où  $\tilde{X}_T^{\ell, h_\ell}$  est obtenu à l'aide du même schéma numérique que  $\hat{X}_T^{\ell, h_\ell}$ , mais en transposant les paires d'accroissements browniens successifs. Ce schéma avec couplage antithétique permet alors d'obtenir un estimateur multi-pas corrigé avec une variance qui décroît en fonction du niveau  $\ell$  au même ordre que la différence  $f(\tilde{X}_T^{\ell, h_\ell}) - f(\tilde{X}_T^{\ell, h_{\ell-1}})$  calculée avec le schéma de Milstein. Ils obtiennent alors une variance MLMC en  $\mathcal{O}(h^2)$  pour des fonctions  $f$  régulières. Cela permet donc à la variance de décroître à l'ordre  $\beta = 2$  tandis que  $\alpha = 1 = \gamma$  et ainsi d'atteindre une complexité en  $\mathcal{O}(\epsilon^{-2})$ .

### 1.1.5 Contribution de la thèse

Dans le chapitre 2 de cette thèse, nous nous intéressons à la discrétisation d'Euler-Maruyama d'une EDS  $d$ -dimensionnelle avec un coefficient de diffusion constant et un coefficient de dérive  $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  mesurable borné par rapport aux variables d'espace définie par :

$$X_t = X_0 + W_t + \int_0^t b(s, X_s) ds, \quad t \in [0, T].$$

L'intérêt porté aux EDS avec des coefficients non-réguliers découle notamment du domaine des mathématiques financières, où l'on cherche à modéliser par exemple un processus de prix d'actions dont la tendance change radicalement lorsqu'un facteur baisse par rapport à une valeur seuil, ou bien en théorie du contrôle stochastique lorsque l'on choisit un processus de contrôle qui minimise le coût actualisé attendu.

Nous utilisons le schéma d'Euler-Maruyama avec une randomisation de la variable temporelle afin de nous débarrasser de toute hypothèse de régularité de la dérive par rapport à cette variable. Nous définissons alors l'interpolation continue du schéma avec pas de temps  $h \in (0, T]$  comme suit :

$$X_t^h = X_0 + W_t + \int_0^t b(\delta_{\ell_s}, X_{\tau_s^h}^h) ds, \quad t \in [0, T] \text{ où } \tau_s^h = \lfloor s/h \rfloor h \text{ et } \ell_s = \lfloor s/h \rfloor;$$

avec  $(\delta_k)_{k \in \llbracket 0, \lfloor \frac{T}{h} \rfloor \rrbracket}$  des variables aléatoires indépendantes et identiquement distribuées suivant la loi uniforme sur  $[kh, (k+1)h]$  et indépendantes de  $(X_0, (W_t)_{t \geq 0})$ . Pour  $s \in (0, t]$ , les variables  $\tau_s^h$  et  $\ell_s$  représentent respectivement le dernier instant de discrétisation avant  $s$  et l'indice correspondant à l'intervalle temporel  $s \in [kh, (k+1)h)$  tel que  $k \leq \lfloor T/h \rfloor - 1$ .

Pour  $t \geq 0$ , nous notons  $\mu_t$  et  $\mu_t^h$  les lois de probabilités respectives de la solution  $X_t$  et de sa discrétisation d'Euler-Maruyama  $X_t^h$  avec  $\mu_0 = \mu_0^h$ . Nous étudions alors l'ordre de convergence de la distance en variation totale entre  $\mu_t$  et  $\mu_t^h$ . Pour  $t > 0$ , nous montrons que  $\mu_t$  et  $\mu_t^h$  admettent des densités  $p(t, \cdot)$  et  $p^h(t, \cdot)$  par rapport à la mesure de Lebesgue et par conséquent, notre approche revient à étudier la vitesse de convergence de la norme  $L^1$  de la différence entre les densités  $p(t, \cdot)$  et  $p^h(t, \cdot)$ . Pour ce faire, nous faisons appel aux formules de Duhamel vérifiées par chacune des densités.

Nous commençons par prouver une convergence faible en variation totale d'ordre 1/2 lorsque  $b$  est uniquement mesurable bornée :

**Theorem 1.1.4.** *Supposons que la dérive  $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  est mesurable bornée, alors :*

$$\exists C < +\infty, \forall h \in (0, T], \forall k \left\| 0, \left\lfloor \frac{T}{h} \right\rfloor \right\|, \quad \|\mu_{kh} - \mu_{kh}^h\|_{TV} \leq C\sqrt{h}.$$

Ensuite, lorsque nous supposons plus de régularité sur le coefficient de dérive, à savoir que ses dérivées spatiales au sens des distributions sont  $L^\rho$  en espace uniformément en temps pour un certain  $\rho \geq d$ , nous améliorons la convergence en variation totale pour atteindre un ordre 1 (à un facteur logarithmique près) au temps terminal  $T$  :

**Theorem 1.1.5.** *Supposons que la dérive  $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  est mesurable bornée. Supposons également que  $\sup_{t \in [0, T]} \|\nabla \cdot b(t, \cdot)\|_{L^\rho} < +\infty$  pour un certain  $\rho \in [d, +\infty]$  ou bien si  $d = 1$ ,  $\sup_{t \in [0, T]} \|\partial_x b(t, \cdot)\|_{TV} < +\infty$ ; où  $\nabla \cdot b(t, \cdot)$  et  $\partial_x b(t, \cdot)$  sont respectivement la divergence spatiale et la dérivée spatiale de  $b$  au sens des distributions. Alors :*

$$\exists \tilde{C} < +\infty, \forall h \in (0, T], \forall k \left\| 0, \left\lfloor \frac{T}{h} \right\rfloor \right\|, \quad \|\mu_{kh} - \mu_{kh}^h\|_{TV} \leq \frac{\tilde{C}}{\sqrt{kh}} (1 + \ln(k)) h.$$

Supposons que  $X_0$  admet une densité régulière en plus de la régularité supposée sur  $b$ , alors nous améliorons la vitesse de convergence précédente en éliminant le préfacteur  $\frac{1}{\sqrt{kh}}$ .

Nous étudions aussi à travers la transformation de Lamperti l'application de ces théorèmes à des EDS unidimensionnelles avec un coefficient de diffusion  $\sigma$  non constant. Nous complétons notre analyse théorique par des expériences numériques.

## 1.2 Équations différentielles stochastiques de type McKean-Vlasov

Il existe un lien étroit entre certaines Équations aux Dérivées Partielles, les problèmes physiques dont elles sont issues et les EDS associées. L'EDP décrit un comportement moyen macroscopique et l'EDS un comportement microscopique et aléatoire. Nous pouvons traduire cette relation en des méthodes de résolution approchée pour certaines EDP, fondées sur des simulations de processus de diffusion. Cette même idée a permis de montrer que les mesures empiriques de certaines familles de processus de diffusion en interaction faible approchent la solution d'EDP non linéaires. Ceci a donné naissance aux méthodes dites particulières probabilistes.

On appelle processus ou diffusions de type McKean-Vlasov des processus stochastiques qui peuvent être décrits par des EDS de la forme :

$$\begin{cases} X_t = X_0 + \int_0^t b(s, \mu_s, X_s) ds + \int_0^t \sigma(s, \mu_s, X_s) dW_s, \\ \mu_t = \mathcal{L}(X_t). \end{cases} \quad (1.2.1)$$

Ces processus sont des diffusions dites non linéaires au sens de McKean car les coefficients présentent la particularité de dépendre de la loi de la variable aléatoire  $X_t$ , et non seulement de sa valeur. Ces EDS ont initialement été étudiées par Henry McKean dans le cas particulier de l'évolution temporelle de la distribution de particules chargées dans un plasma (modèle de Vlasov). Depuis, les EDS interviennent dans de nombreux champs d'application comme la physique, les neurosciences et la finance.

À l'exception de quelques cas très particuliers, ces EDS ne peuvent être résolues explicitement. On fait alors appel à la théorie de champ moyen. C'est une approche efficace qui permet de passer d'une description macroscopique d'un phénomène physique où chaque particule ressent l'influence de toutes les autres, aussi éloignées soient-elles, à une description microscopique décrivant le comportement de ces particules en interaction. Les méthodes de particules à champ moyen peuvent alors être considérées comme une linéarisation stochastique d'équations non linéaires au sens de McKean. Intervient alors la limite de champ moyen dans laquelle le nombre de particules qu'on notera dorénavant  $N$  devient très grand (formalisé par l'étude asymptotique  $N \rightarrow +\infty$ ), mais au sein duquel les interactions entre particules sont suffisamment faibles pour que les forces s'exerçant sur une particule donnée restent finies à la limite.

Une procédure de linéarisation de l'EDS (1.2.1) consiste à introduire  $N$  copies  $\left( (X_t^{i,N})_{t \geq 0} \right)_{1 \leq i \leq N}$  dirigées par  $N$  mouvements browniens indépendants  $\left( (W_t^i)_{t \geq 0} \right)_{1 \leq i \leq N}$ , et dans lesquelles la loi  $\mu_t$  est remplacée par la mesure empirique  $\mu_t^N = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i,N}}$ . On utilise cette dernière car dans un cadre où les particules sont indépendantes et identiquement distribuées, la mesure empirique converge vers la loi commune d'après la loi des grands nombres :  $\mu_t^N$  fournit alors une bonne notion de "densité de particules". On dispose donc du système de particules en interaction suivant :

$$\begin{cases} X_t^{i,N} = X_0^i + \int_0^t b(s, \mu_s^N, X_s^{i,N}) ds + \int_0^t \sigma(s, \mu_s^N, X_s^{i,N}) dW_s^i, \\ \mu_t^N = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i,N}}. \end{cases} \quad (1.2.2)$$

On se pose alors ces deux questions :

- Sommes-nous en mesure de déterminer la loi des positions des particules à un temps ultérieur  $t$  connaissant les positions des particules au temps initial  $t = 0$  dans la limite où  $N$  tend vers l'infini ?
- Si on suppose que les états  $(X_0^i)_{1 \leq i \leq N}$  des particules au temps initial  $t = 0$  sont indépendants, est-ce que cette indépendance reste vraie aux temps ultérieurs dans la limite où  $N$  tend vers l'infini ? On parlera alors de propagation du chaos.

### 1.2.1 Propagation du chaos et analyse de l'erreur forte de discrétisation en particules

L'approximation du processus non linéaire par les systèmes de particules décrite ci-dessus est généralement rendue valide par un résultat de propagation du chaos : lorsque le nombre de particules augmente, la mesure empirique  $\mu_t^N$  échantillonne de mieux en mieux la loi  $\mu_t$ . L'intérêt des méthodes particulières réside donc dans cette convergence : nous sommes en mesure d'approcher la loi de la solution d'équations de McKean à l'aide de la mesure empirique du système à  $N$  particules lorsque l'interaction entre deux particules données qui est d'ordre  $\frac{1}{N}$  disparaît, si bien que leurs trajectoires deviennent indépendantes.

On obtient alors des vitesses de convergence du système de particules vers le processus limite. Le premier résultat apparaît dans le cours fondateur de Sznitman [84] où il étudie des EDS non linéaires au sens de McKean avec des coefficients qui sont réguliers et où la mesure intervient de façon linéaire au travers d'une intégrale :

$$\begin{pmatrix} \sigma \\ b \end{pmatrix} (s, \mu, x) = \varsigma \left( s, \int f(x, y) \mu(dy), x \right).$$

La fonction  $f$  est lipschitzienne et la fonction  $\varsigma$  est lipschitzienne à croissance linéaire par rapport à ses deuxième et troisième variables uniformément en temps. Sznitman couple alors la particule  $X^{i,N}$  avec le processus  $X^i$  qui est solution de (1.2.1) où le mouvement brownien  $W$  est remplacé par  $W^i$  et  $X_0$  par  $X_0^i$ . En exploitant ce couplage, il prouve que l'ordre fort de convergence est en  $\mathcal{O}(N^{-1/2})$  :

**Theorem 1.2.1.** *Pour tout  $i \in \llbracket 1, N \rrbracket$  :*

$$\exists C < +\infty, \forall N \in \mathbb{N}^*, \quad \mathbb{E} \left[ \sup_{t \in [0, T]} |X_t^{i,N} - X_t^i|^2 \right] \leq \frac{C}{N}.$$

Il aura fallu attendre plusieurs années et l'arrivée de la théorie des dérivées par rapport aux mesures, notamment la  $L$ -dérivée et qui a été introduite pour les jeux à champ moyens par P. -L Lions dans ses cours au Collège de France [66], pour étendre ce résultat de convergence forte. Szpruch et Tse [79] généralisent alors ce résultat de propagation du chaos à des EDS dont les coefficients sont homogènes en temps et ne dépendent pas nécessairement de la mesure au travers d'une intégrale. Ils prouvent, dans un cadre très régulier, qu'en moyenne la distance de Wasserstein quadratique sur  $\mathcal{C}([0, T] \times \mathbb{R}^d)$  entre les mesures empiriques de  $X^{i,N}$  et  $X^i$  est d'ordre  $\mathcal{O}(N^{-1/2})$  :

**Theorem 1.2.2.** *On suppose que la loi initiale de  $X_0$  admet des moments d'ordre 12 finis. De plus, on suppose que  $b$  et  $\sigma$  appartiennent à  $\mathcal{M}_3(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$  où  $\mathcal{P}_2(\mathbb{R}^d)$  désigne l'espace des mesures de probabilité de carré intégrable et  $\mathcal{M}_3(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$  désigne l'espace des fonctions 3 fois différentiables en espace et en mesure (pour la notion de  $L$ -différentiabilité introduite plus haut) avec des dérivées à tous ordres bornées et lipschitziennes. On obtient alors :*

$$\exists C < +\infty, \forall N \in \mathbb{N}^*, \quad \mathbb{E} \left[ \mathcal{W}_2 \left( \mu^N, \frac{1}{N} \sum_{i=1}^N \delta_{X^i} \right)^4 \right] \leq \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^N \left( \sup_{t \in [0, T]} |X_t^i - X_t^{i, N}|^4 \right) \right] \leq \frac{C}{N^2}.$$

On peut alors déduire du Théorème 1.2.1 que pour toute fonction  $\varphi$  lipschitzienne, l'erreur quadratique moyenne se comporte en  $\mathcal{O}(N^{-1})$  :

$$\exists C < +\infty, \forall N \in \mathbb{N}^*, \quad \sup_{t \in [0, T]} \mathbb{E} \left[ \left( \frac{1}{N} \sum_{i=1}^N \varphi(X_t^{i, N}) - \mathbb{E}[\varphi(X_t^1)] \right)^2 \right] \leq \frac{C}{N}. \quad (1.2.3)$$

Lorsque les coefficients des EDS ne satisfont plus les conditions classiques de type Lipschitz, on parle de problèmes de McKean singuliers. On cherche alors à exploiter la structure particulière de chaque type d'interaction pour montrer la propagation du chaos. Des EDS de McKean singulières qui reviennent souvent en finance ou en mécanique des fluides sont les EDS unidimensionnelles où  $\mu_t$  intervient dans les coefficients au travers de sa fonction de répartition  $F(t, \cdot)$ . Elles sont de la forme suivante :

$$\begin{cases} X_t = X_0 + \int_0^t \lambda(F(s, X_s)) ds + \int_0^t \sigma(F(s, X_s)) dW_s, \\ F(t, x) = \mathbb{P}(X_t \leq x) \quad \forall (t, x) \in [0, T] \times \mathbb{R}. \end{cases} \quad (1.2.4)$$

On dit du système de particules associé qu'il interagit à travers le rang. On l'appelle plus communément en anglais modèle Rank-based. Kolli et Shkolnikov [56] prouvent alors une propagation du chaos avec vitesse optimale en  $\mathcal{O}(N^{-1/2})$  pour ce type d'EDS lorsqu'ils supposent que :

- Les positions initiales  $(X_0^i)_{1 \leq i \leq N}$  du système particulaire sont indépendantes et identiquement distribuées suivant une loi à densité bornée avec des moments d'ordre  $(2 + \eta)$  finis pour un certain  $\eta > 0$ .
- Les coefficients  $\lambda$  et  $\sigma$  sont différentiables avec des dérivées localement höldériennes.

### 1.2.2 Analyse de l'erreur faible de discrétisation en particules : Biais

On a vu plus haut que la vitesse de convergence de l'erreur quadratique moyenne (1.2.3) est du même ordre que l'erreur statistique et on se demande alors si le biais converge plus rapidement. L'idée est de dresser une analogie entre la discrétisation en nombre de particules et la discrétisation en temps des EDS à l'aide du schéma d'Euler-Maruyama où la vitesse forte de convergence se comporte généralement en  $\mathcal{O}(\sqrt{h})$  et la vitesse faible de convergence en  $\mathcal{O}(h)$ .

L'intérêt pour le biais introduit par les approximations particulières est d'abord motivé par une efficacité numérique. Nous avons observé via des simulations numériques que le biais se comporte généralement en  $\mathcal{O}(N^{-1})$  même pour des EDS de McKean singulières (cf. à la section numérique 3.3 du chapitre 3). De plus, si l'on suppose que le biais se comporte en  $\mathcal{O}(N^{-1})$  au lieu de  $\mathcal{O}(N^{-1/2})$ , alors on réduit le coût algorithmique. En effet, simuler  $\sqrt{N}$  copies indépendantes d'un système à  $\sqrt{N}$  particules entraîne le même ordre d'erreur  $\mathcal{O}(N^{-1/2})$  qu'effectuer une seule simulation d'un système avec  $N$  particules mais induit une réduction de coût de  $\mathcal{O}(N^2)$  à  $\mathcal{O}(N^{3/2})$  quand la simulation de l'interaction a un coût quadratique. On peut expliquer tout cela comme suit :

Soit  $t \in [0, T]$  et  $p$  l'indice des  $\sqrt{N}$  simulations indépendantes. On commence par une décomposition

biais-variance de l'erreur quadratique moyenne :

$$\begin{aligned} & \mathbb{E} \left[ \left( \frac{1}{\sqrt{N}} \sum_{p=1}^{\sqrt{N}} \frac{1}{\sqrt{N}} \sum_{i=1}^{\sqrt{N}} \varphi \left( X_t^{p,i,\sqrt{N}} \right) - \mathbb{E} [\varphi (X_t^1)] \right)^2 \right] \\ &= \left( \mathbb{E} [\varphi (X_t^{1,\sqrt{N}})] - \mathbb{E} [\varphi (X_t^1)] \right)^2 + \mathbb{E} \left[ \left( \frac{1}{N} \sum_{p=1}^{\sqrt{N}} \sum_{i=1}^{\sqrt{N}} \varphi \left( X_t^{p,i,\sqrt{N}} \right) - \mathbb{E} [\varphi (X_t^{1,\sqrt{N}})] \right)^2 \right]. \end{aligned}$$

Le terme du biais au carré contribue en  $\mathcal{O}(N^{-1})$  car on rappelle qu'on dispose d'un système de  $\sqrt{N}$  particules. Concernant le terme de variance, on tire avantage de l'indépendance des  $\sqrt{N}$  simulations et on le réécrit sous la forme suivante:

$$\frac{1}{\sqrt{N}} \mathbb{E} \left[ \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{\sqrt{N}} \varphi \left( X_t^{i,\sqrt{N}} \right) - \mathbb{E} [\varphi (X_t^{1,\sqrt{N}})] \right)^2 \right].$$

L'espérance globale contribue en  $\mathcal{O}(N^{-1/2})$  et comme on multiplie par  $\frac{1}{\sqrt{N}}$  alors la variance contribue au total en  $\mathcal{O}(N^{-1})$ . Par conséquent, l'erreur quadratique moyenne se comporte en  $\mathcal{O}(N^{-1})$ . Ainsi pour une précision  $\epsilon$ , une erreur quadratique moyenne en  $\mathcal{O}(\epsilon^2)$  nécessite que  $N = \mathcal{O}(\epsilon^{-2})$ . Le biais du système à  $\sqrt{N}$  particules nous coûte donc  $\epsilon^{-1}$  mais l'interaction des particules ne nous coûte dorénavant que  $N$  au lieu de  $N^2$ . On finit alors avec un coût total en  $\epsilon^{-3}$  au lieu de  $\epsilon^{-4}$ .

L'ordre de convergence faible de la discrétisation particulière des diffusions non linéaires au sens de McKean a récemment fait l'objet de plusieurs études. On cite tout d'abord les travaux de Mischler, Mouhot et Wennberg [67] qui prouvent pour une fonction  $\varphi$  lipschitzienne et ayant une certaine régularité Sobolev que l'erreur faible se comporte en  $\mathcal{O}(N^{-1})$  :

$$\exists C < +\infty, \forall N \in \mathbb{N}^*, \quad \sup_{t \in [0, T]} \left| \mathbb{E} [\varphi (X_t^{1,N})] - \int_{\mathbb{R}^n} \varphi(x) \mu_t(dx) \right| \leq \frac{C}{N}$$

lorsque le coefficient de dérive  $b$  est lipschitzien avec une dépendance linéaire par rapport à la mesure, et le coefficient de diffusion  $\sigma$  ne dépend ni de la mesure ni du temps et est uniformément elliptique. Quand  $n = d$ , Chassagneux, Szpruch et Tse [13] établissent un développement du biais en nombre de particules, semblable au développement de l'erreur faible en pas de temps  $h$  établi par Talay et Tubaro plus haut (cf. au théorème 1.1.2). Ainsi, pour  $k \in \mathbb{N}^*$ , ils obtiennent que :

$$\exists C_1, \dots, C_{k-1} < +\infty, \forall N \in \mathbb{N}^*, \quad \mathbb{E} [\Phi (\mu_T^N)] - \Phi (\mu_T) = \sum_{j=1}^{k-1} \frac{C_j}{N^j} + \mathcal{O} \left( \frac{1}{N^k} \right),$$

lorsque les coefficients  $b$  et  $\sigma$  sont homogènes en temps et très réguliers. En effet, ils supposent que  $b$  et  $\sigma$  sont différentiables  $(2k+1)$ -fois par rapport aux variables spatiales et à la mesure (pour la notion de  $L$ -différentiabilité introduite plus haut) et  $\sigma$  est bornée. De plus, la valeur initiale  $X_0$  admet un moment d'ordre  $(2k+1)$  fini. Ils supposent la même régularité pour la fonction test  $\Phi$  sur l'espace des probabilités sur  $\mathbb{R}^d$  et où la mesure n'intervient pas forcément de façon linéaire au travers d'une intégrale. Sous des conditions d'ellipticité uniforme, Chaudru de Raynal et Frikha [14] obtiennent, quant à eux, le résultat de convergence faible suivant :

$$\exists C < +\infty, \forall N \in \mathbb{N}^*, \quad \left| \mathbb{E} [\Phi (\mu_T^N)] - \Phi (\mu_T) \right| \leq \frac{C}{N}$$

lorsque la fonction test  $\Phi$  sur l'espace des probabilités sur  $\mathbb{R}^d$  admet deux dérivées fonctionnelles linéaires qui sont höldériennes et bornées. Ils supposent, de plus, que les coefficients  $\sigma \sigma^*$  et  $b$  sont bornés, globalement höldériens par rapport aux variables spatiales et admettent deux dérivées fonctionnelles linéaires höldériennes et bornées par rapport à la variable de mesure. La dérivée de Lions correspond à la dérivée spatiale de la dérivée par rapport à la variable qui apparaît lors de cette dérivation linéaire.

### 1.2.3 Échantillonnage antithétique

Par analogie avec la discrétisation temporelle, on peut se demander si la variance de la différence entre une fonctionnelle de la mesure empirique d'un système à  $2N$  particules et la même fonctionnelle de celle du système à  $N$  particules peut-être réduite par une approche antithétique. Il s'agit alors de générer deux systèmes à  $N$  particules tous deux construits à partir des couples i.i.d. de conditions initiales et mouvements browniens  $(X_0^i, W^i)_{1 \leq i \leq 2N}$  comme suit :

- le système à  $N$  particules  $(X_{T,(1)}^{i,N})_{1 \leq i \leq N}$  généré par la famille de couples i.i.d.  $(X_0^i, W^i)_{1 \leq i \leq N}$ ,
- le système à  $N$  particules  $(X_{T,(2)}^{i,N})_{1 \leq i \leq N}$  généré par la famille de couples i.i.d.  $(X_0^{N+i}, W^{N+i})_{1 \leq i \leq N}$ .

Soit une fonctionnelle  $\Phi$ . Afin de mesurer la pertinence de la technique d'échantillonnage antithétique pour la réduction de variance, on calcule la variance de la différence entre  $\Phi(\mu_T^{2N})$  et la moyenne de  $\Phi(\mu_{T,(1)}^N)$  et  $\Phi(\mu_{T,(2)}^N)$  pour les deux échantillons. Szpruch et Tse [79] montrent que cette variance se comporte en  $\mathcal{O}(\frac{1}{N^2})$  :

**Theorem 1.2.3.** *On suppose que la loi initiale de l'EDS admet des moments d'ordre 8 finis et que  $\Phi \in \mathcal{M}_4^L(\mathcal{P}_2(\mathbb{R}^d))$  où  $\mathcal{M}_4^L(\mathcal{P}_2(\mathbb{R}^d))$  désigne l'espace des fonctions différentiables jusqu'à l'ordre 4 au sens des dérivées fonctionnelles linéaires et où chaque dérivée est à croissance polynomiale. Alors :*

$$\exists C < +\infty, \forall N \in \mathbb{N}^*, \quad \text{Var} \left[ \Phi(\mu_T^{2N}) - \frac{1}{2} \left( \Phi(\mu_{T,(1)}^N) + \Phi(\mu_{T,(2)}^N) \right) \right] \leq \frac{C}{N^2}.$$

D'autre part, lorsque l'on discrétise en temps les EDS, on obtient un développement de l'erreur faible en fonction du pas de discrétisation  $h$  du schéma d'Euler-Maruyama et au travers des méthodes de Monte Carlo multi-pas introduites par Giles [24] et décrites plus haut, on réduit la complexité de  $\Phi(\mu_T)$  à précision donnée. On se demande alors si on dispose de méthodes analogues mais par rapport au nombre de particules et qui combinent aussi l'échantillonnage antithétique afin de faire décroître plus vite la variance. Une étude encourageante sur le MLMC antithétique dans le contexte d'EDS non-linéaires au sens de McKean a été introduite par Haji-Ali et Tempone [35]. Szpruch et Tse [79] prouvent alors un résultat portant sur la complexité de la méthode MLMC antithétique analogue au résultat de Giles mais pour les particules. Ils construisent l'estimateur MLMC antithétique  $\hat{Z}$  à partir de niveaux de discrétisation  $\ell \in \{0, \dots, L\}$  et la suite géométrique  $N_\ell = 2^\ell$  tels que :

$$\hat{Z} = \frac{1}{M_0} \sum_{m=1}^{M_0} \Phi(\mu_T^{m,N_0}) + \sum_{\ell=1}^L \left\{ \frac{1}{M_\ell} \sum_{m=1}^{M_\ell} \left[ \Phi(\mu_T^{\ell,m,N_\ell}) - \frac{1}{2} \left( \Phi(\mu_{T,(1)}^{\ell,m,N_\ell}) - \Phi(\mu_{T,(2)}^{\ell,m,N_{\ell-1}}) \right) \right] \right\}$$

où  $M_\ell \in \mathbb{N}^*$  correspond au nombre de tirages pour le niveau de discrétisation  $\ell \in \{0, \dots, L\}$ . Ils obtiennent alors le résultat suivant :

**Theorem 1.2.4.** *Soit  $\Phi(\mu_T)$  une fonctionnelle de la loi d'une EDS non linéaire au sens de McKean et  $\hat{Z}$  son estimateur MLMC antithétique. On suppose que la loi initiale de  $X_0$  admet des moments d'ordre 12 finis. De plus, on suppose que  $b$  et  $\sigma$  appartiennent à  $\mathcal{M}_4(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$  et  $\Phi \in \mathcal{M}_4(\mathcal{P}_2(\mathbb{R}^d))$  où  $\mathcal{M}_4(\mathcal{P}_2(\mathbb{R}^d))$  désigne l'espace des fonctions 4 fois différentiables en mesure (pour la notion de  $L$ -différentiabilité introduite plus haut) avec des dérivées à tous ordres bornées et lipschitziennes. Alors il existe des constantes positives  $c_1, c_2$  telles que pour tout  $\epsilon < e^{-1}$ , il y a des valeurs  $L$  et  $M_\ell$  pour lesquelles l'estimateur  $\hat{Z}$  admet une erreur quadratique moyenne bornée strictement par  $\epsilon^2$  :*

$$E \left[ \hat{Z} - \Phi(\mu_T^N) \right] \leq c_1 \epsilon^2$$

et sa complexité  $C$  est bornée par :

$$C \leq c_2 \epsilon^{-2} \log(\epsilon)^2.$$



### 1.2.4 Discrétisation temporelle

En pratique,  $X_t^{i,N}$  ne peut être calculée de manière exacte. On construit alors, à nouveau, un schéma d'approximation sur l'intervalle  $[0, T]$  du système de particules (1.2.2). On se donne une subdivision  $(t_k = kh)_{k \in \llbracket 0, \lfloor \frac{T}{h} \rfloor \rrbracket}$  avec  $h$  comme pas de temps et on approxime notre solution aux points de la subdivision par le schéma d'Euler-Maruyama défini, pour chaque  $i \in \llbracket 1, N \rrbracket$ , par :

$$X_{t_{k+1}}^{i,N,h} = X_{t_k}^{i,N,h} + b\left(t_k, \mu_{t_k}^{N,h}, X_{t_k}^{i,N,h}\right)h + \sigma\left(t_k, \mu_{t_k}^{N,h}, X_{t_k}^{i,N,h}\right)\left(W_{t_{k+1}}^i - W_{t_k}^i\right), \quad (1.2.5)$$

où  $\mu_t^{N,h} = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i,N,h}}$  représente la mesure empirique de  $X_t^{i,N,h}$ . Pour effectuer les preuves de convergence, nous introduisons l'interpolation temporelle continue du schéma d'Euler-Maruyama définie par :

$$X_t^{i,N,h} = X_0^i + \int_0^t b\left(\tau_s^h, \mu_{\tau_s^h}^{N,h}, X_{\tau_s^h}^{i,N,h}\right) ds + \int_0^t \sigma\left(\tau_s^h, \mu_{\tau_s^h}^{N,h}, X_{\tau_s^h}^{i,N,h}\right) dW_s, \quad (1.2.6)$$

avec  $\tau_s^h = \lfloor s/h \rfloor h$  et  $t \in [0, T]$ .

Bossy et Talay [9] étudient l'approximation numérique de solutions d'EDS non linéaires au sens de McKean unidimensionnelles avec des coefficients homogènes en temps et où la mesure intervient de façon linéaire à travers une intégrale, de la forme suivante :

$$X_t = X_0 + \int_0^t \left( \int_{\mathbb{R}} f(X_s, y) d\mu_s(y) \right) ds + \int_0^t \left( \int_{\mathbb{R}} g(X_s, y) d\mu_s(y) \right) dW_s.$$

Les fonctions  $f$  et  $g$  sont supposées uniformément bornées sur  $\mathbb{R}^2$ ,  $f$  est globalement lipschitzienne et  $g$  admet une borne inférieure strictement positive ainsi que des dérivées partielles à l'ordre 1 bornées uniformément. De plus, soit la loi initiale est une mesure de Dirac, soit elle admet une densité continue qui a des queues de distribution gaussiennes. Sous ces conditions de forte régularité, ils prouvent que l'espérance de la norme  $L^1$  de la différence des fonctions de répartition de  $X_t$  et  $X_t^{i,N,h}$  se comporte en  $\mathcal{O}\left(\frac{1}{\sqrt{N}} + \sqrt{h}\right)$ . Les simulations numériques montrent que ce résultat est optimal en nombre de particules mais que la dépendance en le pas de temps est plutôt en  $h$ . Antonelli et Kohatsu-Higa [2] démontrent ce résultat par la suite dans un cadre d'hypothèses comparable.

Bossy [8] prouve cette vitesse de convergence optimale en  $\mathcal{O}\left(\frac{1}{\sqrt{N}} + h\right)$  pour des EDS non linéaires où la mesure intervient à travers la fonction de répartition du processus. Plus spécifiquement, il s'agit des lois de conservation scalaire visqueuses de la forme :

$$X_t = X_0 + \sigma W_t + \int_0^t \lambda(F(s, X_s)) ds$$

où  $F(t, \cdot)$  représente la fonction de répartition de  $X_t$  et  $\lambda$  est la dérivée d'une fonction  $\Lambda : [0, 1] \rightarrow \mathbb{R}$  de classe  $\mathcal{C}^1$ . Sous des conditions de régularité sur la loi initiale ainsi que  $\Lambda$ , elle obtient le résultat suivant :

**Theorem 1.2.5.** *On suppose que  $\Lambda$  est  $\mathcal{C}^3$ , que la fonction de répartition de  $X_0$  qu'on note  $F_0$  est  $\mathcal{C}^2$  bornée avec des dérivées première et seconde bornées par rapport à la variable d'espace et il existe  $M, C, \beta > 0$ ,  $\alpha \geq 0$  tels que  $|\partial_x F_0(x)| \leq \alpha \exp(-\beta x^2/2)$  lorsque  $|x| > M$ . Alors pour tous  $t, h \in [0, T], n \in \mathbb{N}^*$ :*

$$\mathbb{E} \left[ \|F^{N,h}(t, \cdot) - F(t, \cdot)\|_{L^1} \right] + \sup_{x \in \mathbb{R}} (\mathbb{E} [|F^{N,h}(t, x) - F(t, x)|]) = \mathcal{O}\left(\frac{1}{\sqrt{N}} + h\right)$$

où  $F^{N,h}$  représente la fonction de répartition empirique de  $X_t^{i,N,h}$ .

Ce résultat améliore la vitesse  $\mathcal{O}\left(\frac{1}{\sqrt{N}} + \sqrt{h}\right)$  obtenue précédemment par Bossy et Talay [9] dans le cas particulier de l'équation de Burgers où  $\Lambda(u) = u^2/2$ .



La méthode de Monte Carlo multi-indices introduite par Haji-Ali et Tempone [35] pour réduire la variance repose sur la méthode de Monte Carlo multi-pas de Giles [24] mais combine à la fois la discrétisation en temps et en particules. Szpruch et Tse [79] proposent alors un deuxième résultat concernant la réduction de coût pour la méthode multi-indices antithétique en utilisant l'estimateur  $\tilde{Z}$  suivant :

$$\tilde{Z} = \frac{1}{M_0} \sum_{m=1}^{M_0} f\left(\mu_T^{m, N_0, h_0}\right) + \sum_{\ell=1}^L \left\{ \frac{1}{M_\ell} \sum_{m=1}^{M_\ell} \left[ f\left(\mu_T^{\ell, m, N_\ell, h_\ell}\right) - \frac{1}{2} \left( f\left(\mu_{T, (1)}^{\ell, m, N_\ell, h_\ell}\right) - f\left(\mu_{T, (2)}^{\ell, m, N_{\ell-1}, h_{\ell-1}}\right) \right) \right] \right\}.$$

Ils obtiennent alors le résultat suivant :

**Theorem 1.2.6.** *Soit  $\Phi(\mu_T)$  une fonctionnelle de la loi d'une EDS non linéaire au sens de McKean et  $\tilde{Z}$  son estimateur Monte Carlo multi-indices antithétique. On suppose que la loi initiale de  $X_0$  admet des moments d'ordre 12 finis. De plus, on suppose que  $b$  appartient à  $\mathcal{M}_4(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$ , que le coefficient de diffusion  $\sigma$  est constant et que  $\Phi \in \mathcal{M}_4(\mathcal{P}_2(\mathbb{R}^d))$ . Alors il existe des constantes positives  $c_1, c_2$  telles que pour tout  $\epsilon < e^{-1}$ , il y a des valeurs  $L$  et  $M_\ell$  pour lesquelles l'estimateur  $\tilde{Z}$  admet une erreur quadratique moyenne bornée strictement par  $\epsilon^2$  :*

$$E \left[ \tilde{Z} - \Phi(\mu_T^N) \right] \leq c_1 \epsilon^2$$

et sa complexité  $C$  est bornée par :

$$C \leq c_2 \epsilon^{-3}.$$

### 1.2.5 Contributions de la thèse

Dans les chapitres 3 et 4 de cette thèse, nous nous intéressons à la vitesse de convergence faible de systèmes de particules vers leurs processus limites pour deux classes d'EDS non linéaires au sens de McKean. L'idée est de se placer dans des cas de diffusions spéciales où nous exploitons la structure particulière de chaque type d'interaction pour montrer un ordre de convergence faible optimal en  $\mathcal{O}(\frac{1}{N} + h)$ .

Nous obtenons, dans le chapitre 3, un premier résultat de convergence faible lorsque la dépendance en loi des EDS non linéaires au sens de McKean intervient dans les coefficients au travers des moments. Soient  $\alpha : \mathbb{R}^n \rightarrow \mathbb{R}^p$ ,  $\sigma = (\sigma_j^l)_{1 \leq j \leq n, 1 \leq l \leq d} : [0, T] \times \mathbb{R}^p \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times d}$  et  $b = (b_j)_{1 \leq j \leq n} : [0, T] \times \mathbb{R}^p \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  des fonctions lipschitziennes telles que  $\sup_{t \in [0, T]} (|\sigma(t, 0, 0)| + |b(t, 0, 0)|) < +\infty$ . Nous considérons alors l'EDS  $n$ -dimensionnelle suivante :

$$X_t = X_0 + \int_0^t \sigma(s, \mathbb{E}[\alpha(X_s)], X_s) dW_s + \int_0^t b(s, \mathbb{E}[\alpha(X_s)], X_s) ds, \quad t \in [0, T] \quad (1.2.7)$$

avec  $X_0$  une variable aléatoire de carré intégrable indépendante du mouvement brownien  $d$ -dimensionnel  $(W_t)_{t \geq 0}$ . L'approximation particulière de l'EDS non linéaire au sens de McKean (1.2.7) est alors donnée par le système suivant :

Pour  $t \in [0, T]$  et  $1 \leq i \leq N$ ,

$$X_t^{i, N} = X_0^i + \int_0^t \sigma\left(s, \frac{1}{N} \sum_{j=1}^N \alpha(X_s^{j, N}), X_s^{i, N}\right) dW_s^i + \int_0^t b\left(s, \frac{1}{N} \sum_{j=1}^N \alpha(X_s^{j, N}), X_s^{i, N}\right) ds.$$

Nous définissons alors l'interpolation continue du schéma d'Euler-Maruyama avec pas de temps  $h \in (0, T]$  du système particulière associé comme suit :

Pour  $t \in [0, T]$  et  $1 \leq i \leq N$ ,

$$X_t^{i, N, h} = X_0^i + \int_0^t \sigma\left(\tau_s^h, \frac{1}{N} \sum_{j=1}^N \alpha(X_{\tau_s^h}^{j, N, h}), X_{\tau_s^h}^{i, N, h}\right) dW_s^i + \int_0^t b\left(\tau_s^h, \frac{1}{N} \sum_{j=1}^N \alpha(X_{\tau_s^h}^{j, N, h}), X_{\tau_s^h}^{i, N, h}\right) ds. \quad (1.2.8)$$

Nous prouvons, sous des contraintes de régularité supplémentaires sur les coefficients, que l'erreur faible se comporte en  $\mathcal{O}(\frac{1}{N} + h)$  :

**Theorem 1.2.7.** *Supposons que :*

- la diffusion  $\sigma$  est lipschitzienne en ses  $p + n$  dernières composantes uniformément en sa première composante et telle que  $\sup_{t \in [0, T]} (|\sigma(t, 0, 0)| + |b(t, 0, 0)|) < +\infty$ ,
- les fonctions  $\alpha$  et  $\psi$  sont deux fois continûment différentiables avec des dérivées bornées jusqu'à l'ordre 2 et des dérivées secondes lipschitziennes,
- les coefficients  $a = \sigma\sigma^*$  et  $b$  sont globalement lipschitziens, continûment différentiables par rapport à leurs composantes d'indices  $\{2, \dots, 1 + p\}$  avec des dérivées lipschitziennes par rapport à leurs  $p + n$  dernières composantes uniformément en leur première composante,
- il existe  $\tilde{d} \in \mathbb{N}^*$ ,  $\tilde{\sigma} : [0, T] \times \mathbb{R}^p \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times \tilde{d}}$  tels que pour tout  $(t, y, x) \in [0, T] \times \mathbb{R}^p \times \mathbb{R}^n$ ,  $a(t, y, x) = \tilde{\sigma}\tilde{\sigma}^*(t, y, x)$  et  $\tilde{\sigma}, b$  sont continues et deux fois continûment différentiables par rapport à leurs  $n$  dernières composantes avec des dérivées bornées jusqu'à l'ordre 2 et des dérivées secondes lipschitziennes par rapport à leurs  $n$  dernières composantes uniformément en leurs  $1 + p$  premières composantes.

Alors :

$$\exists C < +\infty, \forall h \in (0, T], \forall N \in \mathbb{N}^*, \sup_{t \in [0, T]} \left| \mathbb{E} \left[ \psi \left( X_t^{1, N, h} \right) \right] - \mathbb{E} [\psi(X_t)] \right| \leq C \left( \frac{1}{N} + h \right).$$

Nous prouvons ce résultat en utilisant l'EDP de Feynman-Kac associée. Bien entendu, nous ne tirons pas avantage de la réduction de coût expliquée plus haut car la complexité du calcul de l'interaction est linéaire en le nombre de particules. Néanmoins, ce résultat peut être considéré comme une première étape avant d'aborder des interactions plus générales qui pourraient nécessiter des outils plus avancés comme la master équation pour les jeux à champ moyens introduite par Lions [66].

Nous complétons notre analyse théorique par des expériences numériques. Ces dernières confirment le comportement en  $\mathcal{O}(N^{-1})$  de l'erreur faible même pour des EDS de McKean singulières. Nous vérifions aussi sur ces mêmes EDS que la variance antithétique décroît, dans les bons cas, en  $\mathcal{O}(N^{-1})$ .

Dans un cadre plus singulier, nous obtenons, dans le chapitre 4, un second résultat de convergence faible lorsque la dépendance en loi intervient dans les coefficients au travers de la fonction de répartition. Soit  $\Lambda : [0, 1] \rightarrow \mathbb{R}$  une fonction  $\mathcal{C}^1$  de dérivée  $\lambda : \mathbb{R} \rightarrow \mathbb{R}$  et  $\sigma > 0$ . Nous considérons alors l'EDS unidimensionnelle ( $n = d = 1$ ) suivante :

$$\begin{cases} X_t = X_0 + \sigma W_t + \int_0^t \lambda(F(s, X_s)) ds, & t \in [0, T] \\ F(s, x) = \mathbb{P}(X_s \leq x), & \forall (s, x) \in [0, T] \times \mathbb{R}. \end{cases} \quad (1.2.9)$$

Nous supposons que  $X_0$  admet la fonction de répartition  $F_0$ . Pour  $t > 0$ , nous montrons que la loi  $\mu_t$  de  $X_t$  admet une densité  $p(t, \cdot)$  par rapport à la mesure de Lebesgue. Nous déduisons par la suite que la fonction de répartition de  $X_t$  est solution faible de la loi de conservation scalaire visqueuse suivante :

$$\begin{cases} \partial_t F(t, x) + \partial_x (\Lambda(F(t, x))) = \frac{\sigma^2}{2} \partial_{xx} F(t, x), \\ F_0(x) = F(0, x). \end{cases} \quad (1.2.10)$$

En approchant  $F(t, x)$  par la fonction de répartition empirique  $F^N(t, x) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{X_t^{i, N} \leq x\}}$  et en effectuant une légère correction sur le coefficient de dérive  $\lambda$ , nous introduisons le système particulaire associé :

$$X_t^{i, N} = X_0^i + \sigma W_t^i + \int_0^t \lambda^N \left( \sum_{j=1}^N \mathbf{1}_{\{X_s^{j, N} \leq X_s^{i, N}\}} \right) ds, \quad 1 \leq i \leq N, \quad t \in [0, T]; \quad (1.2.11)$$

où pour  $1 \leq i \leq N$ , le coefficient  $\lambda(i/N)$  est approché par :

$$\lambda^N(i) = N \left( \Lambda \left( \frac{i}{N} \right) - \Lambda \left( \frac{i-1}{N} \right) \right). \quad (1.2.12)$$

Cette modification de la dérive est motivée par le fait que lorsque nous disposons de particules ordonnées  $y^1 < y^2 < \dots < y^N$  alors la dérivée au sens des distributions de  $x \mapsto \Lambda \left( \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{y^i \leq x\}} \right)$  correspond à  $x \mapsto \frac{1}{N} \sum_{i=1}^N \lambda^N(i) \mathbf{1}_{\{y^i \leq x\}}$  et non pas (lorsque  $\Lambda$  n'est pas affine) à  $x \mapsto \frac{1}{N} \sum_{i=1}^N \lambda \left( \frac{i}{N} \right) \mathbf{1}_{\{y^i \leq x\}}$ . Comme notre analyse de l'erreur est fondée sur une comparaison entre la formulation de Duhamel de l'EDP (1.2.10) et la formule de Duhamel perturbée satisfaite par la fonction de répartition empirique de la discrétisation d'Euler du système de particules, nous nous concentrons sur le système de particules avec une dérive corrigée (1.2.11) pour lequel aucun terme d'erreur supplémentaire n'apparaît. Concernant les valeurs initiales  $(X_0^i)_{1 \leq i \leq N}$ , nous considérons les deux cas d'une initialisation aléatoire où les variables sont i.i.d. suivant la loi de  $X_0$  ainsi qu'une initialisation déterministe optimale pour la distance de Wasserstein d'indice 1 où les variables sont égales à  $(F_0^{-1}(\frac{2i-1}{2N}))_{1 \leq i \leq N}$ .

Nous introduisons alors la discrétisation d'Euler-Maruyama avec pas de temps  $h \in (0, T]$  du système de particules (1.2.11) :

$$X_t^{i,N,h} = X_0^i + \sigma W_t^i + \int_0^t \lambda^N \left( \sum_{j=1}^N \mathbf{1}_{\{X_{\tau_s^h}^{j,N,h} \leq X_{\tau_s^h}^{i,N,h}\}} \right) ds, \quad 1 \leq i \leq N, t \in [0, T]. \quad (1.2.13)$$

Nous prouvons alors que l'erreur faible  $L^1$  entre la fonction de répartition empirique  $F^{N,h}(t, \cdot)$  de  $X_t^{i,N,h}$  et sa limite  $F(t, \cdot)$  se comporte en  $\mathcal{O}(\frac{1}{N} + h)$ . La norme  $L^1$  de la différence entre  $F^{N,h}(t, \cdot)$  et  $F(t, \cdot)$  s'interprète comme la distance de Wasserstein d'indice 1 entre la mesure empirique  $\mu_t^{N,h} = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i,N,h}}$  de  $X_t^{i,N,h}$  et la loi  $\mu_t$  du processus  $X_t$ .

**Theorem 1.2.8.** *Supposons que  $\lambda$  est lipschitzienne et que les positions initiales  $(X_0^i)_{1 \leq i \leq N}$  sont :*

- soit i.i.d. avec  $\int_{\mathbb{R}} \sqrt{F_0(x)(1-F_0(x))} dx < +\infty$ ,
- ou bien déterministes optimales avec la loi de  $X_0$  qui est à support compact.

Alors :

$$\exists C_b < +\infty, \forall N \in \mathbb{N}^*, \forall h \in [0, T], \sup_{t \leq T} \mathcal{W}_1 \left( \mathbb{E} [\mu_t^{N,h}], \mu_t \right) \leq C_b \left( \frac{1}{N} + h \right).$$

Afin de prouver ce résultat de convergence faible, nous avons besoin du résultat de convergence forte suivant :

**Theorem 1.2.9.** *Supposons que les positions initiales  $(X_0^i)_{1 \leq i \leq N}$  sont :*

- soit i.i.d. avec  $\int_{\mathbb{R}} \sqrt{F_0(x)(1-F_0(x))} dx < +\infty$ ,
- ou bien déterministes optimales avec  $\sup_{x \geq 1} \int_x^{+\infty} (F_0(-y) + 1 - F_0(y)) dy < +\infty$ .

Alors :

$$\exists C < +\infty, \forall N \in \mathbb{N}^*, \sup_{t \leq T} \mathbb{E} [\mathcal{W}_1 (\mu_t^N, \mu_t)] \leq \frac{C}{\sqrt{N}}.$$

De plus, si la fonction  $\lambda$  est lipschitzienne, nous avons :

$$\exists C < +\infty, \forall N \in \mathbb{N}^*, \forall h \in (0, T], \sup_{t \leq T} \mathbb{E} [\mathcal{W}_1 (\mu_t^{N,h}, \mu_t)] \leq C \left( \frac{1}{\sqrt{N}} + h \right).$$

Ce théorème de convergence forte est une généralisation du théorème de Bossy [8] cité plus haut dans la mesure où nous supposons beaucoup moins de régularité sur les positions initiales quand celles-ci sont aléatoires. Nous traitons aussi le cas où les conditions initiales sont déterministes optimales et nous relaxons, en général, les hypothèses supposées sur  $\lambda$ . Toutefois, Bossy prouve un résultat plus large car l'estimation porte non seulement sur la norme  $L^1$  de la différence des fonctions de répartition mais aussi sur la norme  $L^\infty$ .

Profitant de la constance du coefficient de diffusion  $\sigma$ , nous prouvons aussi un résultat de propagation du chaos avec vitesse optimale en  $\mathcal{O}(N^{-1/2})$  similaire à celui obtenu par Kolli et Shkolnikov [56] plus haut mais sans hypothèse de densité bornée pour la loi initiale et en assouplissant les hypothèses sur  $\lambda$ . Leur résultat traite d'un cas plus général et difficile car le coefficient de diffusion dépend également du rang et nécessite donc plus de régularité sur la loi initiale et  $\lambda$ .

**Theorem 1.2.10.** *Supposons que les conditions initiales  $X_0^i$  sont i.i.d. et le processus  $(X_t^i)_{t \geq 0}$  est solution de l'EDS non linéaire au sens de McKean (1.2.9) avec le mouvement brownien  $(W_t)_{t \geq 0}$  remplacé par  $(W_t^i)_{t \geq 0}$  et  $X_0$  remplacé par  $X_0^i$ . Si  $\lambda$  est lipschitzienne, alors :*

$$\forall \rho, T > 0, \exists C < \infty, \forall N \in \mathbb{N}^*, \forall i \in \{1, \dots, N\}, \quad \mathbb{E} \left[ \sup_{t \in [0, T]} |X_t^i - X_t^{i, N}|^\rho \right] \leq CN^{-\rho/2}.$$

Il est à noter que nous montrons que ces trois résultats de convergence restent vrais lorsque nous choisissons la discrétisation en temps et en particules de l'EDS non linéaire au sens de McKean (1.2.10) sans correction du coefficient de dérive, à savoir  $\lambda^N(i)$  remplacée par  $\lambda(i/N)$ .

Nous complétons notre étude théorique par des expériences numériques pour le cas particulier de l'équation de Burgers où, à travers la transformation de Cole-Hopf, nous disposons d'une formule explicite de la fonction de répartition  $F(t, \cdot)$ . Nous nous plaçons dans le cas où  $\Lambda(u) = -(1-u)^2/2$  et  $F_0$  est la fonction d'Heaviside.

## Part I

# Standard Stochastic differential Equations



## CHAPTER 2

# CONVERGENCE IN TOTAL VARIATION OF THE EULER-MARUYAMA SCHEME APPLIED TO DIFFUSION PROCESSES WITH MEASURABLE DRIFT COEFFICIENT AND ADDITIVE NOISE

*Ce chapitre est un article écrit avec Benjamin Jourdain. Une version raccourcie a été soumise à SIAM Journal on Numerical Analysis.*

### Abstract

We are interested in the Euler-Maruyama discretization of a stochastic differential equation in dimension  $d$  with constant diffusion coefficient and bounded measurable drift coefficient. In the scheme, a randomization of the time variable is used to get rid of any regularity assumption of the drift in this variable. We prove weak convergence with order  $1/2$  in total variation distance. When the drift has a spatial divergence in the sense of distributions with  $\rho$ -th power integrable with respect to the Lebesgue measure in space uniformly in time for some  $\rho \geq d$ , the order of convergence at the terminal time improves to 1 up to some logarithmic factor. In dimension  $d = 1$ , this result is preserved when the spatial derivative of the drift is a measure in space with total mass bounded uniformly in time. We confirm our theoretical analysis by numerical experiments.

**Keywords:** diffusion processes, Euler scheme, weak error analysis.

## 2.1 Introduction

In numerous areas such as mathematical finance when, for example, modelling a stock price process whose trend dramatically changes when a factor goes down a threshold value, or in stochastic control theory when choosing a control process that minimizes the expected discounted cost, we end-up dealing with diffusions that do not show a smooth behavior which results in Stochastic Differential Equations with discontinuous coefficients. In the present paper, we are interested in the Euler-Maruyama discretization of the stochastic differential equation

$$X_t = X_0 + W_t + \int_0^t b(s, X_s) ds, \quad t \in [0, T] \quad (2.1.1)$$

where  $(W_t)_{t \geq 0}$  is a  $d$ -dimensional Brownian motion independent from the initial  $\mathbb{R}^d$ -valued random vector  $X_0$ ,  $T \in (0, +\infty)$  is a finite time horizon and the drift coefficient  $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  is merely measurable and bounded.

While the convergence properties of the Euler-Maruyama scheme are well understood for SDEs with smooth coefficients, the case of irregular coefficients is still an active field of research. Concerning the strong error, the additive noise case is investigated in [36] where Halidias and Kloeden only prove convergence and in [16, 74] where rates are derived. Dareiotis and Gerencsér [16] obtain convergence with  $L^2$ -order  $1/2-$  (meaning  $1/2-\epsilon$  for arbitrarily small  $\epsilon > 0$ ) in the time-step for bounded and Dini-continuous time-homogeneous drift coefficients and check that this order is preserved in dimension  $d = 1$  when the Dini-continuity assumption is relaxed to mere measurability. In the scalar  $d = 1$  case, Neuenkirch and Szölgényi [74] assume that the drift coefficient is the sum of a  $\mathcal{C}_b^2$  part and a bounded integrable irregular part with a finite Sobolev-Slobodeckij semi-norm of index  $\kappa \in (0, 1)$ . They prove  $L^2$ -convergence with order  $\frac{3}{4} \wedge \frac{1+\kappa}{2}-$  for the equidistant Euler-Maruyama scheme, the cutoff of this order at  $\frac{3}{4}$  disappearing for a suitable non-equidistant time-grid. Note that an exact simulation algorithm has been proposed by Étoré and Martinez [19] for one-dimensional SDEs with additive noise and time-homogeneous and smooth except at one discontinuity point drift coefficient.

More papers have been devoted to the strong error of the Euler scheme for SDEs with a non constant diffusion coefficient. The first result goes back to Gyöngy and Krylov [31] who established convergence in probability (without any rate) when the coefficients are continuous in space and pathwise uniqueness holds for the stochastic differential equation. Gyöngy [32] proves almost sure convergence with order  $1/4-$  when the diffusion coefficient is locally Lipschitz in space and the drift coefficient locally one-sided Lipschitz in space uniformly in time and some Lyapunov condition holds. Yan [90] investigates conditions under which the Euler scheme converges to the unique weak solution of the SDE. In dimension  $d = 1$ , this author also shows the  $L^1$ -order  $\beta_1 \wedge \frac{\alpha}{2} \wedge \frac{\alpha}{1+\alpha} \beta_2$  when the drift coefficient is Lipschitz in the spatial variables and  $\beta_1$ -Hölder in time while the diffusion coefficient is  $(\frac{1}{2} + \alpha)$ -Hölder in space and  $\frac{\beta_2}{2}$ -Hölder in time. Still in dimension one, Gyongy and Rasonyi [33] obtain the  $L^1$ -order  $\alpha \wedge \frac{\gamma}{2}$  when the diffusion coefficient is  $(\frac{1}{2} + \alpha)$ -Hölder continuous in space and the drift coefficient is the sum of a function Lipschitz continuous in space and a function non-increasing and  $\gamma$ -Hölder continuous in space. Like in [31, 32], the discretization only concerns the spatial variable of the coefficients while the time variable still moves continuously in the scheme analysed. In [75], Ngo and Taguchi prove  $L^1$ -order  $1/2$  when (resp.  $\alpha \in (0, 1/2]$ , when  $d = 1$ ,) the diffusion coefficient is uniformly elliptic, bounded, Lipschitz (resp.  $(\frac{1}{2} + \alpha)$ -Hölder) in space and the drift coefficient one-sided Lipschitz, bounded and with bounded variation in space with respect to some Gaussian measure. The coefficients are assumed to be  $1/2$ -Hölder with respect to the time variable. In a second paper [76] specialized to dimension  $d = 1$  with time-homogeneous coefficients, they show  $L^1$ -order  $\frac{\beta}{2} \wedge \alpha$  under the same assumption on the diffusion coefficient and when the drift is the sum of a bounded  $\beta$ -Hölder function and a bounded function with bounded variation with respect to some Gaussian measure. When the diffusion coefficient is uniformly elliptic, Lipschitz continuous in space, Dini-continuous in time and the drift coefficient is Dini-continuous in both variables, Bao, Huang and Yuan [5] prove  $L^2$ -convergence with an order expressed in terms of the Dini modulus of continuity. Recent attention has been paid to the Euler-Maruyama discretization of SDEs with a piecewise Lipschitz drift coefficient and a globally Lipschitz diffusion coefficient which satisfies some non-degeneracy condition on the discontinuity hypersurface of the drift coefficient. Leobacher and Szölgényi [65] prove convergence with  $L^2$ -order  $1/4-$  of the Euler-Maruyama method. This result is proved by comparison with a scheme with  $L^2$ -order  $1/2$  [63, 64] obtained by the Euler discretization of a transformation of the



original SDE which permits to remove the discontinuity of the drift. In dimension  $d = 1$ , Müller-Gronbach and Yaroslavl'tseva [71] recover for each  $p \in [1, \infty)$  the  $L^p$ -order of convergence  $1/2$  valid when the drift coefficient is globally Lipschitz. In higher dimension, this order  $1/2$  (up to some logarithmic factor) is proved by Neuenkirch, Szölgyenyi and Szpruch [73] for the  $L^2$ -error of an Euler-Maruyama scheme with adaptive time-stepping.

Concerning the weak error, Mikulevicius and Platen [68] prove that, under uniform ellipticity, when the coefficients are  $\alpha$ -Hölder with respect to the spatial variables and  $\frac{\alpha}{2}$ -Hölder with respect to the time variable for  $\alpha \in (0, 1)$ , then  $\mathbb{E}[f(X_T)]$  is approximated with order  $\frac{\alpha}{2}$  by replacing  $X_T$  by the Euler-Maruyama scheme at time  $T$  when the test function  $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is twice continuously differentiable with  $\alpha$ -Hölder second order derivatives. In the additive noise case, Kohatsu-Higa, Lejay and Yasuda [58], prove that for  $f$  thrice continuously differentiable with polynomially growing derivatives, the convergence holds with order  $1/2$ — when  $d \geq 2$  (resp.  $1/3$ — when  $d = 1$ ) and the drift coefficient is time homogeneous, bounded and Lipschitz except on a set  $G$  such that  $\varepsilon^{-d}$  times the Lebesgue measure of  $\{x \in \mathbb{R}^d : \inf_{y \in G} |x - y| \leq \varepsilon\}$  is bounded. Their approach, which consists in regularizing the drift coefficient and considering both the stochastic differential equation and the Euler scheme for the regularized coefficient, is also applied in [59] to the case of time-dependent, bounded, uniformly elliptic and continuous diffusion coefficients. In [60], Konakov and Menozzi regularize both coefficients to obtain that the absolute difference between the densities with respect to the Lebesgue measure of the solution and its Euler-Maruyama discretization is bounded from above by a Gaussian density multiplied by a factor with order  $\frac{\alpha}{2}$ —, when these coefficients are uniformly elliptic, bounded and  $\frac{\alpha}{2}$ -Hölder continuous in time and  $\alpha$ -Hölder continuous in space. The order of the factor is  $\frac{1}{2d}$ — when the coefficients are bounded, uniformly elliptic, continuously differentiable with bounded derivatives up to the order 2 in time and the order 4 in space at the possible exception, for the drift coefficient, of a finite union of time-independent smooth submanifolds where it can be discontinuous. In [22], Frikha deals with time-homogeneous one-dimensional stochastic differential equations possibly involving a local time term in addition to a bounded measurable drift coefficient and a bounded uniformly elliptic and  $\alpha$ -Hölder diffusion coefficient. He proves that the absolute difference between the densities is smaller than a Gaussian density multiplied by a factor with order  $\frac{\alpha}{2}$  in the time-step. The latest work in this field is by Suo, Yuan and Zhang [83] who study in a multidimensional setting stochastic differential equations with additive noise and time-homogeneous drift coefficients with at most linear growth and satisfying an integrated against some Gaussian measure  $\alpha$ -Hölder type regularity condition. When this coefficient has sublinear growth (and under some restriction on the time-horizon when it has linear growth), they prove convergence in total variation with order  $\frac{\alpha}{2}$ .

In the current paper, we consider the stochastic differential equation (2.1.1) with additive noise and bounded and measurable drift function  $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ . We are interested in estimating the spatial integral of the absolute difference between the densities with respect to the Lebesgue measure of the solution and its Euler-Maruyama discretization with time-step  $h \in (0, T]$ . This integral is equal to the total variation distance between the probability measures that admit these densities with respect to the Lebesgue measure. Note that the approximation of a Markovian semi-group in total variation distance has been investigated by Bally and Rey [3] who apply their results to the Ninomiya discretization scheme.

To get rid with any assumption stronger than mere measurability concerning the regularity of the drift coefficient with respect to the time variable, we consider the Euler-Maruyama discretization with randomized time variable of (2.1.1). It evolves inductively on the regular time-grid  $(kh)_{k \in \llbracket 0, \lfloor \frac{T}{h} \rfloor \rrbracket}$  by:

$$X_{(k+1)h}^h = X_{kh}^h + (W_{(k+1)h} - W_{kh}) + b(\delta_k, X_{kh}^h)h, \quad (2.1.2)$$

where the random variables  $(\delta_k)_{k \in \llbracket 0, \lfloor \frac{T}{h} \rfloor \rrbracket}$  are independent, respectively distributed according to the uniform law on  $[kh, (k+1)h]$  and are independent from  $(X_0, (W_t)_{t \geq 0})$ . Notice that this sequence is of course not needed to randomize the time variable when  $b$  is time-homogeneous. To our knowledge, such randomization techniques have been proposed so far to improve the strong convergence properties of discretizations of ordinary differential equations [81, 82, 37, 41, 17] or stochastic differential equations [61]. They also happen to be quite efficient in terms of weak error. Indeed, the above randomization turns out to enable convergence in total variation distance with order 1 up to some logarithmic factor. For  $s \in [0, T)$ , we denote by  $\ell_s = \lfloor s/h \rfloor$  the index of the corresponding time interval  $s \in [kh, (k+1)h)$ .

s.t.  $k \leq \lfloor T/h \rfloor - 1$ . We consider then the following continuous time interpolation of the scheme:

$$X_t^h = X_0 + W_t + \int_0^t b(\delta_{\ell_s}, X_{\tau_s^h}^h) ds, \quad t \in [0, T] \text{ where } \tau_s^h = \lfloor s/h \rfloor h. \quad (2.1.3)$$

For  $t \geq 0$ , we denote by  $\mu_t$  the law of  $X_t$  and by  $\mu_t^h$  the law of  $X_t^h$ . We have  $\mu_0 = \mu_0^h = m$ . For  $t > 0$ , according to Proposition 2.2.10 below,  $\mu_t$  and  $\mu_t^h$  admit densities  $p(t, \cdot)$  and  $p^h(t, \cdot)$  with respect to the Lebesgue measure. Therefore, our approach amounts to study the rate of convergence of the  $L^1$ -norm of the difference between  $p(t, \cdot)$  and  $p^h(t, \cdot)$ . We assume, in what is next, that  $X_0$  is distributed according to a probability measure  $m$  on  $\mathbb{R}^d$  and the drift  $b = (b_i)_{1 \leq i \leq d} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  is a measurable function bounded by  $B < +\infty$  when  $\mathbb{R}^d$  is endowed with the  $L^\infty$ -norm.

The paper is organized as follows. In Section 2.2, we state our main results. We first obtain the convergence of the weak error in total variation in  $\mathcal{O}(\sqrt{h})$  when  $b$  is measurable and bounded. When assuming more regularity on  $b$  with respect to the space variables, namely that the divergence in the sense of distributions of  $b$  with respect to these variables is in  $L^\rho(\mathbb{R}^d)$  for some  $\rho \geq d$  uniformly with respect to the time variable, the weak rate of convergence  $\|\mu_{kh} - \mu_{kh}^h\|_{TV}$  is  $\frac{1}{\sqrt{kh}}(1 + \ln(k))h$  and it improves, at the terminal time, to 1 up to a logarithmic factor. Furthermore, when assuming more regularity on the probability measure  $m$  in addition to the spatial regularity on  $b$ , we improve the previous weak rate of convergence by eliminating the prefactor  $\frac{1}{\sqrt{kh}}$ . We obtain these results by comparing the mild equation satisfied by  $p(t, \cdot)$  and the perturbed mild equation satisfied by  $p^h(t, \cdot)$ . We investigate through the Lamperti transform the application of those theorems to one-dimensional SDEs with a non-constant diffusion coefficient. Sections 2.3 and 2.4 are dedicated to the proofs of the main results in Section 2.2. We finally provide numerical experiments in Section 2.5 to illustrate our results.

Beforehand, note that our results also apply to the more general case of SDEs with constant and non degenerate diffusion coefficient  $\sigma \in \mathbb{R}^d \times \mathbb{R}^d$ :

$$Y_t = Y_0 + \sigma W_t + \int_0^t \tilde{b}(s, Y_s) ds, \quad t \in [0, T]$$

where the  $\mathbb{R}^d$ -valued random variable  $Y_0$  is independent from  $(W_t)_{t \geq 0}$  and  $\tilde{b} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  is measurable and bounded. Indeed, our results remain true for this type of diffusions since the transformation  $(X_t = \sigma^{-1}Y_t)_{t \in [0, T]}$  is solution to the dynamics (2.1.1) initialized by  $X_0 = \sigma^{-1}Y_0$  for the choice of  $b(t, x) = \sigma^{-1}\tilde{b}(t, \sigma x)$ . The associated Euler scheme evolving inductively on the time grid  $(kh)_{k \in [0, \lfloor \frac{T}{h} \rfloor]}$  is defined by:

$$X_{(k+1)h}^h = X_{kh}^h + (W_{(k+1)h} - W_{kh}) + \sigma^{-1}\tilde{b}(\delta_k, \sigma X_{kh}^h)h,$$

and we can clearly see that  $(\sigma X_{kh}^h)_{k \in [0, \lfloor \frac{T}{h} \rfloor]}$  coincides exactly with the Euler scheme  $(Y_{kh}^h)_{k \in [0, \lfloor \frac{T}{h} \rfloor]}$  of the process  $(Y_t)_{t \in [0, T]}$ . Denoting by  $\tilde{\mu}_t$  the law of  $Y_t$  and by  $\tilde{\mu}_t^h$  the law of  $Y_t^h$  the continuous time interpolation of the Euler scheme, we have that  $\|\tilde{\mu}_t - \tilde{\mu}_t^h\|_{TV} = \|\mu_t - \mu_t^h\|_{TV}$ . Moreover, Lemma A.1.1 in the appendix, which relates the spatial divergences in the sense of distributions of  $y \mapsto \tilde{b}(t, y)$  and  $x \mapsto \sigma^{-1}\tilde{b}(t, \sigma x)$ , ensures that when the drift coefficient  $\tilde{b}(t, y)$  satisfies the strengthened hypotheses in Theorem 2.2.3 below, then so does  $\sigma^{-1}\tilde{b}(t, \sigma x)$ .

Before going any further, we introduce some additional notation.

### Notation:

- For  $x \in \mathbb{R}^d$ , we denote by  $|x| = \left( \sum_{i=1}^d x_i^2 \right)^{1/2}$  the euclidean norm of  $x$ .
- For  $1 \leq p < \infty$ , we denote by  $L^p(\mathbb{R}^d)$  the space of measurable functions on  $\mathbb{R}^d$  which are  $L^p$ -integrable for the Lebesgue measure i.e.  $f \in L^p$  if  $\|f\|_{L^p} = \left( \int_{\mathbb{R}^d} |f(x)|^p dx \right)^{\frac{1}{p}} < +\infty$ .

- The space  $L^\infty(\mathbb{R}^d)$  refers to the space of almost everywhere bounded measurable functions on  $\mathbb{R}^d$  endowed with the norm  $\|f\|_{L^\infty} = \inf \{C \geq 0 : |f(x)| \leq C \text{ dx a.e. on } \mathbb{R}^d\}$ .
- For notational simplicity, when a function  $g$  is defined on  $[0, T] \times \mathbb{R}^d$  and  $x \in \mathbb{R}^d$ , we may use sometimes the notation  $g_0(x) := g(0, x)$ .
- We denote by  $W^{1,1}(\mathbb{R}^d)$  the Sobolev space over  $\mathbb{R}^d$  defined as:

$$W^{1,1}(\mathbb{R}^d) \equiv \left\{ u \in L^1(\mathbb{R}^d) : \nabla u \in L^1(\mathbb{R}^d)^d \right\}$$

where  $\nabla u$  refers to the spatial derivative of  $u$  in the sense of distributions. The space is endowed with the norm  $\|u\|_{W^{1,1}} = \|u\|_{L^1} + \sum_{i=1}^d \|\partial_{x_i} u\|_{L^1}$ .

- For any open subset  $A \subset \mathbb{R}^d$ , we denote by  $\mathcal{C}_c^k(A)$  the space of real functions continuously differentiable in  $A$  up to the order  $k \in \mathbb{N}$ , with compact support on  $A$ .
- We denote by  $BV(\mathbb{R})$  the space of functions with bounded variation on  $\mathbb{R}$ . For a function  $f \in L_{\text{loc}}^1(\mathbb{R})$ , if  $f \in BV(\mathbb{R})$  then the derivative of  $f$  in the sense of distributions is a finite measure in  $\mathbb{R}$ .
- Let  $(\mu_1, \dots, \mu_{d^*})$  be signed bounded measures on  $\mathbb{R}^d$  and  $f : \mathbb{R}^d \rightarrow \mathbb{R}^{d^*}$  a  $\mathcal{C}^0$ -integrable function, with  $d^* \in \{1, d\}$ . We define the convolution product of  $f$  and  $\mu = (\mu_1, \dots, \mu_{d^*})$  by:

$$(f * \mu)(x) = \sum_{i=1}^{d^*} \int_{\mathbb{R}^d} f_i(x-y) d\mu_i(y) \text{ for } x \in \mathbb{R}^d.$$

When each  $\mu_i$  admits a density  $g_i$  with respect to the Lebesgue measure, we also denote by  $(f * g)$  this convolution product.

## 2.2 Main results

In this section, we give the main results concerning the convergence of  $\mu_t^h$ , the law of the Euler discretization with time-step  $h$  towards its limit  $\mu_t$ . We will make an intensive use of the interpretation of the total variation norm of their difference as the  $L^1$ -norm of the difference between their respective densities  $p^h(t, \cdot)$  and  $p(t, \cdot)$ .

We recall that the total variation norm for a signed measure  $\mu$  on  $\mathbb{R}^d$  is defined as:

$$\|\mu\|_{\text{TV}} = \sup_{\varphi \in \mathcal{L}} \int_{\mathbb{R}^d} \varphi(x) d\mu(x) \quad (2.2.1)$$

where  $\mathcal{L}$  denotes the set of all measurable functions  $\varphi : \mathbb{R}^d \rightarrow [-1, 1]$ . Moreover, when  $\mu$  admits a density  $f_\mu$  with respect to a reference non-negative measure  $\lambda$ , we have the following equality:

$$\|\mu\|_{\text{TV}} = \int_{\mathbb{R}^d} |f_\mu(x)| \lambda(dx). \quad (2.2.2)$$

Let us now state our estimation of the weak convergence rate of the Euler scheme towards its limit.

**Theorem 2.2.1.** *Assume  $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  is measurable and bounded by  $B < +\infty$ . Then:*

$$\exists C < +\infty, \forall h \in (0, T], \forall k \in \left[0, \left\lfloor \frac{T}{h} \right\rfloor\right], \quad \|\mu_{kh} - \mu_{kh}^h\|_{\text{TV}} \leq C\sqrt{h}.$$

**Remark 2.2.2.** • *In dimension  $d = 1$ , when specialized to the constant diffusion coefficient and absence of local time term case, Theorem 2.2 in [22] gives a finer estimation of the absolute difference between the densities by  $C\sqrt{h}$  times some Gaussian density. Our result is recovered by spatial integration. It implies that for any bounded and measurable test function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $\mathbb{E}[f(X_{kh}^h)] - \mathbb{E}[f(X_{kh})] = \mathcal{O}(\sqrt{h})$ .*

- Up to some factor  $h^{-\varepsilon}$  with  $\varepsilon$  arbitrarily small, this behaviour was proved in dimension  $d \geq 2$  for thrice continuously differentiable test functions (with polynomial growth together with their derivatives)  $f$  by Kohatsu-Higa, Lejay and Yasuda [58], when the drift coefficient is time-homogeneous and Lipschitz outside some sufficiently small set.
- Suo, Yuan and Zhang [83] proved convergence with order  $\frac{\alpha}{2}$  in total variation when the drift coefficient is time-homogeneous and satisfies some integrated against a Gaussian measure  $\alpha$ -Hölder type of regularity condition. Since  $\alpha$  appears to take values smaller than one, we obtain the better order of convergence  $1/2$  without any regularity assumption. On the other hand, we need boundedness of the drift coefficient whereas Suo, Yan and Zhang get rid of this assumption and only assume sublinear growth and even linear growth but with some restriction on the time-horizon  $T$ .

Now, when assuming more regularity on  $b$  with respect to the space variables, we obtain a better rate of convergence:

**Theorem 2.2.3.** *Assume  $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  is measurable and bounded by  $B < +\infty$ . If  $\sup_{t \in [0, T]} \|\nabla \cdot b(t, \cdot)\|_{L^\rho} < +\infty$  for some  $\rho \in [d, +\infty]$  or for  $d = 1$ ,  $\sup_{t \in [0, T]} \|\partial_x b(t, \cdot)\|_{TV} < +\infty$ ; where  $\nabla \cdot b(t, \cdot)$  and  $\partial_x b(t, \cdot)$  are respectively the spatial divergence and the spatial derivative of  $b$  in the sense of distributions. Then:*

$$\exists \tilde{C} < +\infty, \forall h \in (0, T], \forall k \in \left[0, \left\lfloor \frac{T}{h} \right\rfloor\right], \quad \|\mu_{kh} - \mu_{kh}^h\|_{TV} \leq \frac{\tilde{C}}{\sqrt{kh}} \left(1 + \ln(k)\right) h.$$

As a consequence, when  $n$  is a positive integer, we have that  $\left\|\mu_T - \mu_T^{T/n}\right\|_{TV} \leq \tilde{C}\sqrt{T} \left(\frac{1+\ln(n)}{n}\right)$ . Therefore, the order of convergence at the terminal time improves to 1 up to some logarithmic factor. This in particular applies to the bounded one-dimensional time-homogeneous drift coefficient with bounded variation defined by Suo, Yuan and Zhang in Example 2.3 [83] using some Cantor set, for which they obtain convergence with order  $1/4$  uniformly in time.

**Remark 2.2.4.** • Of course, the regularity assumption on the drift coefficient  $b$  is satisfied when it is Lipschitz in space (which is equivalent to the boundedness of its spatial gradient in the sense of distributions). When  $d \geq 2$ , the regularity assumption only involves the spatial divergence and not the full spatial gradient  $\nabla b(t, \cdot)$  in the sense of distributions. Note that if we suppose the stronger assumption  $\sup_{t \in [0, T]} \|\nabla b(t, \cdot)\|_{L^\rho} < \infty$  for  $\rho \in (d, +\infty]$ , then, according to the boundedness assumption and Corollary IX.14 [10], the drift is locally  $\left(1 - \frac{d}{\rho}\right)$ -Hölder continuous in space.

- Theorem 1.6 [60] deals with a drift coefficient bounded, continuously differentiable with bounded derivatives up to the order 2 in time and the order 4 in space outside a finite union of time-independent smooth submanifolds where it can be discontinuous. Its constant diffusion statement says that when  $m$  is a Dirac mass, then the absolute difference between the densities is bounded from above by a Gaussian density multiplied by a factor sum of a term with order  $\frac{1}{d}$ — in our total variation rate  $\frac{h}{\sqrt{kh}}$  and a term with order  $1-$  in  $h$  over the distance to the discontinuity set.

Furthermore, if we assume more regularity on  $m$  in addition to the spatial regularity on  $b$ , we obtain the following result:

**Proposition 2.2.5.** *Assume  $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  is measurable and bounded by  $B < +\infty$ . Moreover, assume that  $\sup_{t \in [0, T]} \|\nabla \cdot b(t, \cdot)\|_{L^\rho} < +\infty$  for some  $\rho \in [d, +\infty]$  or that for  $d = 1$ ,  $\sup_{t \in [0, T]} \|\partial_x b(t, \cdot)\|_{TV} < +\infty$ ; where  $\nabla \cdot b(t, \cdot)$  and  $\partial_x b(t, \cdot)$  are respectively the spatial divergence and the spatial derivative of  $b$  in the sense of distributions. If  $m$  admits a density w.r.t. the Lebesgue measure that belongs to  $W^{1,1}(\mathbb{R}^d)$  then:*

$$\exists \hat{C} < +\infty, \forall h \in (0, T], \forall k \in \left[0, \left\lfloor \frac{T}{h} \right\rfloor\right], \quad \|\mu_{kh} - \mu_{kh}^h\|_{TV} \leq \hat{C} \left(1 + \ln(k)\right) h.$$

**Remark 2.2.6.** • We can see in Theorem 2.2.3 that when  $b$  is more regular with respect to the space variables, the weak convergence rate in total variation is bounded by  $(1 + \ln(k)) h$  times a prefactor  $\frac{1}{\sqrt{kh}}$  that decreases over time and explodes in small time. According to Proposition 2.2.5, this prefactor is removed when assuming more regularity on  $m$ .

• For  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ , measurable and bounded and using Equation (2.2.1), we deduce:

- from Theorem 2.2.1 that  $|\mathbb{E}[\varphi(X_{kh}^h)] - \mathbb{E}[\varphi(X_{kh})]| \leq C\|\varphi\|_\infty\sqrt{h}$ ,
- from Theorem 2.2.3 that  $|\mathbb{E}[\varphi(X_{kh}^h)] - \mathbb{E}[\varphi(X_{kh})]| \leq \frac{\hat{C}\|\varphi\|_\infty}{\sqrt{kh}}(1 + \ln(k))h$ ,
- from Proposition 2.2.5 that  $|\mathbb{E}[\varphi(X_{kh}^h)] - \mathbb{E}[\varphi(X_{kh})]| \leq \hat{C}\|\varphi\|_\infty(1 + \ln(k))h$ .

The proofs of the two theorems and the proposition, that we will detail in Sections 2.3 and 2.4, rely on the propositions that we present in Section 2.2.2. Before going any further, we investigate through the Lamperti transform the application of those theorems to SDEs with non-constant diffusion coefficient in dimension  $d = 1$ .

### 2.2.1 Application to one-dimensional SDEs with non-constant diffusion coefficient

Let us consider the one-dimensional stochastic differential equation:

$$Y_t = Y_0 + \int_0^t \sigma(Y_s) dW_s + \int_0^t \beta(s, Y_s) ds, \quad t \in [0, T] \quad (2.2.3)$$

where  $(W_t)_{t \geq 0}$  is a one-dimensional Brownian motion independent from  $Y_0$  which is some  $(l, r)$ -valued random variable with  $-\infty \leq l < r \leq +\infty$ . Let  $z \in (l, r)$ , we assume that  $\sigma : (l, r) \rightarrow \mathbb{R}_+^*$  is a  $\mathcal{C}^1$  function with  $\lim_{y \rightarrow l^+} \int_z^y \frac{dw}{\sigma(w)} = \pm\infty$ ,  $\beta : [0, T] \times (l, r) \rightarrow \mathbb{R}$  is measurable and that  $(t, x) \mapsto \left(\frac{\beta(t, x)}{\sigma(x)} - \frac{\sigma'(x)}{2}\right)$  is bounded on  $[0, T] \times (l, r)$ . We introduce the Lamperti transform  $(X_t = \psi(Y_t))_{t \in [0, T]}$  where  $\psi : (l, r) \rightarrow \mathbb{R}$  is defined by  $\psi(y) = \int_z^y \frac{dw}{\sigma(w)}$ . By Itô's formula,  $(X_t)_{t \in [0, T]}$  is solution to the dynamics (2.1.1) for the choice  $d = 1$ ,  $b(t, \cdot) = \left(\frac{\beta(t, \cdot)}{\sigma} - \frac{\sigma'}{2}\right) \circ \psi^{-1}$  and initialized by  $\psi(Y_0)$ :

$$X_t = \psi(Y_0) + W_t + \int_0^t \left(\frac{\beta(s, \cdot)}{\sigma} - \frac{\sigma'}{2}\right) \circ \psi^{-1}(X_s) ds, \quad t \in [0, T]. \quad (2.2.4)$$

Existence and uniqueness for (2.2.3) can be deduced from the existence and uniqueness for (2.2.4). Indeed, according to [87], the SDE (2.2.4) admits a pathwise unique strong solution  $(X_t)_{t \in [0, T]}$ . By Itô's formula,  $(\psi^{-1}(X_t))_{t \in [0, T]}$  is a solution to (2.2.3). As for the uniqueness, the images of any two solutions of (2.2.3) by  $\psi$  coincide by uniqueness for (2.2.4). Since  $\psi : (l, r) \rightarrow \mathbb{R}$  is one to one, these two solutions coincide. For  $t \geq 0$ , we denote by  $\nu_t$  the probability distribution of  $Y_t$ . The law  $\mu_t$  of  $X_t$  is then the pushforward of  $\nu_t$  by  $\psi$  i.e.  $\mu_t = \psi \# \nu_t$  and conversely  $\nu_t = \psi^{-1} \# \mu_t$ .

We are going to approximate  $(Y_{kh}^h)_{k \in \llbracket 0, \lfloor \frac{T}{h} \rfloor \rrbracket}$  by  $(\psi^{-1}(X_{kh}^h))_{k \in \llbracket 0, \lfloor \frac{T}{h} \rfloor \rrbracket}$  where  $(X_{kh}^h)_{k \in \llbracket 0, \lfloor \frac{T}{h} \rfloor \rrbracket}$  is the Euler scheme of (2.2.4) with time-step  $h \in (0, T]$  initialized by  $X_0^h = \psi(Y_0)$  and evolving inductively on the time grid  $(kh)_{k \in \llbracket 0, \lfloor \frac{T}{h} \rfloor \rrbracket}$  by:

$$X_{(k+1)h}^h = X_{kh}^h + (W_{(k+1)h} - W_{kh}) + \left(\frac{\beta(\delta_k, \cdot)}{\sigma} - \frac{\sigma'}{2}\right) \circ \psi^{-1}(X_{kh}^h) h, \quad (2.2.5)$$

where the random variables  $\delta_k$  are independent, distributed according to the uniform law on  $[kh, (k+1)h]$  and are independent from  $(Y_0, (W_t)_{t \geq 0})$ .

We denote by  $\nu_t^h$  the law of  $\psi^{-1}(X_t^h)$ . Since for  $\varphi \in \mathcal{L}$  defined right after (2.2.1),  $\varphi \circ \psi^{-1} \in \mathcal{L}$ , we have:

$$\begin{aligned} \|\nu_{kh} - \nu_{kh}^h\|_{\text{TV}} &= \|\psi^{-1} \# \mu_{kh} - \psi^{-1} \# \mu_{kh}^h\|_{\text{TV}} = \sup_{\varphi \in \mathcal{L}} \int_{\mathbb{R}} \varphi(\psi^{-1}(x)) (\mu_{kh}(dx) - \mu_{kh}^h(dx)) \\ &\leq \|\mu_{kh} - \mu_{kh}^h\|_{\text{TV}}. \end{aligned}$$

On the other hand, for  $\hat{\mathcal{L}}$  denoting the set of all measurable functions  $\hat{\varphi} : (l, r) \rightarrow [-1, 1]$ , if  $\varphi \in \mathcal{L}$  then  $\hat{\varphi} = \varphi \circ \psi \in \hat{\mathcal{L}}$  and:

$$\begin{aligned} \|\mu_{kh} - \mu_{kh}^h\|_{\text{TV}} &= \sup_{\varphi \in \mathcal{L}} \int_{\mathbb{R}} \varphi(x) (\mu_{kh}(dx) - \mu_{kh}^h(dx)) = \sup_{\varphi \in \mathcal{L}} \int_{\mathbb{R}} \varphi \circ \psi(x) (\nu_{kh}(dx) - \nu_{kh}^h(dx)) \\ &\leq \sup_{\hat{\varphi} \in \hat{\mathcal{L}}} \int_{(l,r)} \hat{\varphi}(x) (\nu_{kh}(dx) - \nu_{kh}^h(dx)) \\ &= \|\nu_{kh} - \nu_{kh}^h\|_{\text{TV}}. \end{aligned}$$

Therefore,  $\|\nu_{kh} - \nu_{kh}^h\|_{\text{TV}} = \|\mu_{kh} - \mu_{kh}^h\|_{\text{TV}}$  and we obtain directly from Theorem 2.2.1 the following result for the weak convergence rate of  $\nu_t^h$  towards  $\nu_t$ :

**Theorem 2.2.7.** Assume  $\sigma : (l, r) \rightarrow \mathbb{R}_+^*$  is  $\mathcal{C}^1$  with  $\lim_{y \rightarrow l} \int_z^y \frac{dw}{\sigma(w)} = \pm\infty$ ,  $\beta : [0, T] \times (l, r) \rightarrow \mathbb{R}$  is measurable and  $(t, x) \mapsto \left( \frac{\beta(t, x)}{\sigma(x)} - \frac{\sigma'(x)}{2} \right)$  is bounded on  $[0, T] \times (l, r)$ . Then:

$$\exists C < +\infty, \forall h \in (0, T], \forall k \in \left[0, \left\lfloor \frac{T}{h} \right\rfloor\right], \quad \|\nu_{kh} - \nu_{kh}^h\|_{\text{TV}} \leq C\sqrt{h}.$$

Let us now discuss the assumptions on  $\beta$  and  $\sigma$  in order to apply Theorem 2.2.3. According to Definition 3.4 [1], the variation  $V\left(\left(\frac{\beta(t, \cdot)}{\sigma} - \frac{\sigma'}{2}\right) \circ \psi^{-1}, \mathbb{R}\right)$  of  $\left(\frac{\beta(t, \cdot)}{\sigma} - \frac{\sigma'}{2}\right) \circ \psi^{-1}$  in  $\mathbb{R}$  is defined by:

$$\begin{aligned} V\left(\left(\frac{\beta(t, \cdot)}{\sigma} - \frac{\sigma'}{2}\right) \circ \psi^{-1}, \mathbb{R}\right) &:= \sup \left\{ \int_{\mathbb{R}} \left(\frac{\beta(t, \cdot)}{\sigma} - \frac{\sigma'}{2}\right) \circ \psi^{-1}(x) \varphi'(x) dx : \varphi \in \mathcal{C}_c^1(\mathbb{R}), \|\varphi\|_{\infty} \leq 1 \right\} \\ &:= \sup \left\{ \int_{(l,r)} \left(\frac{\beta(t, \cdot)}{\sigma} - \frac{\sigma'}{2}\right)(y) (\varphi \circ \psi)'(y) dy : \varphi \in \mathcal{C}_c^1(\mathbb{R}), \|\varphi\|_{\infty} \leq 1 \right\}. \end{aligned}$$

Since  $\psi$  is a  $\mathcal{C}^1$ -diffeomorphism from  $(l, r)$  to  $\mathbb{R}$ :

$$\begin{aligned} V\left(\left(\frac{\beta(t, \cdot)}{\sigma} - \frac{\sigma'}{2}\right) \circ \psi^{-1}, \mathbb{R}\right) &:= \sup \left\{ \int_{(l,r)} \left(\frac{\beta(t, \cdot)}{\sigma} - \frac{\sigma'}{2}\right)(y) \tilde{\varphi}'(y) dy : \tilde{\varphi} \in \mathcal{C}_c^1((l, r)), \|\tilde{\varphi}\|_{\infty} \leq 1 \right\} \\ &= V\left(\left(\frac{\beta(t, \cdot)}{\sigma} - \frac{\sigma'}{2}\right), (l, r)\right). \end{aligned}$$

Moreover, using Proposition 3.6 [1], we have that:

$$\|\partial_x b(t, \cdot)\|_{\text{TV}} = V\left(\left(\frac{\beta(t, \cdot)}{\sigma} - \frac{\sigma'}{2}\right) \circ \psi^{-1}, \mathbb{R}\right) = V\left(\left(\frac{\beta(t, \cdot)}{\sigma} - \frac{\sigma'}{2}\right), (l, r)\right) = \left\| \partial_x \left(\frac{\beta(t, \cdot)}{\sigma} - \frac{\sigma'}{2}\right) \right\|_{\text{TV}}$$

where the spatial derivatives are defined in the sense of distributions on  $\mathbb{R}$  and  $(l, r)$ . Therefore, when assuming more regularity on  $\left(\frac{\beta(t, \cdot)}{\sigma} - \frac{\sigma'}{2}\right)$  with respect to the space variables, we obtain a better rate of convergence:

**Theorem 2.2.8.** Assume  $\sigma : (l, r) \rightarrow \mathbb{R}_+^*$  is  $\mathcal{C}^1$  with  $\lim_{y \rightarrow l} \int_z^y \frac{dw}{\sigma(w)} = \pm\infty$ ,  $\beta : [0, T] \times (l, r) \rightarrow \mathbb{R}$  is measurable and  $(t, x) \mapsto \left( \frac{\beta(t, x)}{\sigma(x)} - \frac{\sigma'(x)}{2} \right)$  is bounded on  $[0, T] \times (l, r)$ . Moreover, assume that  $\sup_{t \in [0, T]} \left\| \partial_x \left( \frac{\beta(t, \cdot)}{\sigma} - \frac{\sigma'}{2} \right) \right\|_{\text{TV}} < +\infty$  where the spatial derivative is defined in the sense of distributions on  $(l, r)$ . Then:

$$\exists \tilde{C} < +\infty, \forall h \in (0, T], \forall k \in \left[0, \left\lfloor \frac{T}{h} \right\rfloor\right], \quad \|\nu_{kh} - \nu_{kh}^h\|_{\text{TV}} \leq \frac{\tilde{C}}{\sqrt{kh}} (1 + \ln(k)) h.$$



We will now discuss the additional assumptions on the law of  $Y_0$  and  $\sigma$  in order to apply Proposition 2.2.5. Let us assume that  $Y_0$  admits a density  $q_0$  w.r.t. the Lebesgue measure. By a change of variables,  $X_0$  admits the density  $(\sigma q_0) \circ \psi^{-1}$  w.r.t. the Lebesgue measure. Since  $q_0 \in L^1((l, r))$  and  $\sigma$  is  $\mathcal{C}^1$  so locally bounded on  $(l, r)$ ,  $(\sigma q_0)$  defines a distribution. Moreover, we assume that  $(\sigma q_0)' \in L^1((l, r))$  where the derivative of  $(\sigma q_0)$  is defined in the sense of distributions. Now, let  $I$  be a compact subinterval of  $(l, r)$  and let  $\varphi$  be a  $\mathcal{C}_c^\infty$  function on  $\mathbb{R}$  such that  $(\varphi \circ \psi)$  is null outside  $I$ . We have, through a change of variables, that:

$$\int_{\mathbb{R}} \varphi'(x) (\sigma q_0) \circ \psi^{-1}(x) dx = \int_{(l, r)} (\varphi \circ \psi)'(y) (\sigma q_0)(y) dy.$$

Since  $(\varphi \circ \psi) \in W^{1,1}(I)$  and  $(\sigma q_0) \in W^{1,1}(I)$ , we can apply Corollary VIII.9 [10] and obtain that:

$$\int_{\mathbb{R}} \varphi'(x) (\sigma q_0) \circ \psi^{-1}(x) dx = - \int_{(l, r)} (\varphi \circ \psi)(y) (\sigma q_0)'(y) dy = - \int_{\mathbb{R}} \varphi(x) (\sigma q_0)' \circ \psi^{-1}(x) \sigma \circ \psi^{-1}(x) dx.$$

Hence, we get, in the sense of distributions, that  $((\sigma q_0) \circ \psi^{-1})' = (\sigma q_0)' \circ \psi^{-1} \times \sigma \circ \psi^{-1}$ . Through a change of variables, we obtain that  $\left\| ((\sigma q_0) \circ \psi^{-1})' \right\|_{L^1(\mathbb{R})} = \left\| (\sigma q_0)' \right\|_{L^1((l, r))}$  and since  $(\sigma q_0)' \in L^1((l, r))$ , we conclude that the density of  $X_0$  is in  $W^{1,1}(\mathbb{R})$ .

**Proposition 2.2.9.** Assume  $\sigma : (l, r) \rightarrow \mathbb{R}_+^*$  is  $\mathcal{C}^1$  with  $\lim_{y \rightarrow l^+} \int_z^y \frac{dw}{\sigma(w)} = \pm\infty$ ,  $\beta : [0, T] \times (l, r) \rightarrow \mathbb{R}$  is measurable and  $(t, x) \mapsto \left( \frac{\beta(t, x)}{\sigma(x)} - \frac{\sigma'(x)}{2} \right)$  is bounded on  $[0, T] \times (l, r)$ . Moreover, assume that  $\sup_{t \in [0, T]} \left\| \partial_x \left( \frac{\beta(t, \cdot)}{\sigma} - \frac{\sigma'}{2} \right) \right\|_{\text{TV}} < +\infty$  where the spatial derivative is defined in the sense of distributions on  $(l, r)$ . If  $Y_0$  admits a density  $q_0$  such that  $\|(\sigma q_0)'\|_{L^1((l, r))} < +\infty$ , where the spatial derivative is defined in the sense of distributions, then:

$$\exists \hat{C} < +\infty, \forall h \in (0, T], \forall k \in \left[ 0, \left\lfloor \frac{T}{h} \right\rfloor \right], \quad \|\nu_{kh} - \nu_{kh}^h\|_{\text{TV}} \leq \hat{C} (1 + \ln(k)) h.$$

### 2.2.2 Existence of densities and mild equations

We are going to state, in the next result, the existence for  $t > 0$  of the densities  $p(t, \cdot)$  and  $p^h(t, \cdot)$  by showing that  $p(t, \cdot)$  solves a mild equation and  $p^h(t, \cdot)$  solves a perturbed version of this mild equation. Let  $G_t(x) = \exp\left(-\frac{|x|^2}{2t}\right) / \sqrt{(2\pi t)^d}$  denote the heat kernel in  $\mathbb{R}^d$ , we have:

**Proposition 2.2.10.** For each  $t \in (0, T]$ ,  $\mu_t$  and  $\mu_t^h$  admit densities  $p(t, \cdot)$  and  $p^h(t, \cdot)$  with respect to the Lebesgue measure on  $\mathbb{R}^d$  s.t. we have dx a.e.:

$$p(t, x) = G_t * m(x) - \int_0^t \nabla G_{t-s} * (b(s, \cdot) p(s, \cdot))(x) ds, \quad (2.2.6)$$

$$\forall h \in (0, T], \quad p^h(t, x) = G_t * m(x) - \int_0^t \mathbb{E} \left[ \nabla G_{t-s}(x - X_s^h) \cdot b(\delta_{\ell_s}, X_{\tau_s^h}^h) \right] ds. \quad (2.2.7)$$

*Proof.* Let  $t > 0$ ,  $f$  be a  $\mathcal{C}^2$  and compactly supported function on  $\mathbb{R}^d$ . We set  $\varphi(s, x) = G_{t-s} * f(x)$  for  $(s, x) \in [0, t] \times \mathbb{R}^d$  and  $\varphi(t, x) = f(x)$ . The function  $\varphi(s, x)$  is continuously differentiable w.r.t.  $s$  and twice continuously differentiable w.r.t.  $x$  on  $[0, t] \times \mathbb{R}^d$  with bounded derivatives and solves

$$\partial_s \varphi(s, x) + \frac{1}{2} \Delta \varphi(s, x) = 0 \text{ for } (s, x) \in [0, t] \times \mathbb{R}^d. \quad (2.2.8)$$

We compute  $\mathbb{E}[\varphi(t, X_t)]$  where  $(X_s)_{s \geq 0}$  solves (2.1.1). Using (2.2.8), applying Ito's formula and taking expectations, we obtain that:

$$\mathbb{E}[\varphi(t, X_t)] = \mathbb{E} \left[ \varphi(0, X_0) + \int_0^t \nabla \varphi(s, X_s) \cdot b(s, X_s) ds \right].$$

By Fubini's Theorem and since  $G_t$  is even, we obtain:

$$\begin{aligned}\mathbb{E}[f(X_t)] &= \int_{\mathbb{R}^d} G_t * f(x) m(dx) + \int_{(0,t] \times \mathbb{R}^d} \sum_{i=1}^d \int_{\mathbb{R}^d} \partial_{x_i} G_{t-s}(x-y) f(y) dy b_i(s, x) \mu_s(dx) ds \\ &= \int_{\mathbb{R}^d} f(x) \left( G_t * m(x) - \int_0^t \nabla G_{t-s} * (b(s, \cdot) p(s, \cdot)) (x) ds \right) dx.\end{aligned}$$

Since  $f$  is arbitrary, we conclude that  $X_t$  admits a density that we denote by  $p(t, \cdot)$  and that satisfies the mild formulation (2.2.6).

Let us establish that  $X_t^h$  admits a density  $p^h(t, \cdot)$  that satisfies a perturbed version of the previous equation. Using similar arguments, we get:

$$\mathbb{E}[\varphi(t, X_t^h)] = \mathbb{E}\left[\varphi(0, X_0) + \int_0^t \nabla \varphi(s, X_s^h) \cdot b(\delta_{\ell_s}, X_{\tau_s^h}^h) ds\right].$$

Once again, by Fubini's Theorem and since  $G_t$  is even, we obtain:

$$\begin{aligned}\mathbb{E}[f(X_t^h)] &= \int_{\mathbb{R}^d} G_t * f(x) m(x) dx + \mathbb{E}\left[\int_{(0,t] \times \mathbb{R}^d} (\nabla G_{t-s}(X_s^h - x) f(x) dx) \cdot b(\delta_{\ell_s}, X_{\tau_s^h}^h) ds\right] \\ &= \int_{\mathbb{R}^d} f(x) \left( G_t * m(x) - \int_0^t \mathbb{E}[\nabla G_{t-s}(x - X_s^h) \cdot b(\delta_{\ell_s}, X_{\tau_s^h}^h)] ds \right) dx.\end{aligned}$$

The function  $f$  being arbitrary, we can conclude.  $\square$

Now, let us put (2.2.7) in a form closer to (2.2.6) but with an additional perturbation term that we control in the following proposition.

**Proposition 2.2.11.** *Assume  $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  is measurable and bounded by  $B < +\infty$ . Then:*

$$\begin{aligned}\forall h \in (0, T], \forall k \in \left[1, \left\lfloor \frac{T}{h} \right\rfloor\right], \quad p^h(kh, \cdot) &= G_{kh} * m - \int_0^{kh} \nabla G_{kh-\tau_s^h} * \left(b(s, \cdot) \mu_{\tau_s^h}^h\right) ds + R^h(k, \cdot), \\ \text{where} \quad \|R^h(k, \cdot)\|_{L^1} &\leq 2dB^2 \left(1 + \frac{d-1}{\pi}\right) \left(\frac{1}{2} + \ln(k)\right) h.\end{aligned}$$

*Proof.* Let  $k \in \left[1, \left\lfloor \frac{T}{h} \right\rfloor\right]$ . By (2.2.7) written for  $t = kh$  and since  $X_s^h = X_{jh}^h + (W_s - W_{jh}) + b(\delta_j, X_{jh}^h)(s - jh)$  for  $s \in [jh, (j+1)h)$ , we have  $dx$  a.e.:

$$\begin{aligned}p^h(kh, x) &= G_{kh} * m(x) - \sum_{j=0}^{k-1} \int_{jh}^{(j+1)h} \mathbb{E}\left[\mathbb{E}\left[\nabla G_{kh-s}(x - X_s^h) \cdot b(\delta_j, X_{jh}^h) \mid X_{jh}^h, \delta_j\right]\right] ds \\ &= G_{kh} * m(x) - \sum_{j=0}^{k-1} \int_{jh}^{(j+1)h} \mathbb{E}\left[\mathbb{E}\left[\nabla G_{kh-s}\left(x - X_{jh}^h - (W_s - W_{jh}) - b(\delta_j, X_{jh}^h)(s - jh)\right) \mid X_{jh}^h, \delta_j\right] \cdot b(\delta_j, X_{jh}^h)\right] ds.\end{aligned}$$

Using the independence between the increments  $(W_s - W_{jh})_{s \geq jh}$  and  $(X_{jh}^h, \delta_j)$  as well as the fact that the heat kernel is a convolution semi-group s.t. for  $0 \leq u < s < t$ ,  $\nabla G_{t-s} * G_{s-u} = \nabla G_{t-u}$ , we deduce:

$$p^h(kh, x) = G_{kh} * m(x) - \sum_{j=0}^{k-1} \int_{jh}^{(j+1)h} \mathbb{E}\left[\nabla G_{kh-jh}\left(x - X_{jh}^h - b(\delta_j, X_{jh}^h)(s - jh)\right) \cdot b(\delta_j, X_{jh}^h)\right] ds.$$

Using Taylor's formula with integral reminder at first order, we obtain:

$$\begin{aligned}\nabla G_{kh-jh}\left(x - X_{jh}^h - b(\delta_j, X_{jh}^h)(s - jh)\right) &= \nabla G_{kh-jh}\left(x - X_{jh}^h\right) - (s - jh) \int_0^1 \sum_{i=1}^d \partial_{x_i} \nabla G_{kh-jh}\left(x - X_{jh}^h - \alpha b(\delta_j, X_{jh}^h)(s - tj)\right) b_i(\delta_j, X_{jh}^h) d\alpha.\end{aligned}$$



We plug this equality in the previous equation. We can easily see by induction using Equation (2.1.2) that  $\delta_j$  is independent from  $X_{jh}^h$  for  $j \leq \lfloor \frac{T}{h} \rfloor$ . Therefore, we obtain:

$$\begin{aligned} p^h(kh, x) - G_{kh} * m(x) &+ \sum_{j=0}^{k-1} \int_{jh}^{(j+1)h} \nabla G_{kh-jh} * \left( b(s, x) \mu_{jh}^h(dx) \right) ds \\ &= \sum_{j=0}^{k-1} \int_{jh}^{(j+1)h} (s - jh) \mathbb{E} \left[ \int_0^1 \sum_{i=1}^d \sum_{l=1}^d \frac{\partial^2}{\partial x_i \partial x_l} G_{kh-jh} \left( x - X_{jh}^h - \alpha b(\delta_j, X_{jh}^h) (s - jh) \right) b_i(\delta_j, X_{jh}^h) b_l(\delta_j, X_{jh}^h) d\alpha \right] ds. \end{aligned}$$

We denote by  $R^h(k, x)$  the right-hand side of the previous equation. To upper-bound the  $L^1$ -norm of  $R^h$ , we use the estimates (A.1.4) and (A.1.5) from Lemma A.1.2, and the boundedness of  $b$  to obtain:

$$\begin{aligned} \|R^h(k, \cdot)\|_{L^1} &\leq \int_0^{kh} (s - \tau_s^h) \sum_{i=1}^d \sum_{l=1}^d \left\| \frac{\partial^2}{\partial x_i \partial x_l} G_{kh-\tau_s^h} \right\|_{L^1} \|b_i\|_{L^\infty} \|b_l\|_{L^\infty} ds \\ &\leq 2dB^2 \left( 1 + \frac{d-1}{\pi} \right) \int_0^{kh} \frac{s - \tau_s^h}{kh - \tau_s^h} ds \\ &\leq 2dB^2 \left( 1 + \frac{d-1}{\pi} \right) \left( \int_0^{(k-1)h} \frac{h}{kh - s} ds + \int_{(k-1)h}^{kh} \frac{s - (k-1)h}{h} ds \right) \\ &= 2dB^2 \left( 1 + \frac{d-1}{\pi} \right) \left( -h \ln(h) + h \ln(kh) + \frac{h}{2} \right). \end{aligned}$$

One can easily conclude.  $\square$

## 2.3 Proof of the convergence rate in total variation

To prove Theorem 2.2.1, we need the following Lemma that gives an estimation of the regularity of  $p(t, \cdot)$  with respect to the time variable.

**Lemma 2.3.1.** *Assume  $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  is measurable and bounded by  $B < +\infty$ .*

$$\exists Q < +\infty, \forall 0 < r \leq s \leq T, \quad \|p(s, \cdot) - p(r, \cdot)\|_{L^1} \leq Q \left( \ln(s/r) + \sqrt{s-r} \right).$$

*Proof.* Let  $0 < r \leq s \leq T$ . We have:

$$p(s, \cdot) - p(r, \cdot) = (G_s - G_r) * m - \int_0^s \nabla G_{s-u} * \left( b(u, \cdot) p(u, \cdot) \right) du + \int_0^r \nabla G_{r-u} * \left( b(u, \cdot) p(u, \cdot) \right) du.$$

Using the estimates (A.1.2), (A.1.3) and (A.1.6) from Lemma A.1.2, we obtain:

$$\begin{aligned} &\|p(s, \cdot) - p(r, \cdot)\|_{L^1} \\ &\leq \|(G_s - G_r) * m\|_{L^1} + \left\| \int_0^r (\nabla G_{s-u} - \nabla G_{r-u}) * \left( b(u, \cdot) p(u, \cdot) \right) du \right\|_{L^1} + \left\| \int_r^s \nabla G_{s-u} * \left( b(u, \cdot) p(u, \cdot) \right) du \right\|_{L^1} \\ &\leq \left\| \int_r^s (\partial_u G_u * m) du \right\|_{L^1} + \left\| \int_0^r \int_{r-u}^{s-u} \partial_\theta \nabla G_\theta * \left( b(u, \cdot) p(u, \cdot) \right) d\theta du \right\|_{L^1} + dB \int_r^s \sum_{i=1}^d \|\partial_{x_i} G_{s-u}\|_{L^1} du \\ &\leq \int_r^s \|\partial_u G_u\|_{L^1} du + \frac{B}{2} \int_0^r \int_{r-u}^{s-u} \sum_{i=1}^d \sum_{j=1}^d \left\| \frac{\partial^3 G_\theta}{\partial x_i \partial x_j^2} \right\|_{L^1} d\theta du + 2\sqrt{\frac{2}{\pi}} dB \sqrt{s-r} \\ &\leq d \ln(s/r) + 2\sqrt{\frac{2}{\pi}} (2d+3) dB \left( \sqrt{s-r} - (\sqrt{s} - \sqrt{r}) \right) + 2\sqrt{\frac{2}{\pi}} dB \sqrt{s-r}. \end{aligned}$$

The conclusion holds with  $Q = d \max \left( 1, 4\sqrt{\frac{2}{\pi}} (d+2) B \right)$ .  $\square$

We are now ready to prove Theorem 2.2.1. Since  $\mu_0 = \mu_0^h = m$ , using Equality (2.2.2) and Proposition 2.2.10, to prove the theorem amounts to prove that:

$$\exists C < +\infty, \forall h \in (0, T], \forall k \in \left[1, \left\lfloor \frac{T}{h} \right\rfloor\right], \quad \|p^h(kh, \cdot) - p(kh, \cdot)\|_{L^1} \leq C\sqrt{h}.$$

Let  $k \in \left[1, \left\lfloor \frac{T}{h} \right\rfloor\right]$ , we have:

$$p^h(kh, \cdot) - p(kh, \cdot) = V^h(k, \cdot) + R^h(k, \cdot) - \int_0^h \left( \nabla G_{kh-s} * (b(s, \cdot)p(s, \cdot)) - \nabla G_{kh} * (b(s, \cdot)m) \right) ds$$

where  $R^h(k, \cdot)$  is defined in Proposition 2.2.11 and  $V^h(k, \cdot)$  is defined by:

$$V^h(k, \cdot) = \int_h^{kh} \left( \nabla G_{kh-s} * (b(s, \cdot)p(s, \cdot)) - \nabla G_{kh-\tau_s^h} * (b(s, \cdot)p^h(\tau_s^h, \cdot)) \right) ds.$$

Since  $V^h(1, \cdot) = 0$ , we suppose  $k \geq 2$  and express  $V^h(k, \cdot)$  as  $V^h(k, \cdot) = \sum_{p=1}^3 V_p^h(k, \cdot)$  where:

$$\begin{aligned} V_1^h(k, \cdot) &= \int_h^{kh} \left( \nabla G_{kh-s} - \nabla G_{kh-\tau_s^h} \right) * (b(s, \cdot)p(s, \cdot)) ds, \\ V_2^h(k, \cdot) &= \int_h^{kh} \nabla G_{kh-\tau_s^h} * \left( b(s, \cdot) [p(s, \cdot) - p(\tau_s^h, \cdot)] \right) ds, \\ V_3^h(k, \cdot) &= \int_h^{kh} \nabla G_{kh-\tau_s^h} * \left( b(s, \cdot) [p(\tau_s^h, \cdot) - p^h(\tau_s^h, \cdot)] \right) ds. \end{aligned}$$

On the one hand, using the estimate (A.1.3) from Lemma A.1.2, we obtain immediately that:

$$\begin{aligned} \|V_3^h(k, \cdot)\|_{L^1} &\leq \sqrt{\frac{2}{\pi}} dB \int_h^{kh} \frac{1}{\sqrt{kh - \tau_s^h}} \|p(\tau_s^h, \cdot) - p^h(\tau_s^h, \cdot)\|_{L^1} ds \\ &= \sqrt{\frac{2}{\pi}} dB \sum_{j=1}^{k-1} \frac{h}{\sqrt{kh - jh}} \|p(jh, \cdot) - p^h(jh, \cdot)\|_{L^1}. \end{aligned} \quad (2.3.1)$$

On the other hand, using the estimate (A.1.3) from Lemma A.1.2, we have for the first time-step that:

$$\begin{aligned} \left\| \int_0^h \left( \nabla G_{kh-s} * (b(s, \cdot)p(s, \cdot)) - \nabla G_{kh} * (b(s, \cdot)m) \right) ds \right\|_{L^1} &\leq B \int_0^h \sum_{i=1}^d \left( \|\partial_{x_i} G_{kh-s}\|_{L^1} + \|\partial_{x_i} G_{kh}\|_{L^1} \right) ds \\ &= \sqrt{\frac{2}{\pi}} dB \left( 2 \left( \sqrt{kh} - \sqrt{(k-1)h} \right) + \frac{h}{\sqrt{kh}} \right) \\ &\leq \sqrt{\frac{2}{\pi}} dB \frac{3h}{\sqrt{kh}}. \end{aligned} \quad (2.3.2)$$

Now, using Inequality (2.3.1) and Proposition 2.2.11 for the first inequality, Inequality (2.3.2) for the second inequality and finally the fact that  $\ln(k) \leq \ln\left(\frac{T}{h}\right)$  with  $\sup_{x>0} \left\{ \left(\frac{1}{2} + \ln\left(\frac{T}{x}\right)\right) \sqrt{x} \right\}$  is attained for  $x = Te^{-\frac{3}{2}}$  for the third inequality, we obtain:

$$\begin{aligned}
& \|p(kh, \cdot) - p^h(kh, \cdot)\|_{L^1} \\
& \leq \|V_1^h(k, \cdot)\|_{L^1} + \|V_2^h(k, \cdot)\|_{L^1} + \sqrt{\frac{2}{\pi}} dB \sum_{j=1}^{k-1} \frac{h}{\sqrt{kh-jh}} \|p(jh, \cdot) - p^h(jh, \cdot)\|_{L^1} \\
& \quad + 2dB^2 \left(1 + \frac{d-1}{\pi}\right) \left(\frac{1}{2} + \ln(k)\right) h + \left\| \int_0^h \left( \nabla G_{kh-s} * (b(s, \cdot)p(s, \cdot)) - \nabla G_{kh} * (b(s, \cdot)m) \right) ds \right\|_{L^1}
\end{aligned} \tag{2.3.3}$$

$$\begin{aligned}
& \leq \|V_1^h(k, \cdot)\|_{L^1} + \|V_2^h(k, \cdot)\|_{L^1} + \sqrt{\frac{2}{\pi}} dB \sum_{j=1}^{k-1} \frac{h}{\sqrt{kh-jh}} \|p(jh, \cdot) - p^h(jh, \cdot)\|_{L^1} \\
& \quad + 2dB^2 \left(1 + \frac{d-1}{\pi}\right) \left(\frac{1}{2} + \ln(k)\right) h + \sqrt{\frac{2}{\pi}} dB \frac{3h}{\sqrt{kh}}
\end{aligned} \tag{2.3.4}$$

$$\begin{aligned}
& \leq \|V_1^h(k, \cdot)\|_{L^1} + \|V_2^h(k, \cdot)\|_{L^1} + \sqrt{\frac{2}{\pi}} dB \sum_{j=1}^{k-1} \frac{h}{\sqrt{kh-jh}} \|p(jh, \cdot) - p^h(jh, \cdot)\|_{L^1} \\
& \quad + \sqrt{\frac{2}{\pi}} dB \left( 2\sqrt{\frac{2T}{\pi}} B(\pi + d - 1)e^{-\frac{3}{4}} + 3 \right) \sqrt{h}.
\end{aligned} \tag{2.3.5}$$

Let us now estimate  $\|V_1^h(k, \cdot)\|_{L^1}$  and  $\|V_2^h(k, \cdot)\|_{L^1}$  for  $k \geq 2$ .

- For  $p = 1$ , using the estimates (A.1.3) and (A.1.6) from Lemma A.1.2, we obtain:

$$\begin{aligned}
\|V_1^h(k, \cdot)\|_{L^1} & \leq \frac{B}{2} \int_h^{(k-1)h} \int_{kh-s}^{kh-\tau_s^h} \sum_{i=1}^d \sum_{j=1}^d \left\| \frac{\partial^3 G_r}{\partial x_i \partial x_j^2} \right\|_{L^1} dr ds + B \int_{(k-1)h}^{kh} \left( \|\nabla G_{kh-s}\|_{L^1} + \|\nabla G_h\|_{L^1} \right) ds \\
& \leq \frac{1}{2} \sqrt{\frac{2}{\pi}} (2d+3) dB \int_h^{(k-1)h} \int_{kh-s}^{kh-\tau_s^h} \frac{dr}{r^{3/2}} ds + dB \sqrt{\frac{2}{\pi}} \int_{(k-1)h}^{kh} \left( \frac{1}{\sqrt{kh-s}} + \frac{1}{\sqrt{h}} \right) ds \\
& \leq \sqrt{\frac{2}{\pi}} (2d+3) dB \int_h^{(k-1)h} \frac{h}{2(kh-s)^{3/2}} ds + 3dB \sqrt{\frac{2}{\pi}} \sqrt{h} \leq 2\sqrt{\frac{2}{\pi}} (d+3) dB \sqrt{h}.
\end{aligned}$$

- For  $p = 2$ , using the estimate (A.1.3) from Lemma A.1.2 and Lemma 2.3.1, we obtain:

$$\begin{aligned}
\|V_2^h(k, \cdot)\|_{L^1} & \leq \sqrt{\frac{2}{\pi}} dB \int_h^{kh} \frac{1}{\sqrt{kh-\tau_s^h}} \|p(s, \cdot) - p(\tau_s^h, \cdot)\|_{L^1} ds \\
& \leq \sqrt{\frac{2}{\pi}} dB Q \left( \int_h^{kh} \frac{\ln(s/\tau_s^h)}{\sqrt{kh-\tau_s^h}} ds + \int_h^{kh} \frac{\sqrt{s-\tau_s^h}}{\sqrt{kh-\tau_s^h}} ds \right).
\end{aligned}$$

Using the fact that for  $s \geq h$ ,  $\ln\left(\frac{s}{\tau_s^h}\right) \leq \frac{s-\tau_s^h}{\tau_s^h} \leq \frac{2h}{s}$  and Lemma A.1.4, we have that:

$$\int_h^{kh} \frac{\ln(s/\tau_s^h)}{\sqrt{kh-\tau_s^h}} ds \leq \int_h^{kh} \frac{2h}{s\sqrt{kh-s}} ds \leq \frac{2h}{\sqrt{kh}} \ln(4k). \tag{2.3.6}$$

Moreover, using the fact that  $\frac{\sqrt{s-\tau_s^h}}{\sqrt{kh-\tau_s^h}} \leq \frac{\sqrt{h}}{\sqrt{kh-s}}$ , we deduce that:

$$\|V_2^h(k, \cdot)\|_{L^1} \leq 2\sqrt{\frac{2}{\pi}} dB Q \left( \sup_{x \geq 1} \frac{\ln(4x)}{\sqrt{x}} + \sqrt{(k-1)h} \right) \sqrt{h} \leq 2\sqrt{\frac{2}{\pi}} dB Q \left( \frac{4}{e} + \sqrt{T} \right) \sqrt{h}.$$

Hence, using (2.3.5) and the above estimates for the quantities  $\|V_1^h(k, \cdot)\|_{L^1}$  and  $\|V_2^h(k, \cdot)\|_{L^1}$ , we obtain:

$$\|p(kh, \cdot) - p^h(kh, \cdot)\|_{L^1} \leq L\sqrt{h} + \sqrt{\frac{2}{\pi}}dB \sum_{j=1}^{k-1} \frac{h}{\sqrt{kh-jh}} \|p(jh, \cdot) - p^h(jh, \cdot)\|_{L^1}$$

where  $L = 2\sqrt{\frac{2}{\pi}}dB \left( \left(d + \frac{9}{2}\right) + Q\left(\frac{4}{e} + \sqrt{T}\right) + \sqrt{\frac{2T}{\pi}}B(\pi + d - 1)e^{-\frac{3}{4}} \right)$ . We iterate this inequality to obtain:

$$\|p(kh, \cdot) - p^h(kh, \cdot)\|_{L^1} \leq L \left( 1 + 2\sqrt{\frac{2}{\pi}}dB\sqrt{(k-1)h} \right) \sqrt{h} + \frac{2d^2B^2}{\pi} \sum_{j=1}^{k-1} \sum_{l=1}^{j-1} \frac{h}{\sqrt{k-j}\sqrt{j-l}} \|p(lh, \cdot) - p^h(lh, \cdot)\|_{L^1}.$$

We re-write the double-sum the following way:

$$\begin{aligned} \sum_{j=1}^{k-1} \sum_{l=1}^{j-1} \frac{h}{\sqrt{k-j}\sqrt{j-l}} \|p(lh, \cdot) - p^h(lh, \cdot)\|_{L^1} &= \sum_{l=1}^{k-2} \sum_{j=l+1}^{k-1} \frac{h}{\sqrt{k-j}\sqrt{j-l}} \|p(lh, \cdot) - p^h(lh, \cdot)\|_{L^1} \\ &= h \sum_{l=1}^{k-2} \left( \|p(lh, \cdot) - p^h(lh, \cdot)\|_{L^1} \sum_{i=1}^{k-l-1} \frac{1}{\sqrt{i}\sqrt{(k-l)-i}} \right) \\ &\leq \pi \int_h^{kh} \|p(\tau_s^h, \cdot) - p^h(\tau_s^h, \cdot)\|_{L^1} ds, \end{aligned}$$

where we used Lemma A.1.3 for the last inequality. Therefore,

$$\|p(kh, \cdot) - p^h(kh, \cdot)\|_{L^1} \leq L \left( 1 + 2\sqrt{\frac{2T}{\pi}}dB \right) \sqrt{h} + 2d^2B^2h \sum_{j=1}^{k-1} \|p(jh, \cdot) - p^h(jh, \cdot)\|_{L^1}.$$

We apply Lemma A.1.6 and obtain:

$$\begin{aligned} &\|p(kh, \cdot) - p^h(kh, \cdot)\|_{L^1} \\ &\leq L \left( 1 + 2\sqrt{\frac{2T}{\pi}}dB \right) \sqrt{h} + 2d^2B^2L \left( 1 + 2\sqrt{\frac{2T}{\pi}}dB \right) \sqrt{h} \sum_{j=1}^{k-1} h \exp \left( 2d^2B^2(kh - (j+1)h) \right) \\ &\leq L \left( 1 + 2\sqrt{\frac{2T}{\pi}}dB \right) \sqrt{h} + 2d^2B^2L \left( 1 + 2\sqrt{\frac{2T}{\pi}}dB \right) \sqrt{h} \exp \left( 2d^2B^2T \right) (kh - h) \\ &\leq L \left( 1 + 2\sqrt{\frac{2T}{\pi}}dB \right) \left( 1 + 2d^2B^2T \exp \left( 2d^2B^2T \right) \right) \sqrt{h}. \end{aligned}$$

The conclusion holds with  $C = L \left( 1 + 2\sqrt{\frac{2T}{\pi}}dB \right) \left( 1 + 2d^2B^2T \exp \left( 2d^2B^2T \right) \right)$ .

## 2.4 Proof of the convergence rate in total variation when assuming more regularity on $b$ w.r.t. to the space variables

The following proposition, developed in Subsection 2.4.1, enables to establish an estimate of the total variation norm of the divergence of  $b(t, \cdot)p(t, \cdot)$  for  $t > 0$  from the regularity assumed on  $b$  w.r.t. to the space variables. When assuming extra regularity on  $m$ , the estimate is improved.

**Proposition 2.4.1.** *Assume  $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  is measurable and bounded by  $B < +\infty$ . If  $\sup_{t \in [0, T]} \|\nabla \cdot b(t, \cdot)\|_{L^\rho} < +\infty$  for some  $\rho \in [d, +\infty]$  or for  $d = 1$ ,  $\sup_{t \in [0, T]} \|\partial_x b(t, \cdot)\|_{TV} < +\infty$ ; where  $\nabla \cdot b(t, \cdot)$  and  $\partial_x b(t, \cdot)$  are respectively the spatial divergence and the spatial derivative of  $b$  in the sense of distributions. Then:*

$$\exists M < +\infty, \forall t \in (0, T], \quad \left\| \nabla \cdot (b(t, \cdot)p(t, \cdot)) \right\|_{TV} \leq \frac{M}{\sqrt{t}}, \quad (2.4.1)$$

and:

$$\forall t \in (0, T], \quad p(t, \cdot) = G_t * m - \int_0^t G_{t-s} * \nabla \cdot (b(s, \cdot) p(s, \cdot)) ds. \quad (2.4.2)$$

Moreover, if  $m$  admits a density w.r.t. the Lebesgue measure that belongs to  $W^{1,1}(\mathbb{R}^d)$ , we obtain:

$$\exists \tilde{M} < +\infty, \forall t \in [0, T], \quad \left\| \nabla \cdot (b(t, \cdot) p(t, \cdot)) \right\|_{\text{TV}} \leq \tilde{M}. \quad (2.4.3)$$

**Remark 2.4.2.** In fact, when  $\sup_{t \in [0, T]} \|\nabla \cdot b(t, \cdot)\|_{L^\rho} < +\infty$  for some  $\rho \in [d, +\infty]$ , we will prove that for  $t \in (0, T]$ ,  $\nabla \cdot (b(t, \cdot) p(t, \cdot)) \in L^1(\mathbb{R}^d)$  and  $\left\| \nabla \cdot (b(t, \cdot) p(t, \cdot)) \right\|_{\text{TV}} = \left\| \nabla \cdot (b(t, \cdot) p(t, \cdot)) \right\|_{L^1}$ .

The results of Proposition 2.4.1 will be used in the proof of Theorem 2.2.3 detailed in Subsection 2.4.2, and in the proof of Proposition 2.2.5 detailed in Subsection 2.4.3.

We bring to attention that, in this subsection, all the derivatives and divergence are defined in the sense of distributions.

### 2.4.1 Proof of Proposition 2.4.1

Let  $\theta \in (0, T]$ .

We define the Banach spaces  $\bar{\mathcal{C}}((0, \theta], L^1(\mathbb{R}^d)) = \left\{ q \in \mathcal{C}((0, \theta], L^1(\mathbb{R}^d)) : \sup_{t \in (0, \theta]} \|q(t, \cdot)\|_{L^1} < +\infty \right\}$ ,

$\tilde{\mathcal{C}}((0, \theta], W^{1,1}(\mathbb{R}^d)) = \left\{ q \in \mathcal{C}((0, \theta], W^{1,1}(\mathbb{R}^d)) : \|q\| = \sup_{t \in (0, \theta]} \|q(t, \cdot)\|_{L^1} + \sup_{t \in (0, \theta]} \sqrt{t} \|\nabla q(t, \cdot)\|_{L^1} < +\infty \right\}$   
and  $\mathcal{C}([0, \theta], W^{1,1}(\mathbb{R}^d))$  endowed respectively with the norms  $\sup_{t \in (0, \theta]} \|q(t, \cdot)\|_{L^1}$ ,  $\|q\|$  and  $\sup_{t \in [0, \theta]} \|q(t, \cdot)\|_{W^{1,1}}$ .

One has  $\mathcal{C}([0, \theta], W^{1,1}(\mathbb{R}^d)) \subset \tilde{\mathcal{C}}((0, \theta], W^{1,1}(\mathbb{R}^d)) \subset \bar{\mathcal{C}}((0, \theta], L^1(\mathbb{R}^d))$ .

The next theorem states regularity properties of the density  $(p(t, \cdot))_{t \in (0, T]}$  when assuming more regularity on  $b$  w.r.t. the space variables.

**Theorem 2.4.3.** Assume  $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  is measurable and bounded by  $B < +\infty$ . If  $\sup_{t \in [0, T]} \|\nabla \cdot b(t, \cdot)\|_{L^\rho} < +\infty$  for some  $\rho \in [d, +\infty]$  or for  $d = 1$ ,  $\sup_{t \in [0, T]} \|\partial_x b(t, \cdot)\|_{\text{TV}} < +\infty$  then  $p \in \tilde{\mathcal{C}}((0, T], W^{1,1}(\mathbb{R}^d))$ . Moreover, if  $m$  admits a density w.r.t. the Lebesgue measure in  $W^{1,1}(\mathbb{R}^d)$  then  $p \in \mathcal{C}([0, T], W^{1,1}(\mathbb{R}^d))$ .

The proof of Theorem 2.4.3 relies on the uniqueness of the mild equation (2.2.6). This latter can be proved by a fixed-point method. To do so, for  $\theta \in (0, T]$ , we define on the space  $\bar{\mathcal{C}}((0, \theta], L^1(\mathbb{R}^d))$  the map  $\Phi$ :

$$\Phi : q \mapsto \left( \Phi_t(q) = G_t * m - \int_0^t \nabla G_{t-s} * (b(s, \cdot) q(s, \cdot)) ds \right)_{t \in (0, \theta]}.$$

By a slight abuse of notation, we do not make explicit the dependence of the map  $\Phi$  on the time horizon  $\theta$ . Let us check that  $\Phi$  is well-defined. For  $t \in (0, \theta]$ , we have, using the estimate (A.1.3) from Lemma A.1.2, that:

$$\|\Phi_t(q)\|_{L^1} \leq \|G_t * m\|_{L^1} + dB \sqrt{\frac{2}{\pi}} \int_0^t \frac{1}{\sqrt{t-s}} \|q(s, \cdot)\|_{L^1} ds \leq 1 + 2dB \sqrt{\frac{2\theta}{\pi}} \sup_{u \in (0, \theta]} \|q(u, \cdot)\|_{L^1}. \quad (2.4.4)$$

Hence, since  $q \in \bar{\mathcal{C}}((0, \theta], L^1(\mathbb{R}^d))$ , we have that  $\sup_{t \in (0, \theta]} \|\Phi_t(q)\|_{L^1} < +\infty$ .

The following result ensures that the map  $\Phi$  admits a unique fixed-point in  $\bar{\mathcal{C}}((0, \theta], L^1(\mathbb{R}^d))$ .

**Lemma 2.4.4.** Assume  $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  is measurable and bounded by  $B < +\infty$ . For all  $\theta \in (0, T]$ ,  $(p(t, \cdot))_{t \in (0, \theta]}$  is the unique fixed-point of the map  $\Phi$  in  $\bar{\mathcal{C}}((0, \theta], L^1(\mathbb{R}^d))$ .

*Proof.* Let  $q \in \bar{\mathcal{C}}((0, \theta], L^1(\mathbb{R}^d))$ . Using Inequality (2.4.4), we have that  $\sup_{t \in (0, \theta]} \|\Phi_t(q)\|_{L^1} < +\infty$ . For  $0 < r \leq s \leq \theta$ , adapting the proof of Lemma 2.3.1, we obtain that:

$$\|\Phi_s(q) - \Phi_r(q)\|_{L^1} \leq \max \left( 1, \sup_{u \in [r, s]} \|q(u, \cdot)\|_{L^1} \right) Q \left( \ln(s/r) + \sqrt{s-r} \right). \quad (2.4.5)$$

Therefore,  $t \mapsto \Phi_t(q)$  is continuous on  $(0, \theta]$  with values in  $L^1(\mathbb{R}^d)$  and  $\Phi(q) \in \bar{\mathcal{C}}((0, \theta], L^1(\mathbb{R}^d))$ . Now, let  $q, \tilde{q} \in \bar{\mathcal{C}}((0, \theta], L^1(\mathbb{R}^d))$ . Using the same reasoning as for Inequality (2.4.4), we obtain:

$$\|\Phi_t(q) - \Phi_t(\tilde{q})\|_{L^1} \leq dB \sqrt{\frac{2}{\pi}} \int_0^t \frac{1}{\sqrt{t-s}} \|q(s, \cdot) - \tilde{q}(s, \cdot)\|_{L^1} ds. \quad (2.4.6)$$

Let  $n \in \mathbb{N}^*$ , we define  $\Phi^{n+1} = \Phi^n \circ \Phi = \Phi \circ \Phi^n$  and iterate Inequality (2.4.6)  $2n$ -times to obtain:

$$\begin{aligned} \|\Phi_t^{2n}(q) - \Phi_t^{2n}(\tilde{q})\|_{L^1} &\leq \left( dB \sqrt{\frac{2}{\pi}} \right)^{2n} \pi^n \int_0^t \frac{(t-s)^{n-1}}{(n-1)!} \|q(s, \cdot) - \tilde{q}(s, \cdot)\|_{L^1} ds \\ &\leq (dB)^{2n} \frac{(2\theta)^n}{n!} \sup_{u \in (0, \theta]} \|q(u, \cdot) - \tilde{q}(u, \cdot)\|_{L^1}. \end{aligned}$$

Therefore,  $\sup_{t \in (0, \theta]} \|\Phi_t^{2n}(q) - \Phi_t^{2n}(\tilde{q})\|_{L^1} \leq (dB)^{2n} \frac{(2\theta)^n}{n!} \sup_{t \in (0, \theta]} \|q(t, \cdot) - \tilde{q}(t, \cdot)\|_{L^1}$  and for  $n$  big enough,  $\Phi^{2n}$  is a contraction on  $\bar{\mathcal{C}}((0, \theta], L^1(\mathbb{R}^d))$ . By Picard's Theorem,  $\Phi^{2n}$  admits then a unique fixed-point  $q$  in  $\bar{\mathcal{C}}((0, \theta], L^1(\mathbb{R}^d))$ . We have that  $\Phi(q) = \Phi(\Phi^{2n}(q)) = \Phi^{2n}(\Phi(q))$  making  $\Phi(q)$  a fixed-point of  $\Phi^{2n}$ , but since this latter is unique, we conclude that  $\Phi(q) = q$ . For  $t > 0$ ,  $p(t, \cdot)$  is solution to the mild equation (2.2.6) and using Lemma 2.3.1,  $p \in \bar{\mathcal{C}}((0, \theta], L^1(\mathbb{R}^d))$ . Consequently,  $(p(t, \cdot))_{t \in (0, \theta]}$  is the unique fixed-point of the map  $\Phi$  in  $\bar{\mathcal{C}}((0, \theta], L^1(\mathbb{R}^d))$ .  $\square$

Now, we seek to establish more regularity on the fixed-point of the map  $\Phi$ .

**Proposition 2.4.5.** *Assume  $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  is measurable and bounded by  $B < +\infty$ . Moreover, assume that  $\sup_{t \in [0, T]} \|\nabla \cdot b(t, \cdot)\|_{L^\rho} < +\infty$  for some  $\rho \in [d, +\infty]$  or that for  $d = 1$ ,  $\sup_{t \in [0, T]} \|\partial_x b(t, \cdot)\|_{TV} < +\infty$ .*

- If  $m$  admits a density w.r.t the Lebesgue measure in  $W^{1,1}(\mathbb{R}^d)$  then for all  $\theta \in (0, T]$ ,  $\Phi$  admits a unique fixed-point in the space  $\mathcal{C}([0, \theta], W^{1,1}(\mathbb{R}^d))$ .
- Otherwise, there exists  $\theta_0 \in (0, T]$  s.t.  $\Phi$  admits a unique fixed-point in the space  $\tilde{\mathcal{C}}((0, \theta_0], W^{1,1}(\mathbb{R}^d))$ .

Let us deduce Theorem 2.4.3 before giving the proof of Proposition 2.4.5.

*Proof.* • For  $\theta = \theta_0$  given by Proposition 2.4.5:

- Let  $t \in (0, \theta_0]$ . According to Proposition 2.4.5, the map  $\Phi$  admits a unique fixed-point in  $\tilde{\mathcal{C}}((0, \theta_0], W^{1,1}(\mathbb{R}^d))$ . With the inclusion  $\tilde{\mathcal{C}}((0, \theta_0], W^{1,1}(\mathbb{R}^d)) \subset \bar{\mathcal{C}}((0, \theta_0], L^1(\mathbb{R}^d))$ , this fixed-point coincides with the unique fixed-point of  $\Phi$  in  $\bar{\mathcal{C}}((0, \theta_0], L^1(\mathbb{R}^d))$  which is  $(p(t, \cdot))_{t \in (0, \theta_0]}$  according to Lemma 2.4.4. Therefore, we have that  $(p(t, \cdot))_{t \in (0, \theta_0]} \in \tilde{\mathcal{C}}((0, \theta_0], W^{1,1}(\mathbb{R}^d))$  and  $\sup_{t \in (0, \theta_0]} \sqrt{t} \|p(t, \cdot)\|_{W^{1,1}} \leq \max(1, \sqrt{\theta_0}) \|p\| < +\infty$ .
- Now, let  $t \in [\theta_0, T]$ . Using the fact that the heat kernel is a convolution semi-group and that

$$p(\theta_0, \cdot) = G_{\theta_0} * m - \int_0^{\theta_0} \nabla G_{\theta_0-r} * (b(r, \cdot) p(r, \cdot)) dr,$$

we obtain:

$$p(t, \cdot) = G_{t-\theta_0} * p(\theta_0, \cdot) - \int_0^{t-\theta_0} \nabla G_{(t-\theta_0)-s} * (b(\theta_0 + s, \cdot) p(\theta_0 + s, \cdot)) ds$$

such that for  $u = (t - \theta_0) \in [0, T - \theta_0]$ :

$$p(\theta_0 + u, \cdot) = G_u * p(\theta_0, \cdot) - \int_0^u \nabla G_{u-s} * \left( b(\theta_0 + s, \cdot) p(\theta_0 + s, \cdot) \right) ds.$$

Hence, by Lemma 2.4.4,  $(p(\theta_0 + u, \cdot))_{u \in [0, T - \theta_0]}$  is the unique fixed-point in  $\bar{\mathcal{C}}((0, T - \theta_0], L^1(\mathbb{R}^d))$  of the functional defined like  $\Phi$  but with  $m$  replaced by  $p(\theta_0, \cdot)$  and  $b$  shifted by  $\theta_0$  in the time variable. Since  $p(\theta_0, \cdot) \in W^{1,1}(\mathbb{R}^d)$  and according to Proposition 2.4.5, this functional admits a unique fixed-point in  $\mathcal{C}([0, T - \theta_0], W^{1,1}(\mathbb{R}^d))$  and this fixed-point coincides with  $(p(\theta_0 + u, \cdot))_{u \in [0, T - \theta_0]}$ . Therefore,  $(p(t, \cdot))_{t \in [\theta_0, T]} \in \mathcal{C}([\theta_0, T], W^{1,1}(\mathbb{R}^d))$  and

$$\sup_{t \in [\theta_0, T]} \sqrt{t} \|p(t, \cdot)\|_{W^{1,1}} \leq \sqrt{T} \sup_{t \in [\theta_0, T]} \|p(t, \cdot)\|_{W^{1,1}} < +\infty.$$

We can conclude.

• Now, we assume that  $m$  admits a density w.r.t. the Lebesgue measure in  $W^{1,1}(\mathbb{R}^d)$ . For  $\theta = T$  and according to Proposition 2.4.5, the map  $\Phi$  admits a unique fixed-point in  $\mathcal{C}([0, T], W^{1,1}(\mathbb{R}^d))$ . With the inclusion  $\mathcal{C}([0, T], W^{1,1}(\mathbb{R}^d)) \subset \bar{\mathcal{C}}((0, T], L^1(\mathbb{R}^d))$ , this fixed-point coincides with the unique fixed-point of  $\Phi$  in  $\bar{\mathcal{C}}((0, T], L^1(\mathbb{R}^d))$  which is  $(p(t, \cdot))_{t \in (0, T]}$  according to Lemma 2.4.4. Therefore,  $(p(t, \cdot))_{t \in [0, T]} \in \mathcal{C}([0, T], W^{1,1}(\mathbb{R}^d))$ .  $\square$

To prove Proposition 2.4.5, we need the following convolution and derivation in the sense of distributions result.

**Lemma 2.4.6.** *Let  $q : \mathbb{R}^d \rightarrow \mathbb{R}$  be a function in  $W^{1,1}(\mathbb{R}^d)$  and  $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$  a bounded measurable function. We assume either that  $\|\nabla \cdot g\|_{L^\rho} < +\infty$  for some  $\rho \in [d, +\infty]$  or that for  $d = 1$ ,  $\|g'\|_{TV} < +\infty$ .*

• *Under the first assumption, for any  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$   $C^\infty$ -bounded together with its first order derivatives, we have:*

$$\int_{\mathbb{R}^d} \nabla \varphi(x) \cdot (q(x)g(x)) dx = - \int_{\mathbb{R}^d} \varphi(x) (\nabla q(x) \cdot g(x) + q(x) \nabla \cdot g(x)) dx$$

so that in the sense of distributions,  $\nabla \cdot (qg) = q \nabla \cdot g + \nabla q \cdot g$ .

Moreover, for  $\check{C} = \sup_{\substack{f \in W^{1,1} \\ f \neq 0}} \frac{\|f\|_{L^{\frac{\rho}{\rho-1}}(\mathbb{R}^d)}}{\|f\|_{W^{1,1}(\mathbb{R}^d)}}$  which is finite according to Corollary IX.10 [10], we have that:

$$\left\| \nabla \cdot (qg) \right\|_{L^1} \leq \check{C} \|q\|_{W^{1,1}} \|\nabla \cdot g\|_{L^\rho} + \|\nabla q\|_{L^1} \|g\|_{L^\infty}. \quad (2.4.7)$$

• *Under the second assumption, for any  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$   $C^\infty$ -bounded together with its first order derivative, we have:*

$$\int_{\mathbb{R}} \varphi'(x) (q(x)g(x)) dx = - \int_{\mathbb{R}} \varphi(x) q'(x) g(x) dx - \int_{\mathbb{R}} \varphi(x) q(x) g'(x) dx$$

where the continuous representative of  $q$  which exists according to Theorem VIII.2 [10], is chosen to define  $q(x)g'(x)$ . Moreover, the derivative of  $qg$  in the sense of distributions is a bounded measure on the real line and for  $\check{C} = \sup_{\substack{f \in W^{1,1} \\ f \neq 0}} \frac{\|f\|_\infty}{\|f\|_{W^{1,1}(\mathbb{R})}}$  which is finite according to Theorem VIII.7 [10], we have

that:

$$\|(qg)'\|_{TV} \leq \check{C} \|q\|_{W^{1,1}} \|g'\|_{TV} + \|q'\|_{L^1} \|g\|_\infty. \quad (2.4.8)$$

*Proof.* According to Theorem IX.2 [10], there exists a sequence of functions  $(q_n)_n$  in  $\mathcal{C}_c^\infty(\mathbb{R}^d)$  such that when  $n \rightarrow +\infty$ ,  $q_n \rightarrow q$  and  $\nabla q_n \rightarrow \nabla q$  respectively in  $L^1(\mathbb{R}^d)$  and  $L^1(\mathbb{R}^d)^d$ . We have:

$$\int_{\mathbb{R}^d} \nabla \varphi(x) \cdot (q_n(x)g(x)) dx = \int_{\mathbb{R}^d} \nabla (\varphi(x)q_n(x)) \cdot g(x) dx - \int_{\mathbb{R}^d} \varphi(x) \nabla q_n(x) \cdot g(x) dx. \quad (2.4.9)$$

Since  $\nabla\varphi$ ,  $g$  and  $\varphi$  are bounded functions on  $\mathbb{R}^d$ , we have  $\int_{\mathbb{R}^d} \nabla\varphi(x) \cdot (q_n(x)g(x)) dx \xrightarrow{n \rightarrow +\infty} \int_{\mathbb{R}^d} \nabla\varphi(x) \cdot (q(x)g(x)) dx$  and  $\int_{\mathbb{R}^d} \varphi(x)\nabla q_n(x) \cdot g(x) dx \xrightarrow{n \rightarrow +\infty} \int_{\mathbb{R}^d} \varphi(x)\nabla q(x) \cdot g(x) dx$ .

• Under the first assumption, since  $\varphi q_n \in C_c^\infty(\mathbb{R}^d)$ , we have that:

$$\int_{\mathbb{R}^d} \nabla(\varphi(x)q_n(x)) \cdot g(x) dx = - \int_{\mathbb{R}^d} \varphi(x)q_n(x)\nabla \cdot g(x) dx.$$

Using Hölder's inequality, we obtain that:

$$\begin{aligned} \left| \int_{\mathbb{R}^d} \varphi(x)q_n(x)\nabla \cdot g(x) dx - \int_{\mathbb{R}^d} \varphi(x)q(x)\nabla \cdot g(x) dx \right| &\leq \sup_{x \in \mathbb{R}^d} |\varphi(x)| \|q_n - q\|_{L^{\frac{\rho}{\rho-1}}} \|\nabla \cdot g\|_{L^\rho} \\ &\leq \check{C} \sup_{x \in \mathbb{R}^d} |\varphi(x)| \|q_n - q\|_{W^{1,1}} \|\nabla \cdot g\|_{L^\rho} \\ &\xrightarrow{n \rightarrow +\infty} 0. \end{aligned}$$

Hence, taking the limit  $n \rightarrow +\infty$  in Equation (2.4.9), we get, in the sense of distributions, that  $\nabla \cdot (qg) = q\nabla \cdot g + \nabla q \cdot g$ . Now, using once again Hölder's inequality and Corollary IX.10 [10], one has

$$\|q\nabla \cdot g\|_{L^1} \leq \|q\|_{L^{\frac{\rho}{\rho-1}}} \|\nabla \cdot g\|_{L^\rho} \leq \check{C} \|q\|_{W^{1,1}} \|\nabla \cdot g\|_{L^\rho}$$

and one deduces that

$$\left\| \nabla \cdot (qg) \right\|_{L^1} \leq \check{C} \|q\|_{W^{1,1}} \|\nabla \cdot g\|_{L^\rho} + \|\nabla q\|_{L^1} \|g\|_{L^\infty}.$$

• Under the second assumption, since  $\varphi q_n \in C_c^\infty(\mathbb{R})$ , we have that:

$$\int_{\mathbb{R}} (\varphi q_n)'(x)g(x) dx = - \int_{\mathbb{R}} \varphi(x)q_n(x)g'(x) dx.$$

According to Theorem VIII.2 and Theorem VIII.7 [10],  $q$  admits a bounded and continuous representative, and for this representative the integral  $\int_{\mathbb{R}} \varphi(x)q(x)g'(x) dx$  makes sense. Using Hölder's inequality, we have:

$$\begin{aligned} \left| \int_{\mathbb{R}} \varphi(x)q_n(x)g'(x) dx - \int_{\mathbb{R}} \varphi(x)q(x)g'(x) dx \right| &\leq \sup_{x \in \mathbb{R}} |\varphi(x)| \sup_{x \in \mathbb{R}} |q_n(x) - q(x)| \|g'\|_{TV} \\ &\leq \check{C} \sup_{x \in \mathbb{R}} |\varphi(x)| \|q_n - q\|_{W^{1,1}} \|g'\|_{TV} \\ &\xrightarrow{n \rightarrow +\infty} 0. \end{aligned}$$

Hence, taking the limit  $n \rightarrow +\infty$  in Equation (2.4.9), we get, in the sense of distributions, that  $(qg)' = qg' + q'g$ . Now, using once again Hölder's inequality and Theorem VIII.7 [10], one has

$$\|qg'\|_{TV} \leq \sup_{x \in \mathbb{R}} |q(x)| \|g'\|_{TV} \leq \check{C} \|q\|_{W^{1,1}} \|g'\|_{TV}$$

and one deduces that

$$\|(qg)'\|_{TV} \leq \check{C} \|q\|_{W^{1,1}} \|g'\|_{TV} + \|q'\|_{L^1} \|g\|_\infty.$$

□

We are now ready to prove Proposition 2.4.5.

*Proof.* We are going to suppose that  $\sup_{t \in [0, T]} \|\nabla \cdot b(t, \cdot)\|_{L^\rho} < +\infty$  for some  $\rho \in [d, +\infty]$ . When  $d = 1$ ,  $\sup_{t \in [0, T]} \|\partial_x b(t, \cdot)\|_{TV} < +\infty$ , the proof is analogous and the estimations remain valid when replacing  $\|\nabla \cdot b(t, \cdot)\|_{L^\rho}$  by  $\|\partial_x b(t, \cdot)\|_{TV}$ .



For  $\theta \in (0, T]$ , let  $0 \leq s < t \leq \theta$ . If  $q(s, \cdot)$  is in  $W^{1,1}(\mathbb{R}^d)$ , we apply Lemma 2.4.6 with  $\varphi = G_{t-s}$  which is  $C^\infty$ -bounded together with its first order derivatives and  $g = b(s, \cdot)$  to obtain that:

$$\nabla G_{t-s} * (b(s, \cdot)q(s, \cdot)) = G_{t-s} * \nabla \cdot (b(s, \cdot)q(s, \cdot)).$$

If  $q \in \tilde{\mathcal{C}}((0, \theta], W^{1,1}(\mathbb{R}^d))$ , we first have:

$$\forall t \in (0, \theta], \quad \Phi_t(q) = G_t * m - \int_0^t G_{t-s} * \nabla \cdot (b(s, \cdot)q(s, \cdot)) \, ds. \quad (2.4.10)$$

- We start by proving that there exists  $\theta_0$  s.t.  $\Phi$  admits a unique fixed-point in  $\tilde{\mathcal{C}}((0, \theta_0], W^{1,1}(\mathbb{R}^d))$ :

Let  $t \in (0, \theta]$  and  $q \in \tilde{\mathcal{C}}((0, \theta], W^{1,1}(\mathbb{R}^d))$ . We have, using the estimate (A.1.3) from Lemma A.1.2 and Inequality (2.4.7) from Lemma 2.4.6, that:

$$\begin{aligned} & \int_0^t \left\| \nabla G_{t-s} * \nabla \cdot (b(s, \cdot)q(s, \cdot)) \right\|_{L^1} \, ds \\ & \leq d\sqrt{\frac{2}{\pi}} \int_0^t \frac{1}{\sqrt{t-s}} \left( B + \check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right) \left( \|q(s, \cdot)\|_{L^1} + \|\nabla q(s, \cdot)\|_{L^1} \right) \, ds \\ & \leq d\sqrt{2\pi} \max(1, \sqrt{T}) \left( B + \check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right) \left( \sup_{u \in (0, \theta]} \|q(u, \cdot)\|_{L^1} + \sup_{u \in (0, \theta]} \sqrt{u} \|\nabla q(u, \cdot)\|_{L^1} \right). \end{aligned}$$

Therefore,  $\int_0^t \left\| \nabla G_{t-s} * \nabla \cdot (b(s, \cdot)q(s, \cdot)) \right\|_{L^1} \, ds \leq d\sqrt{2\pi} \max(1, \sqrt{T}) \left( B + \check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right) \|q\|$

which is finite. We can then apply Fubini's theorem and obtain that, in the sense of distributions, the gradient of  $\Phi_t(q)$  defined in (2.4.10), is equal to:

$$\nabla \Phi_t(q) = \nabla G_t * m - \int_0^t \nabla G_{t-s} * \nabla \cdot (b(s, \cdot)q(s, \cdot)) \, ds. \quad (2.4.11)$$

We can now estimate  $\|\Phi(q)\|$ . Using the same arguments as before, we have that:

$$\|G * m\| = \sup_{t \in [0, \theta]} \|G_t * m\|_{L^1} + \sup_{t \in [0, \theta]} \sqrt{t} \sum_{i=1}^d \|\partial_{x_i} G_t * m\|_{L^1} \leq 1 + d\sqrt{\frac{2}{\pi}} < +\infty,$$

and that:

$$\begin{aligned}
& \|\Phi_t(q)\|_{L^1} + \sqrt{t}\|\nabla\Phi_t(q)\|_{L^1} \\
& \leq \left(1 + d\sqrt{\frac{2}{\pi}}\right) + d\sqrt{\frac{2}{\pi}} \int_0^t \frac{1}{\sqrt{t-s}} \left\{ B\|q(s, \cdot)\|_{L^1} + \sqrt{t} \left( \check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \|q(s, \cdot)\|_{W^{1,1}} ds \right. \right. \\
& \qquad \qquad \qquad \left. \left. + B\|\nabla q(s, \cdot)\|_{L^1} \right) \right\} \\
& \leq \left(1 + d\sqrt{\frac{2}{\pi}}\right) + d\sqrt{\frac{2}{\pi}} \int_0^t \frac{1}{\sqrt{t-s}} \left\{ \left( B + \sqrt{T}\check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right) \|q(s, \cdot)\|_{L^1} \right. \\
& \qquad \qquad \qquad \left. + \sqrt{\theta} \left( B + \check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right) \|\nabla q(s, \cdot)\|_{L^1} \right\} ds \\
& \leq \left(1 + d\sqrt{\frac{2}{\pi}}\right) + \sqrt{\theta} d\sqrt{\frac{2}{\pi}} \left\{ 2 \left( B + \sqrt{T}\check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right) \sup_{u \in (0, \theta]} \|q(u, \cdot)\|_{L^1} \right. \\
& \qquad \qquad \qquad \left. + \pi \left( B + \check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right) \sup_{u \in (0, \theta]} \sqrt{u} \|\nabla q(u, \cdot)\|_{L^1} \right\} \\
& = \left(1 + d\sqrt{\frac{2}{\pi}}\right) + \sqrt{\theta} d\sqrt{\frac{2}{\pi}} \left\{ B \left( 2 \sup_{u \in (0, \theta]} \|q(u, \cdot)\|_{L^1} + \pi \sup_{u \in (0, \theta]} \sqrt{u} \|\nabla q(u, \cdot)\|_{L^1} \right) \right. \\
& \qquad \qquad \qquad \left. + \check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \left( 2\sqrt{T} \sup_{u \in (0, \theta]} \|q(u, \cdot)\|_{L^1} + \pi \sup_{u \in (0, \theta]} \sqrt{u} \|\nabla q(u, \cdot)\|_{L^1} \right) \right\} \\
& \leq \left(1 + d\sqrt{\frac{2}{\pi}}\right) + d\sqrt{2\pi\theta} \left( B + \max(1, \sqrt{T})\check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right) \left( \sup_{u \in (0, \theta]} \|q(u, \cdot)\|_{L^1} \right. \\
& \qquad \qquad \qquad \left. + \sup_{u \in (0, \theta]} \sqrt{u} \|\nabla q(u, \cdot)\|_{L^1} \right).
\end{aligned}$$

Hence,  $\|\Phi(q)\| \leq \left(1 + d\sqrt{\frac{2}{\pi}}\right) + d\sqrt{2\pi\theta} \left( B + \max(1, \sqrt{T})\check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right) \|q\|$  and  $\|\Phi(q)\| < +\infty$ . Now, let  $0 < r \leq s \leq \theta$ , adapting the proof of Lemma 2.3.1 and using Inequality (2.4.7) from Lemma 2.4.6, we obtain that:

$$\begin{aligned}
& \|\nabla\Phi_s(q) - \nabla\Phi_r(q)\|_{L^1} \\
& \leq d(2d+3)\sqrt{\frac{2}{\pi}} \left( \frac{1}{\sqrt{r}} - \frac{1}{\sqrt{s}} \right) + 4d(d+2)\sqrt{\frac{2}{\pi}} \left( B + \check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right) \sup_{u \in [r, s]} \|q(u, \cdot)\|_{W^{1,1}} \sqrt{s-r}.
\end{aligned}$$

Since  $\|\Phi(q)\| = \sup_{u \in (0, \theta]} \|q(u, \cdot)\|_{L^1} + \sup_{u \in (0, \theta]} \sqrt{u} \|\nabla q(u, \cdot)\|_{L^1}$  and using Inequality (2.4.5), we can conclude that  $t \mapsto \Phi_t(q)$  is continuous on  $(0, \theta]$  with values in  $W^{1,1}(\mathbb{R}^d)$ . Hence,  $\Phi(q) \in \tilde{\mathcal{C}}((0, \theta], W^{1,1}(\mathbb{R}^d))$ . Now, let  $q, \tilde{q} \in \tilde{\mathcal{C}}((0, \theta], W^{1,1}(\mathbb{R}^d))$ , we obtain, with the same reasoning above, that

$$\|\Phi(q) - \Phi(\tilde{q})\| \leq d\sqrt{2\pi\theta} \left( B + \max(1, \sqrt{T})\check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right) \|q - \tilde{q}\|.$$

If:

$$\theta < \frac{1}{2\pi d^2 \left( B + \max(1, \sqrt{T})\check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right)^2} =: 2\theta_0$$

then the map  $\Phi$  is a contraction on the space  $\tilde{\mathcal{C}}((0, \theta_0], W^{1,1}(\mathbb{R}^d))$ . Using Picard's Theorem, the map  $\Phi$  admits then a unique fixed-point on the space  $\tilde{\mathcal{C}}((0, \theta_0], W^{1,1}(\mathbb{R}^d))$ .

• Now, we assume that  $m$  admits a density w.r.t. the Lebesgue measure in  $W^{1,1}(\mathbb{R}^d)$ . Let us prove that the map  $\Phi$  admits a unique-fixed point in  $\mathcal{C}([0, \theta], W^{1,1}(\mathbb{R}^d))$  for all  $\theta \in (0, T]$ :

Let  $t \in (0, \theta]$  and  $q \in \mathcal{C}([0, \theta], W^{1,1}(\mathbb{R}^d))$  for all  $\theta \in (0, T]$ . Using Equation (2.4.11) and the fact that  $m$  admits a density w.r.t. the Lebesgue measure in  $W^{1,1}(\mathbb{R}^d)$ , we obtain:

$$\nabla \Phi_t(q) = G_t * \nabla m - \int_0^t \nabla G_{t-s} * \nabla \cdot (b(s, \cdot) q(s, \cdot)) ds.$$

Using the estimate (A.1.3) from Lemma A.1.2 and Inequality (2.4.7) from Lemma 2.4.6, we have:

$$\begin{aligned} \|\Phi_t(q)\|_{W^{1,1}} &\leq \|G_t * m\|_{W^{1,1}} + d \sqrt{\frac{2}{\pi}} \int_0^t \frac{1}{\sqrt{t-s}} \left( B \|q(s, \cdot)\|_{L^1} + \left\| \nabla \cdot (b(s, \cdot) q(s, \cdot)) \right\|_{L^1} \right) ds \\ &\leq \left( 1 + \sum_{i=1}^d \|\partial_{x_i} m\|_{L^1} \right) + d \sqrt{\frac{2}{\pi}} \int_0^t \frac{1}{\sqrt{t-s}} \left( B \|q(s, \cdot)\|_{L^1} + \check{C} \|\nabla \cdot b(s, \cdot)\|_{L^\rho} \|q(s, \cdot)\|_{W^{1,1}} + B \|\nabla q(s, \cdot)\|_{L^1} \right) ds \\ &\leq \left( 1 + \sum_{i=1}^d \|\partial_{x_i} m\|_{L^1} \right) + d \sqrt{\frac{2}{\pi}} \int_0^t \frac{1}{\sqrt{t-s}} \left( B \|q(s, \cdot)\|_{L^1} + \check{C} \|\nabla \cdot b(s, \cdot)\|_{L^\rho} \|q(s, \cdot)\|_{W^{1,1}} + B \|\nabla q(s, \cdot)\|_{L^1} \right) ds \\ &\leq \left( 1 + \sum_{i=1}^d \|\partial_{x_i} m\|_{L^1} \right) + d \sqrt{\frac{2}{\pi}} \left( B + \check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right) \int_0^t \frac{1}{\sqrt{t-s}} \|q(s, \cdot)\|_{W^{1,1}} ds \\ &\leq \left( 1 + \sum_{i=1}^d \|\partial_{x_i} m\|_{L^1} \right) + 2d \sqrt{\frac{2t}{\pi}} \left( B + \check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right) \sup_{u \in [0, \theta]} \|q(u, \cdot)\|_{W^{1,1}}. \end{aligned}$$

Hence,  $\sup_{t \in [0, \theta]} \|\Phi_t(q)\|_{W^{1,1}} \leq \left( 1 + \sum_{i=1}^d \|\partial_{x_i} m\|_{L^1} \right) + 2d \sqrt{\frac{2\theta}{\pi}} \left( B + \check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right) \sup_{t \in [0, \theta]} \|q(t, \cdot)\|_{W^{1,1}}$  is finite. Now, concerning the continuity of  $t \mapsto \Phi_t(q)$  on  $[0, \theta]$  with values in  $W^{1,1}(\mathbb{R}^d)$ , we already proved above the continuity on  $(0, \theta]$ . As for the continuity at  $t = 0$ , we denote by  $\tau_y w$  the translation of  $w \in L^1(\mathbb{R}^d)$  by  $y \in \mathbb{R}^d$  defined by  $\tau_y w(x) = w(x - y)$  for  $x \in \mathbb{R}^d$ . We have:

$$\begin{aligned} \|G_t * w - w\|_{L^1} &= \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} G_t(y) w(x - y) dy - w(x) \right| dx = \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} G_1(y) w(x - \sqrt{t}y) dy - w(x) \right| dx \\ &\leq \int_{\mathbb{R}^d} G_1(y) \left\| \tau_{\sqrt{t}y} w - w \right\|_{L^1} dy. \end{aligned}$$

Using Lemma IV.4 [10], we have that  $\left\| \tau_{\sqrt{t}y} w - w \right\|_{L^1} \rightarrow 0$  when  $t \rightarrow 0$  and since  $\left\| \tau_{\sqrt{t}y} w - w \right\|_{L^1} \leq 2 \|w\|_{L^1}$ , by dominated convergence, we obtain that  $G_t * w \rightarrow w$  when  $t \rightarrow 0$  in  $L^1(\mathbb{R}^d)$ . We replace  $w$  by  $m$  and  $\nabla m$  since  $m$  admits a density in  $W^{1,1}(\mathbb{R}^d)$  and conclude that  $\|G_t * m - m\|_{W^{1,1}} \rightarrow 0$  when  $t \rightarrow 0$ . Moreover, we have that:

$$\left\| \int_0^t G_{t-s} * \nabla \cdot (b(s, \cdot) q(s, \cdot)) ds \right\|_{W^{1,1}} \leq \left( B + \check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right) \sup_{u \in [0, \theta]} \|q(u, \cdot)\|_{W^{1,1}} \left( t + d \sqrt{\frac{2}{\pi}} \sqrt{t} \right)$$

which converges to 0 when  $t \rightarrow 0$ . Hence,  $\|\Phi_t(q) - m\|_{W^{1,1}}$  converges to 0 when  $t \rightarrow 0$  and  $\Phi \in \mathcal{C}([0, \theta], W^{1,1}(\mathbb{R}^d))$ .

Now, let  $q, \tilde{q} \in \mathcal{C}([0, \theta], W^{1,1}(\mathbb{R}^d))$ . With the same reasoning above, we obtain that:

$$\|\Phi_t(q) - \Phi_t(\tilde{q})\|_{W^{1,1}} \leq d \sqrt{\frac{2}{\pi}} \left( B + \check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right) \int_0^t \frac{1}{\sqrt{t-s}} \|q(s, \cdot) - \tilde{q}(s, \cdot)\|_{W^{1,1}} ds. \quad (2.4.12)$$

As done in the proof of Lemma 2.4.4, for  $n \in \mathbb{N}^*$ , we iterate Inequality (2.4.12)  $2n$ -times and deduce that for  $n$  big enough,  $\Phi^{2n}$  is a contraction on  $\mathcal{C}([0, \theta], W^{1,1}(\mathbb{R}^d))$ . We conclude, through Picard's Theorem, the existence of a unique fixed-point of the map  $\Phi$  on  $\mathcal{C}([0, \theta], W^{1,1}(\mathbb{R}^d))$ .  $\square$

We are now ready to prove Proposition 2.4.1.

*Proof.* We are going to suppose that  $\sup_{t \in [0, T]} \|\nabla \cdot b(t, \cdot)\|_{L^\rho} < +\infty$  for some  $\rho \in [d, +\infty]$ . When  $d = 1$ ,  $\sup_{t \in [0, T]} \|\partial_x b(t, \cdot)\|_{TV} < +\infty$ , the proof is analogous and the estimations remain valid when replacing  $\|\nabla \cdot b(t, \cdot)\|_{L^\rho}$  by  $\|\partial_x b(t, \cdot)\|_{TV}$ .

The proof is an immediate consequence of Theorem 2.4.3 and Lemma 2.4.6. Indeed, assuming the regularity on  $b$  w.r.t. the space variables, we have from Theorem 2.4.3 that  $(p(t, \cdot))_{t \in (0, T]} \in \tilde{\mathcal{C}}((0, T], W^{1,1}(\mathbb{R}^d))$ . We can then apply Lemma 2.4.6 to obtain, for  $t \in (0, T]$ , that:

$$\begin{aligned} & \sqrt{t} \left\| \nabla \cdot (b(t, \cdot) p(t, \cdot)) \right\|_{TV} \\ & \leq \left( B + \check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right) \sqrt{t} \|\nabla p(t, \cdot)\|_{L^1} + \check{C} \sqrt{T} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \|p(t, \cdot)\|_{L^1} \\ & \leq \max \left( B + \check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho}, \check{C} \sqrt{T} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right) \|p\| < +\infty. \end{aligned}$$

The conclusion holds with  $M = \max \left( B + \check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho}, \check{C} \sqrt{T} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right) \|p\|$ . Since  $p \in \tilde{\mathcal{C}}((0, T], W^{1,1}(\mathbb{R}^d))$ , using Lemma 2.4.6, we have as in Equality (2.4.10) that:

$$\forall t \in (0, T], \quad p(t, \cdot) = G_t * m - \int_0^t G_{t-s} * \nabla \cdot (b(s, \cdot) p(s, \cdot)) ds.$$

Moreover, when  $m$  admits a density w.r.t. the Lebesgue measure in  $W^{1,1}(\mathbb{R}^d)$ , according to Theorem 2.4.3,  $(p(t, \cdot))_{t \in [0, T]} \in \mathcal{C}([0, T], W^{1,1}(\mathbb{R}^d))$ . Therefore, using once again Lemma 2.4.6, we obtain for  $t \in [0, T]$  that:

$$\left\| \nabla \cdot (b(t, \cdot) p(t, \cdot)) \right\|_{TV} \leq \left( B + \check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right) \sup_{u \in [0, T]} \|p(u, \cdot)\|_{W^{1,1}} < +\infty.$$

The conclusion holds with  $\tilde{M} = \left( B + \check{C} \sup_{u \in [0, T]} \|\nabla \cdot b(u, \cdot)\|_{L^\rho} \right) \sup_{u \in [0, T]} \|p(u, \cdot)\|_{W^{1,1}}$ .  $\square$

## 2.4.2 Proof of Theorem 2.2.3

We first use Inequality (2.4.1) from Proposition 2.4.1 to obtain a stronger regularity of  $p(t, \cdot)$  with respect to the time variable.

**Lemma 2.4.7.** *Assume Inequality (2.4.1). We have:*

$$\exists \tilde{Q} < +\infty, \forall 0 < r \leq s \leq T, \quad \|p(s, \cdot) - p(r, \cdot)\|_{L^1} \leq \tilde{Q} \left( \ln(s/r) + \frac{s-r}{2\sqrt{r}} \ln \left( \frac{4s}{s-r} \right) + (\sqrt{s} - \sqrt{r}) \right).$$

*Proof.* We will adapt the proof of Lemma 2.3.1. Let  $0 < r \leq s \leq T$ . Using Equality (2.4.2) from Proposition 2.4.1, we have that:

$$p(s, \cdot) - p(r, \cdot) = (G_s - G_r) * m - \int_0^s G_{s-u} * \nabla \cdot (b(u, \cdot) p(u, \cdot)) du + \int_0^r G_{r-u} * \nabla \cdot (b(u, \cdot) p(u, \cdot)) du. \quad (2.4.13)$$

Therefore, using the estimates (A.1.2) and (A.1.3) from Lemma A.1.2, the fact that  $\ln(1+x) \leq x$ ,  $\forall x > 0$  and Lemma A.1.5, we obtain:

$$\begin{aligned}
 & \|p(s, \cdot) - p(r, \cdot)\|_{L^1} \\
 & \leq \|(G_s - G_r) * m\|_{L^1} + \left\| \int_0^r (G_{s-u} - G_{r-u}) * \nabla \cdot (b(u, \cdot)p(u, \cdot)) du \right\|_{L^1} + \left\| \int_r^s G_{s-u} * \nabla \cdot (b(u, \cdot)p(u, \cdot)) du \right\|_{L^1} \\
 & \leq \int_r^s \|\partial_u G_u\|_{L^1} du + \int_0^s \int_{r-u}^{s-u} \|\partial_\theta G_\theta\|_{L^1} \|\nabla \cdot (b(u, \cdot)p(u, \cdot))\|_{TV} d\theta du + M \int_r^s \frac{du}{\sqrt{u}} \\
 & \leq d \ln(s/r) + dM \int_0^r \int_{r-u}^{s-u} \frac{d\theta}{\theta \sqrt{u}} du + 2M(\sqrt{s} - \sqrt{r}) \\
 & = d \ln(s/r) + 2dM \left( \frac{s-r}{\sqrt{s} + \sqrt{r}} \ln \left( \frac{(\sqrt{s} + \sqrt{r})^2}{s-r} \right) + 2\sqrt{r} \ln \left( 1 + \frac{\sqrt{s} - \sqrt{r}}{2\sqrt{r}} \right) \right) + 2M(\sqrt{s} - \sqrt{r}) \\
 & \leq d \ln(s/r) + dM \frac{s-r}{\sqrt{r}} \ln \left( \frac{4s}{s-r} \right) + 2M(1+d)(\sqrt{s} - \sqrt{r}).
 \end{aligned}$$

The conclusion holds with  $\tilde{Q} = \max(d, 2M(1+d))$ . □

We are now ready to prove Theorem 2.2.3. Once again, using Equality (2.2.2) and Proposition 2.2.10, to prove the theorem amounts to prove that:

$$\exists \tilde{C} < +\infty, \forall h \in (0, T], \forall k \in \left[1, \left\lfloor \frac{T}{h} \right\rfloor\right], \quad \|p(kh, \cdot) - p^h(kh, \cdot)\|_{L^1} \leq \frac{\tilde{C}}{\sqrt{kh}} (1 + \ln(k)) h.$$

For  $h \in (0, T]$ ,  $k \in \left[1, \left\lfloor \frac{T}{h} \right\rfloor\right]$ , we recall Inequality (2.3.4):

$$\begin{aligned}
 \|p(kh, \cdot) - p^h(kh, \cdot)\|_{L^1} & \leq \|V_1^h(k, \cdot)\|_{L^1} + \|V_2^h(k, \cdot)\|_{L^1} + \sqrt{\frac{2}{\pi}} dB \sum_{j=1}^{k-1} \frac{h}{\sqrt{kh} - jh} \|p(jh, \cdot) - p^h(jh, \cdot)\|_{L^1} \\
 & \quad + 2dB^2 \left(1 + \frac{d-1}{\pi}\right) \left(\frac{1}{2} + \ln(k)\right) h + \sqrt{\frac{2}{\pi}} dB \frac{3h}{\sqrt{kh}}.
 \end{aligned}$$

Let us estimate  $\|V_1^h(k, \cdot)\|_{L^1}$  and  $\|V_2^h(k, \cdot)\|_{L^1}$  for  $k \geq 2$  by taking advantage of the additional regularity of  $b$  and using Equality (2.4.2) from Proposition 2.4.1:

- We use Lemma 2.4.6 and the additional regularity of  $b$  to transfer the gradient from  $G$  to  $bp$  and rewrite  $V_1^h(k, \cdot)$  as:

$$\begin{aligned}
 V_1^h(k, \cdot) & = \int_h^{kh} \left( G_{kh-s} - G_{kh-\tau_s^h} \right) * \nabla \cdot (b(s, \cdot)p(s, \cdot)) ds \\
 & = \int_h^{(k-1)h} \left( \int_{kh-s}^{kh-\tau_s^h} \partial_u G_u du \right) * \nabla \cdot (b(s, \cdot)p(s, \cdot)) ds + \int_{(k-1)h}^{kh} \left( G_{kh-s} - G_{kh-\tau_s^h} \right) * \nabla \cdot (b(s, \cdot)p(s, \cdot)) ds.
 \end{aligned} \tag{2.4.14}$$

Therefore, using Estimate (A.1.2) from Lemma A.1.2, Inequality (2.4.1) from Proposition 2.4.1, Inequality (2.3.6) and Lemma A.1.4, we obtain:

$$\begin{aligned}
 \|V_1^h(k, \cdot)\|_{L^1} & \leq dM \int_h^{(k-1)h} \frac{1}{\sqrt{s}} \ln \left( 1 + \frac{s - \tau_s^h}{kh - s} \right) ds + 2M \int_{(k-1)h}^{kh} \frac{ds}{\sqrt{s}} \\
 & \leq dM \int_h^{(k-1)h} \frac{h}{(kh-s)\sqrt{s}} ds + 4M \left( \sqrt{kh} - \sqrt{(k-1)h} \right) \\
 & \leq dM \frac{h}{\sqrt{kh}} \ln(4k) + 4M \frac{h}{\sqrt{kh} + \sqrt{(k-1)h}} \\
 & \leq \frac{2M(2 + d \ln(2))}{\sqrt{kh}} (1 + \ln(k)) h.
 \end{aligned}$$

• Using Lemma 2.4.7 and the estimate (A.1.3) from Lemma A.1.2, we have:

$$\|V_2^h(k, \cdot)\|_{L^1} \leq \sqrt{\frac{2}{\pi}} dB \tilde{Q} \left\{ \int_h^{kh} \frac{\ln(s/\tau_s^h)}{\sqrt{kh} - \tau_s^h} ds + \int_h^{kh} \frac{s - \tau_s^h}{2\sqrt{\tau_s^h} \sqrt{kh} - \tau_s^h} \ln\left(\frac{4s}{s - \tau_s^h}\right) ds + \int_h^{kh} \frac{\sqrt{s} - \sqrt{\tau_s^h}}{\sqrt{kh} - \tau_s^h} ds \right\}.$$

We use Inequality (2.3.6), the fact that  $\sup_{0 < x \leq h} \left( x \ln\left(\frac{4kh}{x}\right) \right)$  is attained for  $x = h$  since  $h \leq kh$  and Lemma A.1.3 to obtain:

$$\begin{aligned} \|V_2^h(k, \cdot)\|_{L^1} &\leq \sqrt{\frac{2}{\pi}} dB \tilde{Q} \left\{ \frac{2h}{\sqrt{kh}} \ln(4k) + h \ln(4k) \int_h^{kh} \frac{ds}{2\sqrt{\tau_s^h} \sqrt{kh} - \tau_s^h} + \int_h^{kh} \frac{s - \tau_s^h}{(\sqrt{s} + \sqrt{\tau_s^h}) \sqrt{kh} - \tau_s^h} ds \right\} \\ &\leq \sqrt{\frac{2}{\pi}} dB \tilde{Q} \left\{ \frac{2h}{\sqrt{kh}} \ln(4k) + \frac{h}{2} \ln(4k) \sum_{j=1}^{k-1} \frac{1}{\sqrt{j} \sqrt{k-j}} + \frac{h}{2} \sum_{j=1}^{k-1} \frac{1}{\sqrt{j} \sqrt{k-j}} \right\} \\ &\leq \sqrt{\frac{2}{\pi}} dB \tilde{Q} \left\{ \frac{2}{\sqrt{kh}} \ln(4k) + \frac{\pi}{2} (1 + \ln(4k)) \right\} h. \end{aligned}$$

Therefore,

$$\|p(kh, \cdot) - p^h(kh, \cdot)\|_{L^1} \leq \frac{\tilde{L}}{\sqrt{kh}} (1 + \ln(k)) h + \sqrt{\frac{2}{\pi}} dB \sum_{j=1}^{k-1} \frac{h}{\sqrt{kh} - jh} \|p(jh, \cdot) - p^h(jh, \cdot)\|_{L^1}$$

where  $\tilde{L} = 2M(2 + d \ln(2)) + dB \left\{ \sqrt{\frac{2}{\pi}} (3 + 4 \ln(2) \tilde{Q}) + \left( 2B \left( 1 + \frac{d-1}{\pi} \right) + \sqrt{\frac{\pi}{2}} (1 + 2 \ln(2)) \tilde{Q} \right) \sqrt{T} \right\}$ .

Iterating this inequality, using the fact that for  $j \leq k$ ,  $\ln(j) \leq \ln(k)$  and using Lemma A.1.3, we obtain:

$$\begin{aligned} &\|p(kh, \cdot) - p^h(kh, \cdot)\|_{L^1} \\ &\leq \frac{\tilde{L}}{\sqrt{kh}} (1 + \ln(k)) h + \sqrt{\frac{2}{\pi}} dB \sum_{j=1}^{k-1} \frac{h}{\sqrt{kh} - jh} \left( \frac{\tilde{L}}{\sqrt{jh}} (1 + \ln(j)) h \right. \\ &\quad \left. + \sqrt{\frac{2}{\pi}} dB \sum_{l=1}^{j-1} \frac{h}{\sqrt{jh} - lh} \|p(lh, \cdot) - p^h(lh, \cdot)\|_{L^1} \right) \\ &\leq \tilde{L} \left( \frac{1}{\sqrt{kh}} + \sqrt{\frac{2}{\pi}} dB \sum_{j=1}^{k-1} \frac{h}{\sqrt{kh} - jh \sqrt{jh}} \right) (1 + \ln(k)) h + 2d^2 B^2 h \sum_{j=1}^{k-1} \|p(jh, \cdot) - p^h(jh, \cdot)\|_{L^1} \\ &\leq \tilde{L} \sqrt{h} (1 + dB \sqrt{2\pi T}) \frac{1 + \ln(k)}{\sqrt{k}} + 2d^2 B^2 h \sum_{j=1}^{k-1} \|p(jh, \cdot) - p^h(jh, \cdot)\|_{L^1}. \end{aligned}$$

We apply Lemma A.1.6 and obtain that:

$$\begin{aligned} &\|p(kh, \cdot) - p^h(kh, \cdot)\|_{L^1} \\ &\leq \tilde{L} \sqrt{h} (1 + dB \sqrt{2\pi T}) \frac{1 + \ln(k)}{\sqrt{k}} + 2d^2 B^2 \tilde{L} \sqrt{h} (1 + dB \sqrt{2\pi T}) h \sum_{j=1}^{k-1} \frac{1 + \ln(j)}{\sqrt{j}} \exp\left(2d^2 B^2 (kh - (j+1)h)\right). \end{aligned}$$

Now, using once again the fact that for  $j \leq k$ ,  $\ln(j) \leq \ln(k)$ , we have:

$$\begin{aligned} \sqrt{h} \sum_{j=1}^{k-1} \frac{1 + \ln(j)}{\sqrt{j}} \exp\left(2d^2 B^2 (kh - (j+1)h)\right) &\leq (1 + \ln(k)) \exp\left(2d^2 B^2 T\right) \int_h^{kh} \frac{ds}{\sqrt{s}} \\ &= 2(1 + \ln(k)) \exp\left(2d^2 B^2 T\right) \sqrt{T}. \end{aligned}$$

Therefore, we deduce that:

$$\forall h \in (0, T], \forall k \in \left[1, \left\lfloor \frac{T}{h} \right\rfloor\right], \quad \|p(kh, \cdot) - p^h(kh, \cdot)\|_{L^1} \leq \frac{\tilde{C}}{\sqrt{kh}} \left(1 + \ln(k)\right) h$$

where  $\tilde{C} = \tilde{L} \left(1 + 4d^2 B^2 T \left(1 + dB\sqrt{2\pi T}\right) \exp\left(2d^2 B^2 T\right)\right)$ .

### 2.4.3 Proof of Proposition 2.2.5

We first use Inequality (2.4.3) from Proposition 2.4.1 to obtain a stronger regularity of  $p(t, \cdot)$  with respect to the time variable.

**Lemma 2.4.8.** *Assume  $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  is measurable and bounded by  $B < +\infty$  such that  $\sup_{t \in [0, T]} \|\nabla \cdot b(t, \cdot)\|_{L^\rho} < +\infty$  for some  $\rho \in [d, +\infty]$  or that for  $d = 1$ ,  $\sup_{t \in [0, T]} \|\partial_x b(t, \cdot)\|_{TV} < +\infty$ ; where  $\nabla \cdot b(t, \cdot)$  and  $\partial_x b(t, \cdot)$  are respectively the spatial divergence and the spatial derivative of  $b$  in the sense of distributions. Moreover, assume that  $m$  admits a density w.r.t. the Lebesgue measure that belongs to  $W^{1,1}(\mathbb{R}^d)$ . We have:*

$$\exists \hat{Q} < +\infty, \forall 0 < r \leq s \leq T, \quad \|p(s, \cdot) - p(r, \cdot)\|_{L^1} \leq \hat{Q} \left( (\sqrt{s} - \sqrt{r}) + (s-r) \ln \left( \frac{s}{s-r} \right) + r \ln(s/r) + (s-r) \right).$$

*Proof.* We will adapt, once again, the proof of Lemma 2.3.1. Let  $0 < r \leq s \leq T$ , using Equality (2.4.13), the estimates (A.1.1), (A.1.2) and (A.1.3) from Lemma A.1.2, Inequality (2.4.3) from Proposition 2.4.1 and Lemma A.1.5, we obtain:

$$\begin{aligned} & \|p(s, \cdot) - p(r, \cdot)\|_{L^1} \\ & \leq \|(G_s - G_r) * m\|_{L^1} + \left\| \int_0^r (G_{s-u} - G_{r-u}) * \nabla \cdot (b(u, \cdot) p(u, \cdot)) du \right\|_{L^1} + \left\| \int_r^s G_{s-u} * \nabla \cdot (b(u, \cdot) p(u, \cdot)) du \right\|_{L^1} \\ & \leq \left\| \int_r^s (\partial_u G_u * m) du \right\|_{L^1} + \tilde{M} \int_0^s \int_{r-u}^{s-u} \|\partial_\theta G_\theta\|_{L^1} d\theta du + \tilde{M}(s-r) \\ & \leq \frac{1}{2} \sum_{i=1}^d \|\partial_{x_i} m\|_{L^1} \int_r^s \|\partial_{x_i} G_u\|_{L^1} du + d\tilde{M} \int_0^r \ln \left( \frac{s-u}{r-u} \right) du + \tilde{M}(s-r) \\ & \leq 2\sqrt{\frac{2}{\pi}} \sum_{i=1}^d \|\partial_{x_i} m\|_{L^1} (\sqrt{s} - \sqrt{r}) + d\tilde{M} \left( (s-r) \ln \left( \frac{s}{s-r} \right) + r \ln(s/r) \right) + \tilde{M}(s-r). \end{aligned}$$

The conclusion holds with  $\hat{Q} = \max \left( 2\sqrt{\frac{2}{\pi}} \sum_{i=1}^d \|\partial_{x_i} m\|_{L^1}, d\tilde{M} \right)$ . □

We are now ready to prove Proposition 2.2.5. Once more, using Equality (2.2.2) and Proposition 2.2.10, to prove the theorem amounts to prove that:

$$\exists \hat{C} < +\infty, \forall h \in (0, T], \forall k \in \left[1, \left\lfloor \frac{T}{h} \right\rfloor\right], \quad \|p(kh, \cdot) - p^h(kh, \cdot)\|_{L^1} \leq \hat{C} \left(1 + \ln(k)\right) h.$$

For  $h \in (0, T]$ ,  $k \in \left[1, \left\lfloor \frac{T}{h} \right\rfloor\right]$ , we recall Inequality (2.3.3):

$$\begin{aligned} \|p(kh, \cdot) - p^h(kh, \cdot)\|_{L^1} & \leq \|V_1^h(k, \cdot)\|_{L^1} + \|V_2^h(k, \cdot)\|_{L^1} + \sqrt{\frac{2}{\pi}} dB \sum_{j=1}^{k-1} \frac{h}{\sqrt{kh} - jh} \|p(jh, \cdot) - p^h(jh, \cdot)\|_{L^1} \\ & \quad + 2dB^2 \left(1 + \frac{d-1}{\pi}\right) \left(\frac{1}{2} + \ln(k)\right) h + \left\| \int_0^h \left( \nabla G_{kh-s} * (b(s, \cdot) p(s, \cdot)) - \nabla G_{kh} * (b(s, \cdot) m) \right) ds \right\|_{L^1}. \end{aligned}$$

Concerning the last term of the right-hand side of this previous inequality, we use Lemma 2.4.6 and the additional regularity of  $b$  to transfer the gradient from  $G$  to  $bp$ ; and using Inequality (2.4.3) from Proposition 2.4.1, we obtain that:

$$\begin{aligned} & \left\| \int_0^h \left( G_{kh-s} * \nabla \cdot (b(s, \cdot) p(s, \cdot)) - G_{kh} * \nabla \cdot (b(s, \cdot) m) \right) ds \right\|_{L^1} \\ & \leq \int_0^h \left( \left\| \nabla \cdot (b(s, \cdot) p(s, \cdot)) \right\|_{\text{TV}} + \left\| \nabla \cdot (b(s, \cdot) m) \right\|_{\text{TV}} \right) ds \leq 2\tilde{M}h. \end{aligned}$$

Let us estimate  $\|V_1^h(k, \cdot)\|_{L^1}$  and  $\|V_2^h(k, \cdot)\|_{L^1}$  for  $k \geq 2$  by taking advantage of the additional regularity of  $b$  and  $m$ , and using Equality (2.4.2) from Proposition 2.4.1:

- We recall Equality (2.4.14):

$$V_1^h(k, \cdot) = \int_h^{(k-1)h} \left( \int_{kh-s}^{kh-\tau_s^h} \partial_u G_u du \right) * \nabla \cdot (b(s, \cdot) p(s, \cdot)) ds + \int_{(k-1)h}^{kh} \left( G_{kh-s} - G_{kh-\tau_s^h} \right) * \nabla \cdot (b(s, \cdot) p(s, \cdot)) ds.$$

Therefore, using the fact that  $\ln(1+x) \leq x$ ,  $\forall x > 0$  and Estimate (A.1.2) from Lemma A.1.2, we obtain:

$$\begin{aligned} \|V_1^h(k, \cdot)\|_{L^1} & \leq d\tilde{M} \int_h^{(k-1)h} \ln \left( 1 + \frac{s - \tau_s^h}{kh - s} \right) ds + 2\tilde{M}h \\ & \leq d\tilde{M} \int_h^{(k-1)h} \frac{h}{kh - s} ds + 2\tilde{M}h = \tilde{M} (d \ln(k-1) + 2) h \\ & \leq \tilde{M} (2 + d \ln(k)) h. \end{aligned}$$

- Using Lemma 2.4.8, the estimate (A.1.3) from Lemma A.1.2 and the fact that  $\ln(1+x) \leq x$ ,  $\forall x > 0$ , we have:

$$\begin{aligned} \|V_2^h(k, \cdot)\|_{L^1} & \leq \sqrt{\frac{2}{\pi}} dB \hat{Q} \left\{ \int_h^{kh} \frac{\sqrt{s} - \sqrt{\tau_s^h}}{\sqrt{kh - \tau_s^h}} ds + \int_h^{kh} \frac{s - \tau_s^h}{\sqrt{kh - \tau_s^h}} \ln \left( \frac{s}{s - \tau_s^h} \right) ds \right. \\ & \quad \left. + \int_h^{kh} \frac{\tau_s^h}{\sqrt{kh - \tau_s^h}} \ln \left( 1 + \frac{s - \tau_s^h}{\tau_s^h} \right) ds + \int_h^{kh} \frac{s - \tau_s^h}{\sqrt{kh - \tau_s^h}} ds \right\} \\ & \leq \sqrt{\frac{2}{\pi}} dB \hat{Q} \left\{ \int_h^{kh} \frac{h}{\sqrt{kh - s} \sqrt{s}} ds + \int_h^{kh} \frac{s - \tau_s^h}{\sqrt{kh - s}} \ln \left( \frac{kh}{s - \tau_s^h} \right) ds + 2 \int_h^{kh} \frac{h}{\sqrt{kh - s}} ds \right\}. \end{aligned}$$

The function  $x \mapsto x \ln(kh/x)$  is increasing on the interval  $(0, kh/e]$ , attains its maximum at  $x = kh/e$  and non-increasing on the interval  $[kh/e, +\infty)$ . Therefore, we get that:

$$(s - \tau_s^h) \ln \left( \frac{kh}{s - \tau_s^h} \right) \leq \left( h \ln(k) \mathbf{1}_{\{h \leq \frac{kh}{e}\}} + \frac{kh}{e} \mathbf{1}_{\{h > \frac{kh}{e}\}} \right) \leq (1 + \ln(k)) h.$$

We then deduce that:

$$\|V_2^h(k, \cdot)\|_{L^1} \leq \sqrt{\frac{2}{\pi}} dB \hat{Q} \left( \pi + 2\sqrt{T} (1 + \ln(k)) + 4\sqrt{T} \right) h.$$

Therefore,

$$\|p(kh, \cdot) - p^h(kh, \cdot)\|_{L^1} \leq \hat{L} (1 + \ln(k)) h + \sqrt{\frac{2}{\pi}} dB \sum_{j=1}^{k-1} \frac{h}{\sqrt{kh - jh}} \|p(jh, \cdot) - p^h(jh, \cdot)\|_{L^1}$$

where  $\hat{L} = \max(2, d)\tilde{M} + dB \left( 2B \left( 1 + \frac{d-1}{\pi} \right) + \sqrt{\frac{2}{\pi}} \hat{Q} (\pi + 6\sqrt{T}) \right)$ . Iterating this inequality, using



the fact that for  $j \leq k$ ,  $\ln(j) \leq \ln(k)$  and using Lemma A.1.3, we obtain:

$$\begin{aligned}
& \|p(kh, \cdot) - p^h(kh, \cdot)\|_{L^1} \\
& \leq \hat{L}(1 + \ln(k))h + \sqrt{\frac{2}{\pi}}dB \sum_{j=1}^{k-1} \frac{h}{\sqrt{kh-jh}} \left( \hat{L}(1 + \ln(j))h + \sqrt{\frac{2}{\pi}}dB \sum_{l=1}^{j-1} \frac{h}{\sqrt{jh-lh}} \|p(lh, \cdot) - p^h(lh, \cdot)\|_{L^1} \right) \\
& \leq \hat{L} \left( 1 + \sqrt{\frac{2}{\pi}}dB \int_h^{kh} \frac{ds}{\sqrt{kh-s}} \right) (1 + \ln(k))h + 2d^2B^2h \sum_{j=1}^{k-1} \|p(jh, \cdot) - p^h(jh, \cdot)\|_{L^1} \\
& \leq \hat{L} \left( 1 + 2\sqrt{\frac{2T}{\pi}}dB \right) (1 + \ln(k))h + 2d^2B^2h \sum_{j=1}^{k-1} \|p(jh, \cdot) - p^h(jh, \cdot)\|_{L^1}.
\end{aligned}$$

Finally using Lemma A.1.6 and for  $\hat{C} = \hat{L} \left( 1 + 2\sqrt{\frac{2T}{\pi}}dB \right) \left( 1 + 2d^2B^2T \exp \left( 2d^2B^2T \right) \right)$ , we conclude that:

$$\|p(kh, \cdot) - p^h(kh, \cdot)\|_{L^1} \leq \hat{C}(1 + \ln(k))h.$$

## 2.5 Numerical Experiments

In order to confirm our theoretical estimates for the convergence rate in total variation of  $\mu^h$  to its limit  $\mu$ , we study SDEs with a piecewise constant drift coefficient and additive noise, as done by Götlich, Lux and Neuenkirch in [28]. We consider the special case of one-dimensional SDEs with one drift change at zero:

$$X_t = x + W_t + \int_0^t \left( \alpha \mathbf{1}_{(-\infty, 0)}(X_s) + \beta \mathbf{1}_{[0, +\infty)}(X_s) \right) ds \quad (2.5.1)$$

where  $X_0 = x \in \mathbb{R}$  is the initial value and  $\alpha, \beta \in \mathbb{R}$ . The difference  $(\beta - \alpha)$  represents the height of the jump at the discontinuity point zero. Here, the spatial derivative of the drift in the sense of distributions is equal to  $(\beta - \alpha)\delta_0$  and the drift satisfies the reinforced hypothesis of Theorem 2.2.3.

We analyze how the initial value  $x$  affects the error and how the jump height influences the empirical rate of convergence. We also observe how the drift direction towards or away from the discontinuity point zero influences the error. When  $\alpha > 0 > \beta$ , we speak about inward pointing drift coefficient. Inversely, when  $\alpha < 0 < \beta$ , we speak about outward pointing drift coefficient.

### 2.5.1 The specific case $\alpha = -\beta = \theta > 0$

For  $\theta > 0$ , we study SDEs with inward pointing drift coefficient of the form:

$$X_t = x + W_t - \int_0^t \operatorname{sgn}(X_s) \theta ds.$$

This process is called a Brownian motion with two-valued, state-dependent drift. This example was also used by Kohatsu-Higa, Lejay and Yasuda in [59] to estimate the weak convergence rate of the Euler-Maruyama scheme. According to [53], the transition density function of the process  $(X_t)_{t \geq 0}$  starting at  $x \geq 0$  is the following:

$$p_t(x, z) = \begin{cases} \frac{1}{\sqrt{2\pi t}} \exp \left( -\frac{(x-z-\theta t)^2}{2t} \right) + \frac{\theta e^{-2\theta z}}{\sqrt{2\pi t}} \int_{x+z}^{+\infty} \exp \left( -\frac{(y-\theta t)^2}{2t} \right) dy & \text{when } z > 0, \\ \frac{e^{2\theta x}}{\sqrt{2\pi t}} \exp \left( -\frac{(x-z+\theta t)^2}{2t} \right) + \frac{\theta e^{2\theta z}}{\sqrt{2\pi t}} \int_{x-z}^{+\infty} \exp \left( -\frac{(y-\theta t)^2}{2t} \right) dy & \text{when } z \leq 0. \end{cases}$$

For  $x \leq 0$ , the transition density can be deduced from the symmetry of the Brownian motion that gives  $p_t(x, z) = p_t(-x, -z)$ .

We seek to observe the dependence of the error in total variation at terminal time  $T$ :  $\|\mu_T - \mu_T^h\|_{TV}$  on the time step  $h$  that we choose s.t.  $\frac{T}{h}$  is an integer. To do so, we estimate  $\|p(T, \cdot) - p^h(T, \cdot)\|_{L^1}$  using a kernel density estimator for  $p^h(T, \cdot)$ . We denote by  $N$  the number of random variables  $(X_T^{i,h})_{1 \leq i \leq N}$  that are i.i.d. with density  $p^h(T, \cdot)$ . The kernel density estimator of this latter is defined by:

$$p_{\epsilon, N}^h(T, x) = \frac{1}{\epsilon N} \sum_{j=1}^N K\left(\frac{x - X_T^{j,h}}{\epsilon}\right)$$

where  $K$  represents the kernel and  $\epsilon > 0$  is a smoothing parameter called the bandwidth. The kernel is a non-negative and integrable even function that ensures the required normalization of a density i.e.  $\int_{-\infty}^{+\infty} K(x) dx = 1$ . As for the smoothing parameter, its influence is critical since a very small  $\epsilon$  makes the estimator show insignificant details and a very large  $\epsilon$  causes oversmoothing and may mask some characteristics. So a compromise is needed. The optimal smoothing parameter can be chosen through a minimisation of the asymptotic mean integrated squared error. For an explicit known density, as it is the case here, we have that:

$$\epsilon = cN^{-1/5} \quad \text{with} \quad c = \frac{R(K)^{1/5}}{m_2(K)^{2/5} R(\partial_{xx}^2 p(t, \cdot))^{1/5}} \quad (2.5.2)$$

where for a given function  $g$ ,  $R(g) = \int_{\mathbb{R}} g(x)^2 dx$  and  $m_2(g) = \int_{\mathbb{R}} x^2 g(x) dx$ . We will choose, in what follows, the Epanechnikov kernel defined by:

$$K(x) = \frac{3}{4} (1 - x^2) \mathbf{1}_{\{|x| \leq 1\}}$$

which is known to be theoretically optimal in a mean square error sense with  $R(K) = 3/5$  and  $m_2(K) = 1/5$ . For  $(X_T^{(i),h})_{1 \leq i \leq N}$  denoting the increasing reordering of  $(X_T^{i,h})_{1 \leq i \leq N}$ , we make the following trapezoidal approximation:

$$\|p^h(T, \cdot) - p(T, \cdot)\|_{L^1} \simeq \sum_{i=1}^{N-1} \frac{1}{2} (X_T^{(i+1),h} - X_T^{(i),h}) \left\{ \left| p_{\epsilon, N}^h(T, X_T^{(i+1),h}) - p(T, X_T^{(i+1),h}) \right| + \left| p_{\epsilon, N}^h(T, X_T^{(i),h}) - p(T, X_T^{(i),h}) \right| \right\}.$$

We also define the precision of this estimation as half the width of the 95% confidence interval of the empirical error i.e. Precision =  $1.96 \times \sqrt{\text{Variance}/R}$  where  $R$  denotes the number of Monte-Carlo runs and Variance denotes the empirical variance over these runs of the empirical error.

### Illustration of the theoretical order of convergence in total variation

To observe the convergence rate in total variation for the case  $\theta = 1.0$  and  $x = 0.0$ , we fix the time horizon  $T = 1$  and the number  $N = 500000$  of i.i.d. samples in the kernel density estimator large enough in order to observe the effect of the time-step  $h$  on the error. The simulation is done with  $R = 20$  Monte-Carlo runs. We obtain the following results for the estimation of the error and its associated precision:

| Evolution of the total variation error w.r.t. $h$ |            |                       |                   |                   |
|---------------------------------------------------|------------|-----------------------|-------------------|-------------------|
| Time-step $h$                                     | Estimation | Precision             | Ratio of decrease | Theoretical Ratio |
| $T/4$                                             | 0.2903     | $5.48 \times 10^{-4}$ | $\times$          | $\times$          |
| $T/8$                                             | 0.1680     | $6.96 \times 10^{-4}$ | 1.73              | 1.63              |
| $T/16$                                            | 0.0956     | $4.88 \times 10^{-4}$ | 1.76              | 1.69              |
| $T/32$                                            | 0.0543     | $5.26 \times 10^{-4}$ | 1.76              | 1.73              |
| $T/64$                                            | 0.0314     | $6.53 \times 10^{-4}$ | 1.73              | 1.76              |
| $T/128$                                           | 0.0191     | $3.55 \times 10^{-4}$ | 1.64              | 1.79              |
| $T/256$                                           | 0.0133     | $2.71 \times 10^{-4}$ | 1.43              | 1.80              |
| $T/512$                                           | 0.0101     | $3.10 \times 10^{-4}$ | 1.31              | 1.82              |

- We observe that the ratio of successive estimations  $\frac{\text{Estimation}(h)}{\text{Estimation}(h/2)}$  is roughly around 1.72. But when  $h$  becomes small, the ratio decreases towards 1 (a constant error) because for so small discretizations steps, the effect of the kernel density estimation parameter  $N$  cannot be neglected unless  $N$  is extremely large.
- The last column refers to the theoretical ratios equal to  $2^{\left(\frac{1+\ln(T/h)}{1+\ln(2T/h)}\right)}$  which is the expected behaviour of the error. On the range of values  $\left\{\frac{T}{8}, \frac{T}{16}, \frac{T}{32}, \frac{T}{64}, \frac{T}{128}\right\}$ , both the empirical and the theoretical ratios are equal to 1.72 in average.
- Moreover, the order of convergence in total variation of the Euler scheme is here equal to 0.76. This order is given by the slope of the regression line, which we obtain when plotting  $\log \|p(T, \cdot) - p^h(T, \cdot)\|_{L^1}$  versus  $\log(h)$ .

### Dependence of the order of convergence on the initial value $x$ for fixed $\theta = 1$

To underline the influence of the initial value of the SDE, we start by generating plots of the explicit transition density function for various initializations  $x \in \{-1, 0, 1, 2.5, 5\}$  and different time horizons  $T \in \{1, 3, 6\}$ . We choose a fixed  $\theta = 1$ . We also plot the kernel transition density estimation for the different values of  $x$  at  $T = 1$  for  $N = 100000$  and a time-step  $h = 0.0001$ .

We first observe from Figures 2.1a and 2.1b that the kernel density estimator catches the discontinuity and reproduces well the expected distribution. We also see from Figures 2.1b, 2.1c and 2.1d that when the process starts from the discontinuity point  $x = 0.0$  or close to it  $x \in \{-1, 1\}$ , it visits the discontinuity point several times. When we increase the time horizon  $T$ , the inward pointing drift allows the process to visit the discontinuity point zero when starting far from it.

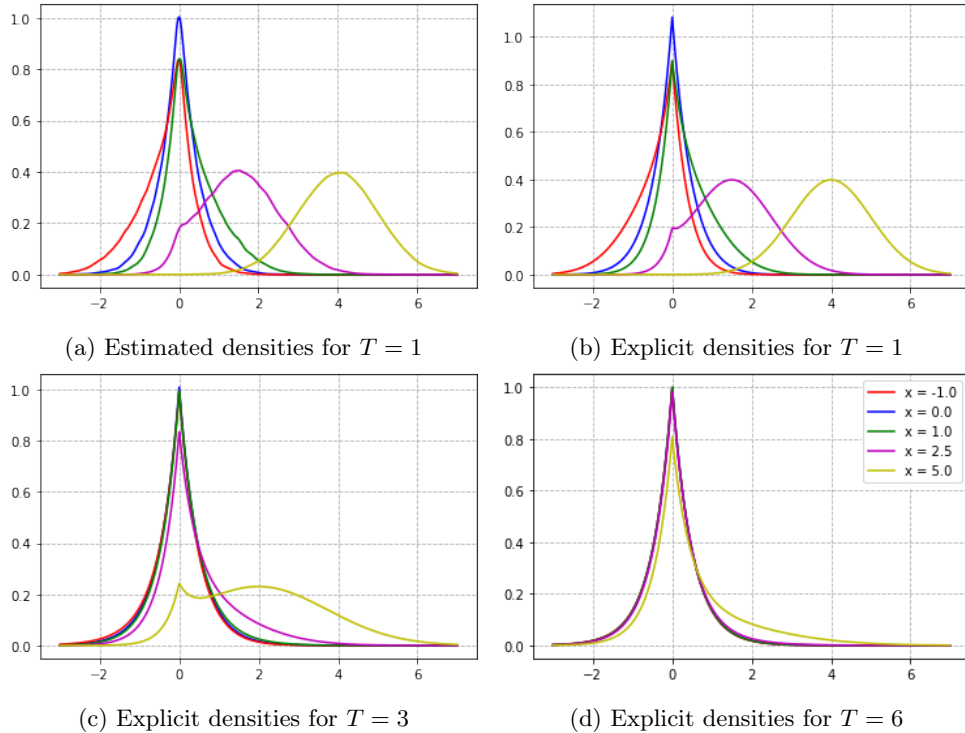


Figure 2.1 – The transition density function for various initializations  $x$

We also generate an example of a solution sample path with time-step  $h = 0.006$  for various initializations  $x$  to confirm that point.

Now, we give the empirical convergence orders obtained for various initializations  $x$ . These orders are given by the slopes of the regression lines in a log-log scale. The parameters used here are  $T = 5$ ,

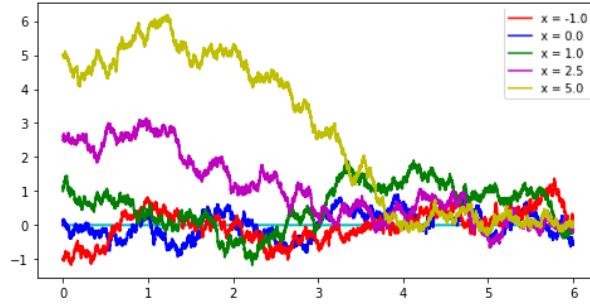


Figure 2.2 – Example of a solution sample paths for various initializations  $x$

$N = 500000$  and step sizes  $h \in \left\{ \frac{T}{8}, \frac{T}{16}, \frac{T}{32}, \frac{T}{64}, \frac{T}{128}, \frac{T}{256}, \frac{T}{512} \right\}$ .

| Initial value $x$           | -1.0 | 0.0  | 1.0  | 2.5  | 5.0  |
|-----------------------------|------|------|------|------|------|
| Empirical convergence order | 0.77 | 0.77 | 0.77 | 0.77 | 0.70 |

We can see from the above table that the empirical convergence orders are stable with respect to the initial value for this type of diffusion. A small difference is observed for the initial value  $x = 5.0$ : then the process starts far from the discontinuity point zero and the time horizon is not long enough for the process to visit it with high probability. When the process does not reach the discontinuity, the Euler scheme is exact making the error smaller and the influence of the kernel estimation error stronger.

### Dependence of the order of convergence on the jump height

To underline the influence of the jump height equal to  $2\theta$ , we start by generating an example of a sample path for two values of  $\theta \in \{1, 10\}$  with fixed time-horizon  $T = 1$ , time-step  $h = 0.003$  and initial value  $x = 0.0$ .

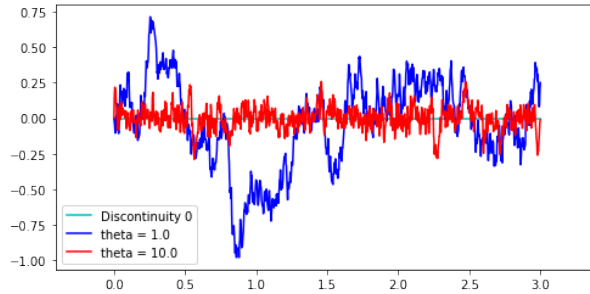


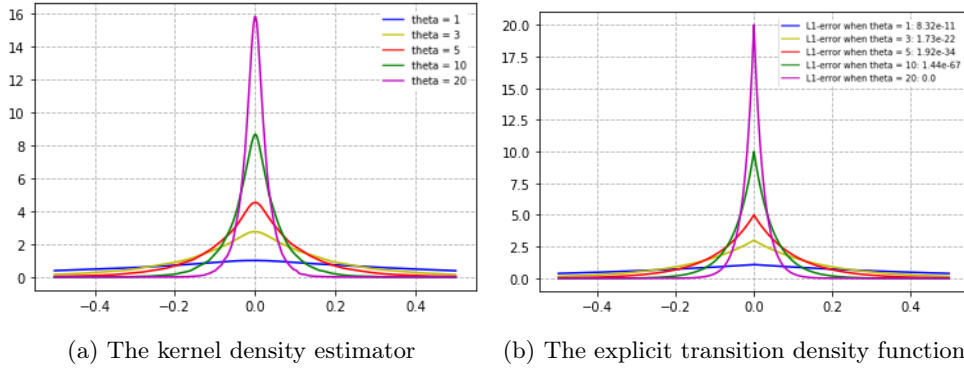
Figure 2.3 – Example of a solution sample paths for various  $\theta$

We observe that when  $\theta$  is big, the process is more likely to visit the discontinuity multiple times than for a smaller value of  $\theta$ . We can explain this by introducing the process  $Y_t = \frac{1}{\theta} X_t$  that starts from  $Y_0 = \frac{x}{\theta}$  and has the following dynamics:

$$Y_t = Y_0 + \frac{1}{\theta} W_t - \int_0^t \text{sgn}(Y_t) dt.$$

The diffusion coefficient equal to  $1/\theta$  becomes very small when  $\theta$  becomes large so that the process has an almost deterministic behaviour and is stuck to the discontinuity point zero by the drift.

Also, the explicit transition density tends to the Laplace density  $\theta e^{-2\theta|x|}$  when  $t \rightarrow +\infty$ . We generate the plots of the kernel density estimate and the explicit transition density function for various  $\theta = \{1, 3, 5, 10, 20\}$ . We choose  $T = 1$ ,  $x = 0.0$ ,  $N = 100000$  and  $h = 0.0001$ .

Figure 2.4 – The transition density function for various  $\theta$ 

We can see that when  $\theta$  is big, the density converges quickly towards the Laplace density. In Figure 2.4b, we confirm this behaviour by giving the  $L^1$ -error between the explicit transition densities and the Laplace densities for each  $\theta$ .

Now, we give the empirical convergence orders obtained for different values of  $\theta$ . The parameters used here are  $T = 1$ ,  $x = 0.0$  and  $N = 800000$ . We choose ranges of step-sizes  $h$  depending on  $\theta$  since for small  $\theta$ , the discretization error are smaller and the kernel estimation error comparatively more influent. For large  $\theta$ , a large time-step implies a very large error because on each time-step, when starting close to the discontinuity point zero, the Euler scheme will move far away to the other side of this discontinuity, a behaviour forbidden for the limiting SDE by the large inward pointing drift.

| Jump-height $\theta$    | 1.0                                                  | 3.0                                                  | 5.0                                                     | 10.0                                                    | 20.0                                                    |
|-------------------------|------------------------------------------------------|------------------------------------------------------|---------------------------------------------------------|---------------------------------------------------------|---------------------------------------------------------|
| time-step range $h$     | $\left\{\frac{1}{2^3}, \dots, \frac{1}{2^8}\right\}$ | $\left\{\frac{1}{2^4}, \dots, \frac{1}{2^9}\right\}$ | $\left\{\frac{1}{2^5}, \dots, \frac{1}{2^{10}}\right\}$ | $\left\{\frac{1}{2^7}, \dots, \frac{1}{2^{12}}\right\}$ | $\left\{\frac{1}{2^8}, \dots, \frac{1}{2^{13}}\right\}$ |
| Empiric. converg. order | 0.76                                                 | 0.77                                                 | 0.77                                                    | 0.77                                                    | 0.72                                                    |

We can see from the above table that the empirical convergence orders are relatively stable with respect to the jump-height for this type of diffusion. A small difference is observed for  $\theta = 20.0$  since the error is still large for the time-steps considered.

### 2.5.2 General case

To our knowledge, no closed-form of a density of  $X_t$  solving (2.5.1) is available for general  $\alpha, \beta \in \mathbb{R}$ . The idea is still to estimate the  $L^1$ -norm of the difference of the densities at maturity  $T$  but this time, instead of comparing  $p^h(T, \cdot)$  to  $p(T, \cdot)$ , we will compare  $p^h(T, \cdot)$  to  $p^{h/2}(T, \cdot)$  and the expected behaviour is:

$$\begin{aligned} \left\| p^h(T, \cdot) - p^{h/2}(T, \cdot) \right\|_{L^1} &\leq \left\| p^h(T, \cdot) - p(T, \cdot) \right\|_{L^1} + \left\| p^{h/2}(T, \cdot) - p(T, \cdot) \right\|_{L^1} \\ &\leq \left( \frac{3}{2} + \ln(2) \right) \tilde{C} \left( 1 + \ln \left( \frac{T}{h} \right) \right) h. \end{aligned}$$

In order to estimate  $p^h(T, \cdot)$ , we use, once again, a kernel density estimator but this time, we choose the Gaussian kernel defined by:

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{x^2}{2} \right) \quad \text{for } x \in \mathbb{R}.$$

We make this choice since no explicit density is available to estimate the bandwidth (2.5.2) and for Gaussian kernels we can obtain using the so-called Silverman's rule of thumb [80]:

$$\epsilon = cN^{-1/5} \quad \text{with} \quad c = 0.9 \times \min \left( \hat{\sigma}, \frac{IQR}{1.34} \right)$$

where the standard deviation  $\hat{\sigma}$  and the interquantile range  $IQR$  are easily computed from the sample of size  $N$ . When the density to estimate is a bimodal mixture, we apply Silverman's rule of thumb on each mode.

For  $(X_T^{i,h/2})_{1 \leq i \leq N}$  i.i.d. variables with density  $p^{h/2}(T, \cdot)$ , the kernel density estimator of this latter is then defined by:

$$p_{\epsilon,N}^{h/2}(T, x) = \frac{1}{\epsilon N} \sum_{j=1}^N K\left(\frac{x - X_T^{j,h/2}}{\epsilon}\right)$$

and we make the following trapezoidal approximation:

$$\begin{aligned} \|p^h(T, \cdot) - p^{h/2}(T, \cdot)\|_{L^1} &\simeq \sum_{i=1}^{N-1} \frac{1}{2} (X_T^{(i+1),h} - X_T^{(i),h}) \left\{ \left| p_{\epsilon,N}^h(T, X_T^{(i+1),h}) - p_{\epsilon,N}^{h/2}(T, X_T^{(i+1),h}) \right| \right. \\ &\quad \left. + \left| p_{\epsilon,N}^h(T, X_T^{(i),h}) - p_{\epsilon,N}^{h/2}(T, X_T^{(i),h}) \right| \right\}. \end{aligned}$$

In what follows, we will study the case of outward pointing diffusions i.e.  $\alpha < 0 < \beta$  and observe how the initial value and the jump-height influences the error. Beforehand, we observe the convergence rate in total variation when varying the time-step  $h$ .

### Illustration of the theoretical order of convergence in total variation

To observe the convergence rate in total variation for the case  $\alpha = -3.0$ ,  $\beta = 4.0$  and  $x = 0.0$ , we fix the time horizon  $T = 1$  and the number  $N = 250000$  of i.i.d. samples in the kernel density estimator large enough in order to observe the effect of the time-step  $h$  on the error. The simulation is done with  $R = 20$  Monte-Carlo runs. We obtain the following results for the estimation of the error and the associated precision:

| Evolution of the total variation error w.r.t. $h$ |            |                       |                   |                   |
|---------------------------------------------------|------------|-----------------------|-------------------|-------------------|
| Time-step $h$                                     | Estimation | Precision             | Ratio of decrease | Theoretical Ratio |
| $T/32$                                            | 0.1105     | $3.58 \times 10^{-4}$ | $\times$          | $\times$          |
| $T/64$                                            | 0.0763     | $3.17 \times 10^{-4}$ | 1.45              | 1.76              |
| $T/128$                                           | 0.0478     | $2.31 \times 10^{-4}$ | 1.60              | 1.79              |
| $T/256$                                           | 0.0279     | $1.91 \times 10^{-4}$ | 1.71              | 1.80              |
| $T/512$                                           | 0.0156     | $1.92 \times 10^{-4}$ | 1.79              | 1.82              |
| $T/1024$                                          | 0.0081     | $1.15 \times 10^{-4}$ | 1.93              | 1.84              |
| $T/2048$                                          | 0.0043     | $1.21 \times 10^{-4}$ | 1.87              | 1.85              |

- We observe that the ratio of successive estimations  $\frac{\text{Estimation}(h)}{\text{Estimation}(h/2)}$  is roughly around 1.72.
- The last column refers to the theoretical ratios equal to  $2^{\left(\frac{1+\ln(T/h)}{1+\ln(2T/h)}\right)}$  which is the expected behaviour of the error. On the range of values  $\left\{\frac{T}{256}, \frac{T}{512}, \frac{T}{1024}, \frac{T}{2048}\right\}$ , both the empirical and the theoretical ratios are equal to 1.82 in average.
- Moreover, the order of convergence in total variation of the Euler scheme is here equal to 0.79. This order is given, once again, by the slope of the regression line in a log-log scale.

### Dependence of the order of convergence on the initial value $x$ for fixed $\alpha = -3.0$ and $\beta = 4.0$

To underline the influence of the initial value of the SDE, we start by generating plots of the estimated transition density function for various initializations  $x \in \{-1.4, -0.4, -0.2, -0.15, 0.0, 0.6\}$ . We choose  $\alpha = -3.0$ ,  $\beta = 4.0$ ,  $T = 1$ ,  $N = 100000$  and  $h = 0.0001$ .

We see from Figure 2.5 that when the process starts from the discontinuity point  $x = 0.0$  or close to it  $x \in \{-0.2, -0.15\}$ , we have a bimodal mixture. When starting far from the discontinuity point zero, we are less likely to visit it and the distribution is Gaussian-like. We also confirm this point by an example

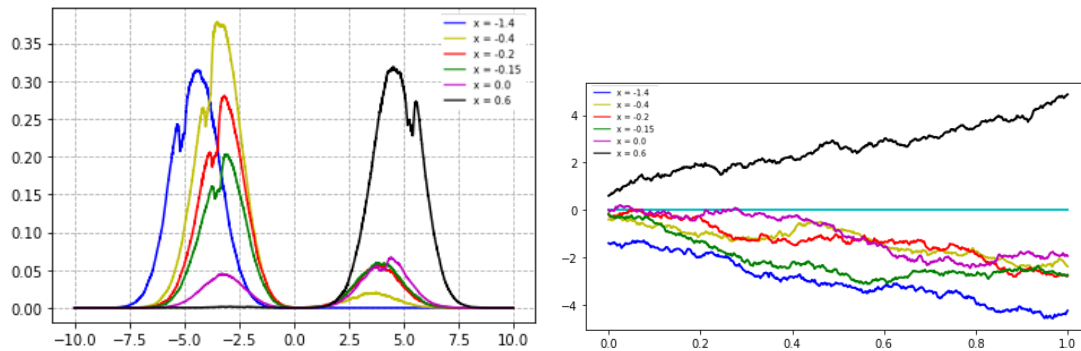


Figure 2.5 – Transition density functions and examples of a solution sample paths for various initializations  $x$

of a solution sample path with time-step  $h = 0.001$  for various initializations  $x$ .

Now, we give the empirical convergence orders obtained for various initializations  $x$ . These orders are given, once again, by the slopes of the regression lines in log-log scales. The parameters used here are  $T = 1$ ,  $N = 500000$  and step sizes  $h \in \left\{ \frac{T}{16}, \frac{T}{32}, \frac{T}{64}, \frac{T}{128}, \frac{T}{256}, \frac{T}{512}, \frac{T}{1024}, \frac{T}{2048} \right\}$ .

| Initial value $x$           | -1.4 | -0.4 | -0.2 | 0.0  | 0.6  |
|-----------------------------|------|------|------|------|------|
| Empirical convergence order | 0.28 | 0.68 | 0.82 | 0.71 | 0.51 |

We can see from the above table that the empirical convergence orders seem to depend on the initial value and the spectrum of orders obtained for different initial values is very broad with values between 0.28 and 0.82. When starting further and further from the discontinuity point zero, we first obtain a better order of convergence but when  $|x|$  becomes large it deteriorates since the kernel estimation error becomes more influent.

### Dependence of the order of convergence on the jump height

To underline the influence of the jump height equal to  $(\beta - \alpha)$ , we start by generating a sample path for two values of  $(\alpha, \beta) \in \{(-4.0, 3.0), (-0.6, 1.0)\}$  with fixed time-horizon  $T = 1$  and initial value  $x = 0.0$ .

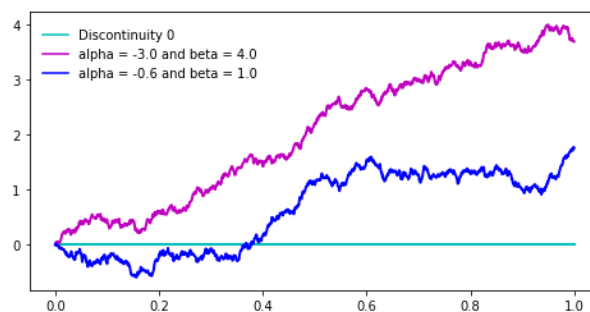


Figure 2.6 – Solution sample paths for various  $(\alpha, \beta)$

We observe that the solution drifts away from the discontinuity point zero and therefore, there are not many chances for a drift correction to take place.

Now, we give the empirical convergence orders obtained for various  $(\alpha, \beta)$ . These orders are given, once again, by the slopes of the regression lines in log-log scales. The parameters used here are  $T = 1$ ,  $N = 500000$ ,  $x = 0.0$  and step-sizes  $h \in \left\{ \frac{T}{8}, \frac{T}{16}, \frac{T}{32}, \frac{T}{64}, \frac{T}{128}, \frac{T}{256}, \frac{T}{512} \right\}$ .

| $(\alpha, \beta)$           | $(-6.0, 8.0)$ | $(-3.0, 4.0)$ | $(-1.5, 2.0)$ | $(-0.75, 1.0)$ | $(-0.375, 0.5)$ |
|-----------------------------|---------------|---------------|---------------|----------------|-----------------|
| Empirical convergence order | 0.51          | 0.66          | 0.90          | 1.02           | 0.78            |

We can see from the table above that the empirical convergence orders are not stable with respect to the jump-height for outward pointing drift diffusions. Enlarging the jump height makes the solution to drift away from the discontinuity point zero and increases the error.

### 2.5.3 Conclusion

We were able, through our numerical experiments, to confirm our theoretical estimates for the convergence rate in total variation of  $\mu_T^h$  to its limit  $\mu_T$  since the order 1 up to a logarithmic factor was recovered.

Moreover, the study conducted when varying the type of drift (inward pointing or outward pointing) has highlighted several features. Our results and interpretations coincide with those obtained by Göttlich, Lux and Neuenkirch in [28] when they estimate the root mean-squared strong error. Like them, we show that for inward pointing drift coefficients, the convergence order is independent of the initial value and the jump-height. This is not the case for outward pointing drift coefficients: the numerical orders are less stable in the initial value and the jump-height. A possible explanation is that the inward pointing drift coefficient engenders many drift changes, while only few drift changes occur in the case of an outward pointing drift coefficient. The solution, in the latter case, can quickly drift away from the discontinuity and because of a small probability of a drift change, the empirical convergence rate might be subject to rare event effects and the linear regression estimates become questionable.



## A.1 Appendix

**Lemma A.1.1.** *Let  $\sigma \in \mathbb{R}^{d \times d}$  be a non-degenerate matrix and  $\tilde{g} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  be a measurable and locally integrable function, the spatial divergence in the sense of distributions of which is a Radon measure denoted by  $\nabla \cdot \tilde{g}(dy)$ . Then, the function  $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$  defined by  $g(x) = \sigma^{-1} \tilde{g}(\sigma x)$  is locally integrable and its spatial divergence  $\nabla \cdot g(dx)$  in the sense of distributions is the image of  $|\det(\sigma^{-1})| \nabla \cdot \tilde{g}$  by  $y \mapsto \sigma^{-1}y$ . In particular, the total mass of  $\nabla \cdot g(dx)$  is equal to  $|\det(\sigma^{-1})|$  times the total mass of  $\nabla \cdot \tilde{g}(dy)$  and when  $\nabla \cdot \tilde{g}(dy)$  admits the density  $f(y)$  with respect to the Lebesgue measure, then  $\nabla \cdot g(dx)$  admits the density  $f(\sigma x)$ .*

*Proof.* The local integrability of  $g$  is easily obtained by the change of variables  $y = \sigma x$ . For any  $\mathcal{C}^\infty$  function  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$  with compact support, we obtain using the same change of variables that

$$\begin{aligned} \int_{\mathbb{R}^d} g(x) \cdot \nabla_x \varphi(x) dx &= \int_{\mathbb{R}^d} \sigma^{-1} \tilde{g}(\sigma x) \cdot \nabla_x \varphi(x) dx = |\det(\sigma^{-1})| \int_{\mathbb{R}^d} \tilde{g}(y) \cdot (\sigma^{-1})^* \nabla_x \varphi(\sigma^{-1}y) dy \\ &= |\det(\sigma^{-1})| \int_{\mathbb{R}^d} \tilde{g}(y) \cdot \nabla_y [\varphi(\sigma^{-1}y)] dy = -|\det(\sigma^{-1})| \int_{\mathbb{R}^d} \varphi(\sigma^{-1}y) \nabla \cdot \tilde{g}(dy), \end{aligned}$$

which implies the first statement. The one concerning the total masses immediately follows and the one concerning the densities is obtained by the inverse change of variables  $x = \sigma^{-1}y$ .  $\square$

For  $t > 0$ , let  $G_t$  denote the heat kernel in  $\mathbb{R}^d$ :  $G_t(x) = \frac{1}{\sqrt{(2\pi t)^d}} \exp\left(-\frac{|x|^2}{2t}\right)$ . The following lemma provides a set of estimates that are very useful:

**Lemma A.1.2.** *The function  $G_t(x)$  solves the heat equation:*

$$\partial_t G_t(x) - \frac{1}{2} \Delta G_t(x) = 0, \quad (t, x) \in [0, +\infty) \times \mathbb{R}^d. \quad (\text{A.1.1})$$

We have estimates of the  $L^1$ -norm of the first order time derivative and the spatial derivatives of  $G$  up to the third order:

$$\|\partial_t G_t\|_{L^1} \leq \frac{d}{t}, \quad (\text{A.1.2})$$

$$\|\partial_{x_i} G_t\|_{L^1} = \sqrt{\frac{2}{\pi t}}, \quad (\text{A.1.3})$$

$$\left\| \frac{\partial^2 G_t}{\partial x_i^2} \right\|_{L^1} \leq \frac{2}{t}, \quad (\text{A.1.4})$$

$$\left\| \frac{\partial^2 G_t}{\partial x_i \partial x_j} \right\|_{L^1} = \frac{2}{\pi t} \quad \text{when } j \neq i, \quad (\text{A.1.5})$$

$$\left\| \frac{\partial^3 G_t}{\partial x_j \partial x_i^2} \right\|_{L^1} \leq \begin{cases} \frac{2}{t^{3/2}} \sqrt{\frac{2}{\pi}} & \text{when } j \neq i, \\ \frac{5}{t^{3/2}} \sqrt{\frac{2}{\pi}} & \text{when } j = i. \end{cases} \quad (\text{A.1.6})$$

*Proof.* Let us compute the estimate (A.1.3). To do so, we use Fubini's theorem and obtain:

$$\begin{aligned} \|\partial_{x_i} G_t\|_{L^1} &= \int_{\mathbb{R}^d} \frac{|x_i|}{t} \frac{1}{(2\pi t)^{d/2}} \exp\left(-\sum_{j=1}^d \frac{x_j^2}{2t}\right) dx_1 \dots dx_d \\ &= \left( \int_{\mathbb{R}} \frac{|y|}{t} \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{y^2}{2t}\right) dy \right) \times \left( \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{y^2}{2t}\right) dy \right)^{d-1} \\ &= \sqrt{\frac{2}{\pi t}}. \end{aligned}$$

We can express the second and third spatial derivatives of  $G$  as:

$$\frac{\partial^2}{\partial x_i \partial x_j} G_t(x) = \begin{cases} \frac{x_i x_j}{t^2} G_t(x) & \text{when } j \neq i, \\ \left(-1 + \frac{x_i^2}{t}\right) \frac{G_t(x)}{t} & \text{when } j = i. \end{cases} \quad \text{and} \quad \frac{\partial^3}{\partial x_j \partial x_i^2} G_t(x) = \begin{cases} \left(1 - \frac{x_i^2}{t}\right) \frac{x_j}{t^2} G_t(x) & \text{when } j \neq i, \\ \left(3 - \frac{x_i^2}{t}\right) \frac{x_i}{t^2} G_t(x) & \text{when } j = i. \end{cases}$$

Using Fubini's theorem as for the estimate (A.1.3),  $\int_{\mathbb{R}} \frac{y^2}{t^2} \frac{e^{-\frac{y^2}{2t}}}{\sqrt{2\pi t}} dy = \frac{1}{t}$  and  $\int_{\mathbb{R}} \frac{|y|^3}{t^3} \frac{e^{-\frac{y^2}{2t}}}{\sqrt{2\pi t}} dy = \frac{2}{t^{3/2}} \sqrt{\frac{2}{\pi}}$ , we obtain the estimates (A.1.4), (A.1.5) and (A.1.6). As for the estimate (A.1.2), we deduce it from the heat equation (A.1.1) and the estimate (A.1.4).  $\square$

**Lemma A.1.3.** *We have:*

$$\forall n \geq 2, \quad \sum_{k=1}^{n-1} \frac{1}{\sqrt{k} \sqrt{n-k}} \leq \pi - \frac{2}{n}.$$

*Proof.* We define the function  $f(x) = \frac{1}{\sqrt{x}\sqrt{1-x}}$  on  $(0, 1)$ . We easily check that  $\forall x \in (0, 1)$ ,  $f(x) \geq f(1/2) = 2$  and that  $\int_0^1 f(x) dx = \pi$ . Using the monotonicity of  $f$  on  $(0, 1/2]$  and  $[1/2, 1)$ , we obtain:

$$\begin{cases} \int_{\frac{k-1}{n}}^{\frac{k}{n}} f(x) dx \geq \frac{1}{n} f\left(\frac{k}{n}\right) & \text{when } 1 \leq k \leq \frac{n}{2}, \\ \int_{\frac{k}{n}}^{\frac{k+1}{n}} f(x) dx \geq \frac{1}{n} f\left(\frac{k}{n}\right) & \text{when } \frac{n}{2} \leq k \leq (n-1). \end{cases}$$

Therefore,

• When  $n$  is even:

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^{\frac{n}{2}} f\left(\frac{k}{n}\right) + \frac{1}{n} \sum_{k=\frac{n}{2}+1}^{n-1} f\left(\frac{k}{n}\right) &\leq \sum_{k=1}^{\frac{n}{2}} \int_{\frac{k-1}{n}}^{\frac{k}{n}} f(x) dx + \sum_{k=\frac{n}{2}+1}^{n-1} \int_{\frac{k}{n}}^{\frac{k+1}{n}} f(x) dx \\ &= \int_0^{\frac{1}{2}} f(x) dx + \int_{\frac{1}{2}+\frac{1}{n}}^1 f(x) dx \\ &= \pi - \int_{\frac{1}{2}}^{\frac{1}{2}+\frac{1}{n}} f(x) dx \leq \pi - \frac{2}{n}. \end{aligned}$$

• When  $n$  is odd:

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^{\frac{n-1}{2}} f\left(\frac{k}{n}\right) + \frac{1}{n} \sum_{k=\frac{n+1}{2}}^{n-1} f\left(\frac{k}{n}\right) &\leq \sum_{k=1}^{\frac{n-1}{2}} \int_{\frac{k-1}{n}}^{\frac{k}{n}} f(x) dx + \sum_{k=\frac{n+1}{2}}^{n-1} \int_{\frac{k}{n}}^{\frac{k+1}{n}} f(x) dx \\ &= \int_0^{\frac{1}{2}-\frac{1}{2n}} f(x) dx + \int_{\frac{1}{2}+\frac{1}{2n}}^1 f(x) dx \\ &= \pi - \int_{\frac{1}{2}-\frac{1}{2n}}^{\frac{1}{2}+\frac{1}{2n}} f(x) dx \leq \pi - \frac{2}{n}. \end{aligned}$$

We can conclude.  $\square$

**Lemma A.1.4.** *For  $0 < a \leq x \leq T$ ,*

$$\int_a^x \frac{dy}{y\sqrt{x-y}} = \frac{1}{\sqrt{x}} \ln \left( \frac{(\sqrt{x} + \sqrt{x-a})^2}{a} \right) \leq \frac{1}{\sqrt{x}} \ln \left( \frac{4x}{a} \right).$$

*Proof.* Using the change of variable  $u = \sqrt{x-y}$  then a partial fraction decomposition, we obtain:

$$\int_a^x \frac{dy}{y\sqrt{x-y}} = 2 \int_0^{\sqrt{x-a}} \frac{du}{u^2-x} = \left[ \frac{1}{\sqrt{x}} \ln \left( \frac{\sqrt{x}+u}{\sqrt{x}-u} \right) \right]_0^{\sqrt{x-a}}. \quad \square$$

**Lemma A.1.5.** For  $0 < r \leq s \leq T$ ,

$$\frac{1}{2} \int_0^r \frac{\ln(s-u) - \ln(r-u)}{\sqrt{u}} du = \frac{s-r}{\sqrt{s}+\sqrt{r}} \ln \left( \frac{(\sqrt{s}+\sqrt{r})^2}{s-r} \right) + 2\sqrt{r} \ln \left( 1 + \frac{\sqrt{s}-\sqrt{r}}{2\sqrt{r}} \right).$$

*Proof.* We start by applying the change the variable  $\theta = \sqrt{u}$  and obtain

$$\begin{aligned} \frac{1}{2} \int_0^r \frac{\ln(s-u) - \ln(r-u)}{\sqrt{u}} du &= \int_0^{\sqrt{r}} \left( \ln(s-\theta^2) - \ln(r-\theta^2) \right) d\theta \\ &= \int_0^{\sqrt{r}} \left( \ln(\sqrt{s}-\theta) + \ln(\sqrt{s}+\theta) - \ln(\sqrt{r}-\theta) - \ln(\sqrt{r}+\theta) \right) d\theta. \end{aligned}$$

A simple integration of  $\ln(x)$  permits us to conclude.  $\square$

The next lemma is a discrete version of Gronwall's lemma and was proved by Holte [38].

**Lemma A.1.6.** If  $(y_n)_{n \in \mathbb{N}}$ ,  $(f_n)_{n \in \mathbb{N}}$  and  $(g_n)_{n \in \mathbb{N}}$  are non-negative sequences and

$$y_n \leq f_n + \sum_{i=0}^{n-1} g_i y_i \quad \text{for } n \in \mathbb{N}$$

then

$$y_n \leq f_n + \sum_{i=0}^{n-1} f_i g_i \exp \left( \sum_{j=i+1}^{n-1} g_j \right) \quad \text{for } n \in \mathbb{N}.$$



## Part II

# McKean-Vlasov Stochastic differential Equations



## CHAPTER 3

# BIAS ESTIMATION AND ANTITHETIC SAMPLING FOR SYSTEMS OF PARTICLES INTERACTING THROUGH MOMENTS

*Ce chapitre est un article écrit avec Benjamin Jourdain. Il a été publié dans ESAIM: Proceedings and Surveys Volume 65, pp.219–235, 2019 sous le nom "Bias behaviour and antithetic sampling in mean-field particle approximations of SDEs nonlinear in the sense of McKean".*

### Abstract

In this paper, we prove that the weak error between a stochastic differential equation with nonlinearity in the sense of McKean given by moments and its approximation by the Euler discretization with time-step  $h$  of a system of  $N$  interacting particles is  $\mathcal{O}(N^{-1} + h)$ . We provide numerical experiments confirming this behaviour and showing that it extends to more general mean-field interaction and study the efficiency of the antithetic sampling technique on the same examples.

**Keywords:** Diffusions non-linear in the sense of McKean, particle approximations, bias estimation, antithetic variance reduction.

### 3.1 Introduction

According to [84], the strong rate of convergence of particle approximations of McKean-Vlasov Stochastic differential equations with Lipschitz coefficients is  $\mathcal{O}(N^{-1/2})$  when  $N$  denotes the number of particles. This rate is driven by the statistical error and one may wonder whether the bias vanishes quicker. A parallel can be drawn with the time discretization of standard stochastic differential equations where, for Lipschitz coefficients, the strong rate of convergence of the explicit Euler-Maruyama scheme is  $\mathcal{O}(\sqrt{h})$  [54] with  $h > 0$  denoting the time-step. Using the Feynman-Kac partial differential equation associated with the stochastic differential equation, Talay and Tubaro [85] checked that, for smooth coefficients, the weak error behaves in  $\mathcal{O}(h)$  and can even be expanded in powers of  $h$ . In the context of particles interacting through jumps, the  $\mathcal{O}(N^{-1})$  behaviour of the bias is known. According to [29], for particle approximations of generalized Boltzmann equations, the total variation distance between the law of the path of a particle and the one of the limiting nonlinear Boltzmann process behaves in  $\mathcal{O}(N^{-1})$ . For Feynman-Kac particle models, expansions in powers of  $1/N$  are obtained in [18].

The interest in the bias introduced by particle approximations is motivated by numerical efficiency. Indeed, the numerical experiments performed in Section 3.3 show a general  $\mathcal{O}(N^{-1})$  behaviour of the bias, even in models with not so smooth coefficients. Under this behaviour, simulating  $\sqrt{N}$  independent copies of the system with  $\sqrt{N}$  particles leads to the same order of error (bias with the same order as the  $\mathcal{O}(N^{-1/4} \times N^{-1/4}) = \mathcal{O}(N^{-1/2})$  statistical error) as one simulation of the system with  $N$  particles (bias smaller than the  $\mathcal{O}(N^{-1/2})$  statistical error). And the former approach is less expensive than the latter as soon as the computational cost of the particle system grows more than linearly with the number of particles. The behaviour of the bias is also of interest in order to adapt to the number of particles the multilevel Monte Carlo method introduced by Giles [24] in the context of time discretization of standard stochastic differential equations. In [35], Haji-Ali and Tempone combine both discretizations through the Multi-index Monte Carlo method. In this perspective, another interesting question is the possibility to take advantage of the antithetic sampling technique introduced in [34] to reduce the variance (see [15, 11] or [25] p57 for the use of this technique in nested Monte Carlo computations). Does the variance of the difference between the empirical mean of the system with  $2N$  particles driven by the i.i.d. couples  $(X_0^i, W^i)_{1 \leq i \leq 2N}$  ( $X_0^i$  and  $W^i$  respectively denote the initial condition and the Brownian motion of the  $i$ -th particle) and the mean of the empirical means of the two independent systems with  $N$  particles respectively driven by  $(X_0^i, W^i)_{1 \leq i \leq N}$  and  $(X_0^i, W^i)_{N+1 \leq i \leq 2N}$  converge quicker to 0 than  $\mathcal{O}(N^{-1})$ ?

In this paper, using the Feynman-Kac partial differential equation associated with the limiting nonlinear process and its moments, we prove the respective  $\mathcal{O}(N^{-1})$  and  $\mathcal{O}(N^{-1} + h)$  behaviour of the bias for systems of diffusive particles interacting through moments and their Euler discretization with time step  $h$ . Of course, the computational cost of such systems is linear in the number  $N$  of particles and the above numerical motivation is not valid. Nevertheless this result can be seen as a first step before addressing more general interactions which could necessitate more advanced tools like the master equation for mean-field games introduced by Lions in his seminal lectures at *Collège de France* and studied in [12] from a probabilistic point of view. Theorem 3.2 [57] is proved using the master equation and implies the  $\mathcal{O}(N^{-1})$  behaviour of the bias for one-dimensional stochastic differential equations with general interaction in the drift coefficient but no interaction in the diffusion coefficient. We provide numerical experiments showing that the  $\mathcal{O}(N^{-1})$  behaviour holds in more general situations including ones with non smooth coefficients. Last, on the same examples, we check that the antithetic variance does not generally decrease quicker than  $\mathcal{O}(N^{-1})$ .

### 3.2 Estimation of the bias for systems of particles interacting through moments

Let  $\alpha : \mathbb{R}^n \rightarrow \mathbb{R}^p$  be Lipschitz continuous and  $\sigma = (\sigma_j^l)_{1 \leq j \leq n, 1 \leq l \leq d} : [0, T] \times \mathbb{R}^p \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times d}$ ,  $b = (b_j)_{1 \leq j \leq n} : [0, T] \times \mathbb{R}^p \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  be Lipschitz continuous in their  $p + n$  last components uniformly in their first component and such that  $\sup_{t \in [0, T]} (|\sigma(t, 0, 0)| + |b(t, 0, 0)|) < \infty$ . We consider the stochastic differential equation in dimension  $n$  nonlinear in the sense of McKean

$$X_t = X_0 + \int_0^t \sigma(s, \mathbb{E}[\alpha(X_s)], X_s) dW_s + \int_0^t b(s, \mathbb{E}[\alpha(X_s)], X_s) ds, \quad t \in [0, T] \quad (3.2.1)$$



where  $X_0$  is some square integrable  $\mathbb{R}^n$ -valued random variable independent from the  $d$ -dimensional Brownian motion  $(W_t)_{t \geq 0}$ . The drift and diffusion coefficient at time  $s$  depend on the law of  $X_s$  through the moments  $\mathbb{E}[\alpha(X_s)]$ . Since  $\sigma$  may only depend on a subset of coordinates of this expectation on  $b$  on another subset, moments in drift and diffusion can differ as well as their respective dimensions. By a solution, we mean a continuous process  $(X_t)_{t \in [0, T]}$  adapted to the filtration generated by the Brownian motion  $W$  and  $X_0$  such that  $\sup_{t \in [0, T]} \mathbb{E}[|\alpha(X_t)|] < \infty$  and the above equation holds with the integrals with respect to  $dW_s$  and  $ds$  making sense. Notice that for any solution,  $x \mapsto (\sigma(s, \mathbb{E}[\alpha(X_s)], x), b(s, \mathbb{E}[\alpha(X_s)], x))$  has affine growth uniformly for  $s \in [0, T]$ . With the square integrability of  $X_0$ , this implies that  $\mathbb{E} \left[ \sup_{t \in [0, T]} |X_t|^2 \right] + \sup_{0 \leq s < t \leq T} \frac{\mathbb{E}[|X_t - X_s|^2]}{t-s} < \infty$  so that, by Lipschitz continuity of  $\alpha$ ,  $t \mapsto \mathbb{E}[\alpha(X_t)]$  is Hölder continuous with exponent  $1/2$  on  $[0, T]$ .

The particle approximation of the SDE nonlinear in the sense of McKean (3.2.1) is given by the system with mean-field interaction for  $1 \leq i \leq N$  and  $t \in [0, T]$ :

$$X_t^{i,N} = X_0^i + \int_0^t \sigma \left( s, \frac{1}{N} \sum_{j=1}^N \alpha(X_s^{j,N}), X_s^{i,N} \right) dW_s^i + \int_0^t b \left( s, \frac{1}{N} \sum_{j=1}^N \alpha(X_s^{j,N}), X_s^{i,N} \right) ds, \quad (3.2.2)$$

with  $(X_0^i, W^i)_{1 \leq i \leq N}$  i.i.d. copies of  $(X_0, W)$ . By the Lipschitz assumptions, existence and trajectorial uniqueness hold for this  $N \times n$  dimensional equation. The Yamada-Watanabe theorem ensures weak uniqueness and therefore exchangeability of the particles  $\left( (X_t^{i,N})_{t \in [0, T]} \right)_{1 \leq i \leq N}$ . Let us also introduce the Euler discretizations with time-step  $h > 0$  of the SDE (3.2.1) and the particle system : For  $t \in [0, T]$ ,  $1 \leq i \leq N$ ,

$$X_t^h = X_0 + \int_0^t \sigma \left( \tau_s^h, \mathbb{E} \left[ \alpha \left( X_{\tau_s^h}^h \right) \right], X_{\tau_s^h}^h \right) dW_s + \int_0^t b \left( \tau_s^h, \mathbb{E} \left[ \alpha \left( X_{\tau_s^h}^h \right) \right], X_{\tau_s^h}^h \right) ds, \text{ where } \tau_s^h = \lfloor s/h \rfloor h, \quad (3.2.3)$$

$$X_t^{i,N,h} = X_0^i + \int_0^t \sigma \left( \tau_s^h, \frac{1}{N} \sum_{j=1}^N \alpha \left( X_{\tau_s^h}^{j,N,h} \right), X_{\tau_s^h}^{i,N,h} \right) dW_s^i + \int_0^t b \left( \tau_s^h, \frac{1}{N} \sum_{j=1}^N \alpha \left( X_{\tau_s^h}^{j,N,h} \right), X_{\tau_s^h}^{i,N,h} \right) ds.$$

It is natural and convenient to consider that  $\tau_s^0 = s$ ,  $(X_t^0)_{t \in [0, T]} = (X_t)_{t \in [0, T]}$  and  $(X_t^{i,N,0})_{t \in [0, T], 1 \leq i \leq N} = (X_t^{i,N})_{t \in [0, T], 1 \leq i \leq N}$  and we use these notations in what follows.

Reasoning like in the laboratory example in [84] or in Theorem 1.3 [47], one easily checks the following result.

**Proposition 3.2.1.** *Strong existence and trajectorial uniqueness hold for the SDE nonlinear in the sense of McKean (3.2.1) and its Euler discretization (3.2.3). Moreover  $\sup_{h \geq 0} \mathbb{E} \left[ \sup_{t \in [0, T]} |X_t^h|^2 \right] \leq C(1 + \mathbb{E}[|X_0|^2])$  where the finite constant  $C$  does not depend on  $X_0$ . Last, if for  $i \in \mathbb{N}^*$ ,  $(X_t^{i,h})_{t \in [0, T]}$  denotes the process obtained by replacement of  $(X_0, W)$  by  $(X_0^i, W^i)$  in (3.2.3), then*

$$\exists C < +\infty, \forall N \in \mathbb{N}^*, \sup_{h \geq 0} \max_{1 \leq i \leq N} \mathbb{E} \left[ \sup_{t \in [0, T]} |X_t^{i,N,h} - X_t^{i,h}|^2 \right] \leq C \frac{1 + \mathbb{E}[|X_0|^2]}{N}.$$

If for  $i \in \mathbb{N}^*$ ,  $(X_t^i)_{t \in [0, T]}$  denotes the process obtained by replacement of  $(X_0, W)$  by  $(X_0^i, W^i)$  in (3.2.1), this implies the following estimation of the bias introduced by the particle discretization: for any

function  $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$  Lipschitz with constant  $L_\psi$  and in particular for each coordinate of  $\alpha$ ,

$$\begin{aligned} \forall s \in [0, T], \quad \left| \mathbb{E}[\psi(X_s^{1,N})] - \mathbb{E}[\psi(X_s)] \right| &= \left| \mathbb{E}[\psi(X_s^{1,N})] - \mathbb{E}[\psi(X_s^1)] \right| \leq L_\psi \mathbb{E}[|X_s^{1,N} - X_s^1|] \\ &\leq L_\psi \left( \sup_{h \geq 0} \mathbb{E} \left[ \sup_{t \in [0, T]} |X_t^{1,N,h} - X_t^{1,h}|^2 \right] \right)^{1/2} \\ &\leq \sqrt{C} L_\psi \frac{\sqrt{1 + \mathbb{E}[|X_0|^2]}}{\sqrt{N}}. \end{aligned}$$

The first inequality is crude since it prevents cancelations in average and one may wonder whether the bias converges faster to 0 as  $N \rightarrow \infty$ . Under additional regularity, the answer is positive, which is our main result.

**Theorem 3.2.2.** *Assume that*

- $\sigma$  is Lipschitz continuous in its  $p + n$  last components uniformly in its first component and such that  $\sup_{t \in [0, T]} (|\sigma(t, 0, 0)| + |b(t, 0, 0)|) < \infty$ ,
- $\alpha, \psi$  are two times continuously differentiable with bounded derivatives up to the order two and Lipschitz continuous second order derivatives,
- $a = \sigma\sigma^*$  and  $b$  are globally Lipschitz continuous, continuously differentiable with respect to their variables with index in  $\{2, \dots, 1+p\}$  with derivatives Lipschitz continuous with respect to their  $p+n$  last variables uniformly in their first variable,
- there exists  $\tilde{d} \in \mathbb{N}^*$ ,  $\tilde{\sigma} : [0, T] \times \mathbb{R}^p \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times \tilde{d}}$  such that for all  $(t, y, x) \in [0, T] \times \mathbb{R}^p \times \mathbb{R}^n$ ,  $a(t, y, x) = \tilde{\sigma}\tilde{\sigma}^*(t, y, x)$  and  $\tilde{\sigma}, b$  are continuous and two times continuously differentiable with respect to their  $n$  last variables with bounded derivatives up to the order two and second order derivatives Lipschitz continuous with respect to their  $n$  last variables uniformly in their  $1+p$  first variables,

Then

$$\exists C < +\infty, \forall h \geq 0, \forall N \in \mathbb{N}^*, \quad \sup_{t \in [0, T]} \left| \mathbb{E}[\psi(X_t^{1,N,h})] - \mathbb{E}[\psi(X_t)] \right| \leq C \left( \frac{1}{N} + h \right).$$

The idea of the proof is, like in [85], to use the Feynman-Kac partial differential equation associated with the nonlinear SDE (3.2.1) to first check that the estimation holds for the coordinates of  $\alpha$  before concluding that it holds for each test function  $\psi$ . The following proposition ensures existence and smoothness to this Feynman-Kac PDE which only depends on  $\sigma$  through  $a = \sigma\sigma^*$ . In its statement and its proof based on a stochastic flow approach, we take advantage of the flexibility given by the choice of a square root of  $a$  which explains the last assumption in Theorem 3.2.2. Let  $(\tilde{W}_t)_{t \geq 0}$  be a  $\tilde{d}$ -dimensional Brownian motion with coordinates  $(\tilde{W}_t^l)_{1 \leq l \leq \tilde{d}}$  and for  $(s, x) \in [0, T] \times \mathbb{R}^n$ ,  $(\tilde{X}_t^{s,x})_{t \in [s, T]}$  denote the solution to

$$\tilde{X}_t^{s,x} = x + \int_s^t \tilde{\sigma} \left( r, \mathbb{E}[\alpha(X_r)], \tilde{X}_r^{s,x} \right) d\tilde{W}_r + \int_s^t b \left( r, \mathbb{E}[\alpha(X_r)], \tilde{X}_r^{s,x} \right) dr, \quad t \in [s, T] \quad (3.2.4)$$

where the coefficients depend on  $\mathbb{E}[\alpha(X_r)]$  and not  $\mathbb{E}[\tilde{X}_r^{s,x}]$ .

**Proposition 3.2.3.** *Let  $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$  be two times continuously differentiable with bounded derivatives up to the order two and Lipschitz continuous second order derivatives. Under the last assumption in Theorem 3.2.2, for each  $t \in [0, T]$ , the function  $[0, t] \times \mathbb{R}^n \ni (s, x) \mapsto u_{t,\psi}(s, x) := \mathbb{E}[\psi(\tilde{X}_t^{s,x})]$  is once (resp. twice) continuously differentiable with respect to the time (resp. space) variable  $s$  (resp.  $x$ ) and such that  $u_{t,\psi}$  together with its spatial derivatives up to the order two are Lipschitz continuous in  $x$*

uniformly in  $0 \leq s \leq t \leq T$ . Moreover,  $u_{t,\psi}$  is a classical solution to the Feynman-Kac PDE

$$\begin{cases} \partial_s u_{t,\psi}(s, x) + \frac{1}{2} \sum_{j,k=1}^n a_{jk}(s, \mathbb{E}[\alpha(X_s)], x) \partial_{jk} u_{t,\psi}(s, x) + \sum_{j=1}^n b_j(s, \mathbb{E}[\alpha(X_s)], x) \partial_j u_{t,\psi}(s, x) = 0, \\ (s, x) \in [0, t] \times \mathbb{R}^n, \\ u_{t,\psi}(t, x) = \psi(x), \quad x \in \mathbb{R}^n. \end{cases} \quad (3.2.5)$$

Here  $\partial_j$  and  $\partial_{jk}$  denote the partial derivative with respect to the  $j$ -th coordinate of  $x$  and with respect to its  $j$ -th and  $k$ -th coordinates.

*Proof.* Even if this result seems to be well-known, we could not find its proof in the literature. Therefore, we are going to give a sketch of its proof. For notational simplicity, we set  $\tilde{\sigma}^0 = b$  and  $\tilde{W}_t^0 = t$ .

Let for  $j \in \{1, \dots, n\}$ ,  $e_j$  denote the  $j$ -th vector of the canonical basis of  $\mathbb{R}^n$  and for  $l \in \{0, \dots, \tilde{d}\}$ ,  $\partial \tilde{\sigma}^l = (\partial_{1+p+k} \tilde{\sigma}_j^l)_{1 \leq j, k \leq n}$  where  $\partial_{1+p+k}$  is the partial derivative with respect to the  $(1+p+k)$ -th variable. Let also for  $l \in \{0, \dots, \tilde{d}\}$  and  $y, z \in \mathbb{R}^n$ ,  $\partial^2 \tilde{\sigma}^l yz \in \mathbb{R}^n$  be defined by  $(\partial^2 \tilde{\sigma}^l yz)_j = \sum_{k,m=1}^n (\partial_{1+p+k} \partial_{1+p+m} \tilde{\sigma}_j^l) y_k z_m$  for  $j \in \{1, \dots, n\}$ . Since  $x \mapsto [\tilde{\sigma}, b](r, \mathbb{E}[\alpha(X_r)], x)$  has affine growth uniformly in  $r \in [0, T]$ , standard moment estimations ensure that

$$\forall q \geq 1, \exists C < +\infty, \forall (s, x) \in [0, T] \times \mathbb{R}^n, \quad \mathbb{E} \left[ \sup_{r \in [s, T]} |\tilde{X}_r^{s,x}|^q \right] \leq C(1 + |x|^q).$$

By [62], Theorem 3.3 p223 and its proof, for  $r \in [s, T]$ ,  $x \mapsto \tilde{X}_r^{s,x}$  is twice continuously differentiable  $\mathbb{P}$ -almost surely with derivatives satisfying for  $j, k \in \{1, \dots, n\}$

$$\begin{aligned} d\partial_j \tilde{X}_r^{s,x} &= \sum_{l=0}^{\tilde{d}} \partial \tilde{\sigma}^l \left( r, \mathbb{E}[\alpha(X_r)], \tilde{X}_r^{s,x} \right) \partial_j \tilde{X}_r^{s,x} d\tilde{W}_r^l, \quad r \in [s, T], \quad \partial_j \tilde{X}_s^{s,x} = e_j \\ d\partial_{jk} \tilde{X}_r^{s,x} &= \sum_{l=0}^{\tilde{d}} \left( \partial \tilde{\sigma}^l \left( r, \mathbb{E}[\alpha(X_r)], \tilde{X}_r^{s,x} \right) \partial_{jk} \tilde{X}_r^{s,x} + \partial^2 \tilde{\sigma}^l \left( r, \mathbb{E}[\alpha(X_r)], \tilde{X}_r^{s,x} \right) \partial_j \tilde{X}_r^{s,x} \partial_k \tilde{X}_r^{s,x} \right) d\tilde{W}_r^l, \quad r \in [s, T], \end{aligned}$$

where  $\partial_{jk} \tilde{X}_s^{s,x} = 0$ . Moreover, for any  $q \in [1, \infty)$ ,

$$\sup_{0 \leq s \leq r \leq T} \sup_{x \neq y \in \mathbb{R}^n} \sum_{j=1}^n \mathbb{E} \left[ \left| \partial_j \tilde{X}_r^{s,x} \right|^q + \sum_{k=1}^n \left( \left| \partial_{jk} \tilde{X}_r^{s,x} \right|^q + \frac{\left| \partial_{jk} \tilde{X}_r^{s,x} - \partial_{jk} \tilde{X}_r^{s,y} \right|^q}{|x - y|^q} \right) \right] < +\infty$$

and  $\exists C < +\infty, \forall x \in \mathbb{R}^n, \forall 0 \leq s \leq \tilde{s} \leq r \leq T$ ,

$$\mathbb{E} \left[ \left| \tilde{X}_r^{s,x} - \tilde{X}_r^{\tilde{s},x} \right|^q + \sum_{j=1}^n \left( \left| \partial_j \tilde{X}_r^{s,x} - \partial_j \tilde{X}_r^{\tilde{s},x} \right|^q + \sum_{k=1}^n \left| \partial_{jk} \tilde{X}_r^{s,x} - \partial_{jk} \tilde{X}_r^{\tilde{s},x} \right|^q \right) \right] \leq C(1 + |x|^q) (\tilde{s} - s)^{q/2}.$$

These properties ensure that  $x \mapsto u_{t,\psi}(s, x)$  is two times continuously differentiable with derivatives:

- $\partial_j u_{t,\psi}(s, x) = \mathbb{E} \left[ \nabla \psi \left( \tilde{X}_t^{s,x} \right) \cdot \partial_j \tilde{X}_t^{s,x} \right]$
- $\partial_{jk} u_{t,\psi}(s, x) = \mathbb{E} \left[ \nabla \psi \left( \tilde{X}_t^{s,x} \right) \cdot \partial_{jk} \tilde{X}_t^{s,x} + \partial_k \tilde{X}_t^{s,x} \cdot \nabla^2 \psi \left( \tilde{X}_t^{s,x} \right) \partial_j \tilde{X}_t^{s,x} \right]$

bounded and Lipschitz continuous in  $x$  uniformly in  $0 \leq s \leq t \leq T$ . Moreover,  $u_{t,\psi}$  and its spatial derivatives up to the order two are continuous in the time variable. Let us now check that  $u_{t,\psi}$  satisfies the Feynman-Kac PDE (3.2.5). Let  $s \in (0, t]$  and  $\varepsilon \in (0, s]$ . By the flow property stated in Theorem 3.3

[62],  $u_{t,\psi}(s - \varepsilon, x) = \mathbb{E} \left[ u_{t,\psi}(s, \tilde{X}_s^{s-\varepsilon, x}) \right]$ . On the other hand, by Taylor expansion,

$$\begin{aligned} & u_{t,\psi}(s, \tilde{X}_s^{s-\varepsilon, x}) - u_{t,\psi}(s, x) \\ &= \nabla_x u_{t,\psi}(s, x) \cdot \left( \int_{s-\varepsilon}^s \tilde{\sigma}(r, \mathbb{E}[\alpha(X_r)], \tilde{X}_r^{s-\varepsilon, x}) d\tilde{W}_r + \int_{s-\varepsilon}^s b(r, \mathbb{E}[\alpha(X_r)], \tilde{X}_r^{s-\varepsilon, x}) dr \right) \\ &+ \frac{1}{2} \sum_{j,k=1}^n \partial_{jk} u_{t,\psi}(s, x) \left( \sum_{l=0}^{\tilde{d}} \int_{s-\varepsilon}^s \tilde{\sigma}_j^l(r, \mathbb{E}[\alpha(X_r)], \tilde{X}_r^{s-\varepsilon, x}) d\tilde{W}_r^l \right) \left( \sum_{m=0}^{\tilde{d}} \int_{s-\varepsilon}^s \tilde{\sigma}_k^m(r, \mathbb{E}[\alpha(X_r)], \tilde{X}_r^{s-\varepsilon, x}) d\tilde{W}_r^m \right) \\ &+ R_\varepsilon \end{aligned}$$

where  $R_\varepsilon$  is some random reminder such that  $|R_\varepsilon| \leq C \left| \tilde{X}_s^{s-\varepsilon, x} - x \right|^3$  with  $C$  a deterministic finite constant. Since  $\mathbb{E} \left[ \int_{s-\varepsilon}^s \tilde{\sigma} \left( r, \mathbb{E}[\alpha(X_r)], \tilde{X}_r^{s-\varepsilon, x} \right) d\tilde{W}_r \right] = 0$  and, by Itô's isometry,

$$\begin{aligned} & \mathbb{E} \left[ \left( \sum_{l=1}^{\tilde{d}} \int_{s-\varepsilon}^s \tilde{\sigma}_j^l \left( r, \mathbb{E}[\alpha(X_r)], \tilde{X}_r^{s-\varepsilon, x} \right) d\tilde{W}_r^l \right) \left( \sum_{m=1}^{\tilde{d}} \int_{s-\varepsilon}^s \tilde{\sigma}_k^m \left( r, \mathbb{E}[\alpha(X_r)], \tilde{X}_r^{s-\varepsilon, x} \right) d\tilde{W}_r^m \right) \right] \\ &= \mathbb{E} \left[ \int_{s-\varepsilon}^s a_{jk} \left( r, \mathbb{E}[\alpha(X_r)], \tilde{X}_r^{s-\varepsilon, x} \right) dr \right], \end{aligned}$$

we deduce that

$$\begin{aligned} & \left| \frac{u_{t,\psi}(s, x) - u_{t,\psi}(s - \varepsilon, x)}{\varepsilon} + \nabla_x u_{t,\psi}(s, x) \cdot b(s, \mathbb{E}[\alpha(X_s)], x) + \frac{1}{2} \sum_{j,k=1}^n \partial_{jk} u_{t,\psi}(s, x) a_{jk}(s, \mathbb{E}[\alpha(X_s)], x) \right| \\ & \leq C \left( \left| b(s, \mathbb{E}[\alpha(X_s)], x) - \frac{1}{\varepsilon} \mathbb{E} \left[ \int_{s-\varepsilon}^s b \left( r, \mathbb{E}[\alpha(X_r)], \tilde{X}_r^{s-\varepsilon, x} \right) dr \right] \right| \right. \\ & \quad + \left| a(s, \mathbb{E}[\alpha(X_s)], x) - \frac{1}{\varepsilon} \mathbb{E} \left[ \int_{s-\varepsilon}^s a \left( r, \mathbb{E}[\alpha(X_r)], \tilde{X}_r^{s-\varepsilon, x} \right) dr \right] \right| + \frac{\mathbb{E} \left[ \left| \tilde{X}_s^{s-\varepsilon, x} - x \right|^3 \right]}{\varepsilon} \\ & \quad \left. + \frac{1}{\varepsilon} \sum_{j,k=1}^n \mathbb{E} \left[ \left| \int_{s-\varepsilon}^s b_j \left( r, \mathbb{E}[\alpha(X_r)], \tilde{X}_r^{s-\varepsilon, x} \right) dr \sum_{m=0}^{\tilde{d}} \int_{s-\varepsilon}^s \tilde{\sigma}_k^m \left( r, \mathbb{E}[\alpha(X_r)], \tilde{X}_r^{s-\varepsilon, x} \right) d\tilde{W}_r^m \right| \right] \right). \end{aligned}$$

By continuity and uniform integrability of  $r \mapsto [\tilde{\sigma}, b] \left( r, \mathbb{E}[\alpha(X_r)], \tilde{X}_r^{s-\varepsilon, x} \right)$  on  $[s - \varepsilon, T]$ , the two first terms in the right-hand side converge to 0 as  $\varepsilon \rightarrow 0$ . Moreover the expectations in the two last terms are smaller than  $C\varepsilon^{3/2}$ . Taking into account the continuity with respect to  $s$  of

$$\nabla_x u_{t,\psi}(s, x) \cdot b(s, \mathbb{E}[\alpha(X_s)], x) + \frac{1}{2} \sum_{j,k=1}^n \partial_{jk} u_{t,\psi}(s, x) a_{jk}(s, \mathbb{E}[\alpha(X_s)], x),$$

we conclude that  $u_{t,\psi}$  is a classical solution of the Feynman-Kac PDE (3.2.5).  $\square$

We are now ready to prove Theorem 3.2.2.

*Proof of Theorem 3.2.2.* Let  $t \in [0, T]$ . Applying Itô's formula and taking into account the Feynman-Kac

PDE (3.2.5), we obtain that:

$$\begin{aligned} u_{t,\psi}(t, X_t^1) - u_{t,\psi}(t, X_t^{1,N,h}) &= u_{t,\psi}(0, X_0^1) - u_{t,\psi}(0, X_0^1) + \int_0^t \nabla_x u_{t,\psi}(s, X_s^1) \cdot \sigma(s, \mathbb{E}[\alpha(X_s)], X_s^1) dW_s^1 \\ &\quad - \int_0^t \nabla_x u_{t,\psi}(s, X_s^{1,N,h}) \cdot \sigma\left(\tau_s^h, \frac{1}{N} \sum_{i=1}^N \alpha(X_{\tau_s^h}^{i,N,h}), X_{\tau_s^h}^{1,N,h}\right) dW_s^1 \\ &\quad + \int_0^t \nabla_x u_{t,\psi}(s, X_s^{1,N,h}) \cdot \left( b(s, \mathbb{E}[\alpha(X_s)], X_s^{1,N,h}) - b\left(\tau_s^h, \frac{1}{N} \sum_{i=1}^N \alpha(X_{\tau_s^h}^{i,N,h}), X_{\tau_s^h}^{1,N,h}\right) \right) ds \\ &\quad + \frac{1}{2} \sum_{j,k=1}^n \int_0^t \partial_{jk} u_{t,\psi}(s, X_s^{1,N,h}) \left( a_{jk}(s, \mathbb{E}[\alpha(X_s)], X_s^{1,N,h}) - a_{jk}\left(\tau_s^h, \frac{1}{N} \sum_{i=1}^N \alpha(X_{\tau_s^h}^{i,N,h}), X_{\tau_s^h}^{1,N,h}\right) \right) ds. \end{aligned}$$

Integrability deduced from Propositions 3.2.1 and 3.2.3 and the properties of  $\sigma$  ensure that the expectations of the stochastic integrals vanish. Therefore, setting for  $s \in [0, t]$

$$\begin{aligned} e^\psi(s) &:= \mathbb{E} \left[ \nabla_x u_{t,\psi}(s, X_s^{1,N,h}) \cdot \left( b(s, \mathbb{E}[\alpha(X_s)], X_s^{1,N,h}) - b\left(\tau_s^h, \frac{1}{N} \sum_{i=1}^N \alpha(X_{\tau_s^h}^{i,N,h}), X_{\tau_s^h}^{1,N,h}\right) \right) \right] \\ e_{jk}^\psi(s) &:= \mathbb{E} \left[ \partial_{jk} u_{t,\psi}(s, X_s^{1,N,h}) \left( a_{jk}(s, \mathbb{E}[\alpha(X_s)], X_s^{1,N,h}) - a_{jk}\left(\tau_s^h, \frac{1}{N} \sum_{i=1}^N \alpha(X_{\tau_s^h}^{i,N,h}), X_{\tau_s^h}^{1,N,h}\right) \right) \right], \end{aligned}$$

for  $1 \leq j, k \leq n$  and using that  $u_{t,\psi}(t, \cdot) = \psi(\cdot)$ , we obtain

$$\mathbb{E}[\psi(X_t)] - \mathbb{E}[\psi(X_t^{1,N,h})] = \int_0^t \left( e^\psi(s) + \frac{1}{2} \sum_{j,k=1}^n e_{jk}^\psi(s) \right) ds. \quad (3.2.6)$$

Let us now estimate  $e_{jk}^\psi(s)$  :

$$\begin{aligned} e_{jk}^\psi(s) &= \mathbb{E} \left[ \partial_{jk} u_{t,\psi}(s, X_s^{1,N,h}) (a_{jk}(s, \mathbb{E}[\alpha(X_s)], X_s^{1,N,h}) - a_{jk}(\tau_s^h, \mathbb{E}[\alpha(X_{\tau_s^h})], X_s^{1,N,h})) \right] \\ &\quad + \mathbb{E} \left[ \left( \partial_{jk} u_{t,\psi}(s, X_s^{1,N,h}) - \partial_{jk} u_{t,\psi}(s, X_{\tau_s^h}^{1,N,h}) \right) \left( a_{jk}(\tau_s^h, \mathbb{E}[\alpha(X_{\tau_s^h})], X_s^{1,N,h}) \right. \right. \\ &\quad \quad \left. \left. - a_{jk}(\tau_s^h, \mathbb{E}[\alpha(X_{\tau_s^h})], X_{\tau_s^h}^{1,N,h}) \right) \right] \\ &\quad + \mathbb{E} \left[ \partial_{jk} u_{t,\psi}(s, X_{\tau_s^h}^{1,N,h}) (a_{jk}(\tau_s^h, \mathbb{E}[\alpha(X_{\tau_s^h})], X_s^{1,N,h}) - a_{jk}(\tau_s^h, \mathbb{E}[\alpha(X_{\tau_s^h})], X_{\tau_s^h}^{1,N,h})) \right] \\ &\quad + \mathbb{E} \left[ \partial_{jk} u_{t,\psi}(s, X_s^{1,N,h}) (a_{jk}(\tau_s^h, \mathbb{E}[\alpha(X_{\tau_s^h})], X_{\tau_s^h}^{1,N,h}) - a_{jk}(\tau_s^h, \mathbb{E}[\alpha(X_{\tau_s^h})], X_{\tau_s^h}^{1,N,h})) \right] \\ &\quad + \mathbb{E} \left[ \partial_{jk} u_{t,\psi}(s, X_s^{1,N,h}) \left( a_{jk}(\tau_s^h, \mathbb{E}[\alpha(X_{\tau_s^h}^{1,N,h})], X_{\tau_s^h}^{1,N,h}) - a_{jk}\left(\tau_s^h, \frac{1}{N} \sum_{i=1}^N \alpha(X_{\tau_s^h}^{i,N,h}), X_{\tau_s^h}^{1,N,h}\right) \right) \right]. \end{aligned}$$

Let us respectively denote by  $e_{jk}^{1,\psi}(s)$ ,  $e_{jk}^{2,\psi}(s)$ ,  $e_{jk}^{3,\psi}(s)$ ,  $e_{jk}^{4,\psi}(s)$  and  $\bar{e}_{jk}^\psi(s)$  the five terms in the right-hand side. Since, by Proposition 3.2.3,  $\partial_{jk} u_{t,\psi}$  is bounded by a finite constant  $M_\psi^{(2)}$  not depending in  $t$  and  $a$  is Lipschitz continuous with constant  $L_a$ ,  $|e_{jk}^{1,\psi}(s)| \leq M_\psi^{(2)} L_a ((s - \tau_s^h) + |\mathbb{E}[\alpha(X_s)] - \mathbb{E}[\alpha(X_{\tau_s^h})]|)$ . Since  $\alpha$  is  $\mathcal{C}^2$  with bounded derivatives and Lipschitz continuous second order derivatives, for  $0 \leq r \leq s \leq T$ , computing  $\alpha(X_s) - \alpha(X_r)$  by Itô's formula, taking expectations and remarking that the expectation of the stochastic integral vanishes, we obtain the existence of a finite constant  $C$  not depending on  $r$  and  $s$  such that  $|\mathbb{E}[\alpha(X_s)] - \mathbb{E}[\alpha(X_r)]| \leq C(s - r)$ . Hence:

$$|e_{jk}^{1,\psi}(s)| \leq M_\psi^{(2)} L_a (1 + C) (s - \tau_s^h) \leq M_\psi^{(2)} L_a (1 + C) h.$$

Since, by Proposition 3.2.3,  $\partial_{jk} u_{t,\psi}$  is Lipschitz continuous in space with constant  $L_\psi^{(2)}$  not depending on  $t$  and  $a_{jk}$  is Lipschitz continuous in space with constant  $L_a$  we have  $|e_{jk}^{2,\psi}(s)| \leq L_\psi^{(2)} L_a \mathbb{E} \left[ |X_s^{1,N,h} - X_{\tau_s^h}^{1,N,h}|^2 \right]$ .

Combining the fact that  $\sup_{N \geq 1, h \geq 0, s \in [0, T]} \mathbb{E} \left[ |X_s^{1, N, h}|^2 \right] < +\infty$  deduced from Proposition 3.2.1, the Lipschitz continuity of  $\alpha$  and the affine growth of  $\sigma, b$  in their  $p + n$  last variables uniform in their first variable, one easily checks that there is a finite constant  $C$  not depending on  $N$  and  $h$  such that for all  $0 \leq r \leq s \leq T$ ,  $\mathbb{E} \left[ |X_s^{1, N, h} - X_r^{1, N, h}|^2 \right] \leq C(s - r)$  and deduce that:

$$|e_{jk}^{2, \psi}(s)| \leq L_\psi^{(2)} L_a C h.$$

Remarking that the first and second order derivatives of  $a_{jk}$  with respect to its  $n$  last variables have affine growth under our assumptions, computing  $a_{jk}(\tau_s^h, \mathbb{E}[\alpha(X_{\tau_s^h})], X_s^{1, N, h}) - a_{jk}(\tau_s^h, \mathbb{E}[\alpha(X_{\tau_s^h})], X_{\tau_s^h}^{1, N, h})$  by Itô's formula, multiplying by the random variable  $\partial_{jk} u_{t, \psi}(s, X_{\tau_s^h}^{1, N, h})$  bounded by  $M_\psi^{(2)}$ , taking expectations and remarking that the contribution of the stochastic integral vanishes, we obtain that:

$$|e_{jk}^{3, \psi}(s)| \leq C h$$

with  $C$  not depending on  $N, s, t, h$ . Last,

$$|e_{jk}^{4, \psi}(s)| \leq M_\psi^{(2)} L_a \left| \mathbb{E}[\alpha(X_{\tau_s^h})] - \mathbb{E}[\alpha(X_{\tau_s^h}^{1, N, h})] \right|.$$

Hence there is a finite constant  $C$  not depending on  $N, s, t, h$  such that

$$\forall 0 \leq s \leq t \leq T, \quad |e_{jk}^\psi(s)| \leq C \left( h + \left| \mathbb{E}[\alpha(X_{\tau_s^h})] - \mathbb{E}[\alpha(X_{\tau_s^h}^{1, N, h})] \right| \right) + |\bar{e}_{jk}^\psi(s)|. \quad (3.2.7)$$

Let us now estimate  $|\bar{e}_{jk}^\psi(s)|$ . Denoting by  $\nabla_2 a_{jk}$  the partial gradient of  $a$  with respect to its variables with indices in  $\{2, \dots, 1 + p\}$ , we have:

$$\begin{aligned} \bar{e}_{jk}^\psi(s) &= \mathbb{E} \left[ \partial_{jk} u_{t, \psi}(s, X_s^{1, N, h}) \left( \mathbb{E}[\alpha(X_{\tau_s^h}^{1, N, h})] - \frac{1}{N} \sum_{i=1}^N \alpha(X_{\tau_s^h}^{i, N, h}) \right) \right. \\ &\quad \cdot \int_0^1 \left\{ \nabla_2 a_{jk} \left( \tau_s^h, v \mathbb{E}[\alpha(X_{\tau_s^h}^{1, N, h})] + \frac{1-v}{N} \sum_{i=1}^N \alpha(X_{\tau_s^h}^{i, N, h}), X_{\tau_s^h}^{1, N, h} \right) - \nabla_2 a_{jk} \left( \tau_s^h, \mathbb{E}[\alpha(X_{\tau_s^h}^{1, N, h})], X_{\tau_s^h}^{1, N, h} \right) \right\} dv \\ &\quad \left. + \mathbb{E} \left[ \left( \mathbb{E}[\alpha(X_{\tau_s^h}^{1, N, h})] - \frac{1}{N} \sum_{i=1}^N \alpha(X_{\tau_s^h}^{i, N, h}) \right) \cdot \partial_{jk} u_{t, \psi}(s, X_s^{1, N, h}) \nabla_2 a_{jk} \left( \tau_s^h, \mathbb{E}[\alpha(X_{\tau_s^h}^{1, N, h})], X_{\tau_s^h}^{1, N, h} \right) \right] \right]. \end{aligned}$$

Let us respectively denote by  $e_{jk}^{5, \psi}(s)$  and  $e_{jk}^{6, \psi}(s)$  the two terms in the right-hand side. By Proposition 3.2.3,  $\partial_{jk} u_{t, \psi}$  is bounded by  $M_\psi^{(2)}$  and Lipschitz continuous in space with constant  $L_\psi^{(2)}$  with  $(M_\psi^{(2)}, L_\psi^{(2)})$  not depending on  $t$ . By assumption  $\nabla_2 a_{jk}$  is bounded by  $M_a^{(2)}$  and Lipschitz continuous in its  $p + n$  last variables with constant  $L_a^{(2)}$ . Therefore,

$$\begin{aligned} \frac{1}{M_\psi^{(2)} L_a^{(2)}} |e_{jk}^{5, \psi}(s)| &\leq \frac{1}{2} \mathbb{E} \left[ \left| \frac{1}{N} \sum_{i=1}^N \alpha(X_{\tau_s^h}^{i, N, h}) - \mathbb{E}[\alpha(X_{\tau_s^h}^{1, N, h})] \right|^2 \right] \\ &\leq \mathbb{E} \left[ \left| \frac{1}{N} \sum_{i=1}^N (\alpha(X_{\tau_s^h}^{i, N, h}) - \alpha(X_{\tau_s^h}^{i, h})) - \mathbb{E}[\alpha(X_{\tau_s^h}^{1, N, h}) - \alpha(X_{\tau_s^h}^{1, h})] \right|^2 \right] + \mathbb{E} \left[ \left| \frac{1}{N} \sum_{i=1}^N \alpha(X_{\tau_s^h}^{i, h}) - \mathbb{E}[\alpha(X_{\tau_s^h}^{1, h})] \right|^2 \right]. \end{aligned}$$

The second term in the right-hand side of the inequality is the variance of the empirical mean of i.i.d. random variables. It is therefore equal to  $\frac{1}{N} \left( \mathbb{E} \left[ |\alpha(X_{\tau_s^h}^{1, h})|^2 \right] - \left| \mathbb{E}[\alpha(X_{\tau_s^h}^{1, h})] \right|^2 \right)$ . The first term is also a variance and we upper-bound it by the corresponding second order moment, which is not greater than  $\frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[ |\alpha(X_{\tau_s^h}^{i, h}) - \alpha(X_{\tau_s^h}^{i, N, h})|^2 \right]$  according to Jensen's inequality for the empirical mean. Therefore,

denoting by  $L_\alpha$  the Lipschitz constant of  $\alpha$ , we have:

$$\begin{aligned} |e_{jk}^{5,\psi}(s)| &\leq M_\psi^{(2)} L_a^{(2)} \left( \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[ \left| \alpha(X_{\tau_s^h}^{i,h}) - \alpha(X_{\tau_s^h}^{1,N,h}) \right|^2 \right] + \frac{1}{N} \left( \mathbb{E} \left[ \left| \alpha(X_{\tau_s^h}^{1,h}) \right|^2 \right] - \left| \mathbb{E} \left[ \alpha(X_{\tau_s^h}^{1,h}) \right] \right|^2 \right) \right) \\ &\leq L_\alpha^2 M_\psi^{(2)} L_a^{(2)} \frac{C}{N}, \end{aligned} \quad (3.2.8)$$

with  $C$  finite and not depending on  $N, s, t$  according to Proposition 3.2.1.

Since  $\mathbb{E} \left[ \alpha(X_{\tau_s^h}^{1,N,h}) \right] - \frac{1}{N} \sum_{i=1}^N \alpha(X_{\tau_s^h}^{i,N,h})$  is centered and by exchangeability of  $(X_{\tau_s^h}^{i,N,h})_{1 \leq i \leq N}$ , we may replace  $\partial_{jk} u_{t,\psi}(s, X_s^{1,N,h}) \nabla_2 a_{jk}(\tau_s^h, \mathbb{E}[\alpha(X_{\tau_s^h}^{1,N,h})], X_{\tau_s^h}^{1,N,h})$  by:

$$\begin{aligned} &\frac{1}{N} \sum_{i=1}^N \partial_{jk} u_{t,\psi}(s, X_s^{i,N,h}) \nabla_2 a_{jk}(\tau_s^h, \mathbb{E}[\alpha(X_{\tau_s^h}^{1,N,h})], X_{\tau_s^h}^{i,N,h}) \\ &\quad - \mathbb{E} \left[ \partial_{jk} u_{t,\psi}(s, X_s^{1,N,h}) \nabla_2 a_{jk}(\tau_s^h, \mathbb{E}[\alpha(X_{\tau_s^h}^{1,N,h})], X_{\tau_s^h}^{1,N,h}) \right] \end{aligned}$$

in the expectation defining  $e_{jk}^{6,\psi}(s)$ . With the Cauchy-Schwarz inequality, we deduce that  $|e_{jk}^{6,\psi}(s)|$  is upper-bounded by:

$$\begin{aligned} &\left( \mathbb{E} \left[ \left| \frac{1}{N} \sum_{i=1}^N \alpha(X_{\tau_s^h}^{i,N,h}) - \mathbb{E}[\alpha(X_{\tau_s^h}^{1,N,h})] \right|^2 \right] \right)^{1/2} \\ &\times \left( \mathbb{E} \left[ \left| \frac{1}{N} \sum_{i=1}^N \partial_{jk} u_{t,\psi}(s, X_s^{i,N,h}) \nabla_2 a_{jk}(\tau_s^h, \mathbb{E}[\alpha(X_{\tau_s^h}^{1,N,h})], X_{\tau_s^h}^{i,N,h}) \right. \right. \right. \\ &\quad \left. \left. \left. - \mathbb{E} \left[ \partial_{jk} u_{t,\psi}(s, X_s^{1,N,h}) \nabla_2 a_{jk}(\tau_s^h, \mathbb{E}[\alpha(X_{\tau_s^h}^{1,N,h})], X_{\tau_s^h}^{1,N,h}) \right] \right|^2 \right] \right)^{1/2}. \end{aligned}$$

Since  $\mathbb{R}^n \ni x \mapsto \partial_{jk} u_{t,\psi}(s, x)$  (resp.  $\mathbb{R}^n \ni x \mapsto \nabla_2 a_{jk}(\tau_s^h, \mathbb{E}[\alpha(X_{\tau_s^h}^{1,N,h})], x)$ ) is bounded by  $M_\psi^{(2)}$  (resp.  $M_a^{(2)}$ ) and Lipschitz continuous with constant  $L_\psi^{(2)}$  (resp.  $L_a^{(2)}$ ), reasoning like in the derivation of (3.2.8), we obtain that the second factor in the right-hand side is smaller than  $(M_\psi^{(2)} L_a^{(2)} + M_a^{(2)} L_\psi^{(2)}) \frac{C}{\sqrt{N}}$  so that  $|e_{jk}^{6,\psi}(s)| \leq L_\alpha (M_\psi^{(2)} L_a^{(2)} + M_a^{(2)} L_\psi^{(2)}) \frac{C}{N}$  with  $C$  finite and not depending on  $N, s, t, \psi$ . With (3.2.7), we deduce that:

$$\forall 0 \leq s \leq t \leq T, \quad |e_{jk}^\psi(s)| \leq C \left( \frac{1}{N} + h + \left| \mathbb{E}[\alpha(X_{\tau_s^h})] - \mathbb{E}[\alpha(X_{\tau_s^h}^{1,N,h})] \right| \right).$$

Estimating  $e^\psi(s)$  in a similar way and using (3.2.6), we deduce the existence of a finite constant  $C$  not depending on  $N$  and  $h$  such that for all  $t \in [0, T]$ :

$$\sup_{s \in [0, t]} |\mathbb{E}[\psi(X_s)] - \mathbb{E}[\psi(X_s^{1,N,h})]| \leq C \left( \frac{1}{N} + h \right) + C \int_0^t \left| \mathbb{E}[\alpha(X_{\tau_s^h})] - \mathbb{E}[\alpha(X_{\tau_s^h}^{1,N,h})] \right| ds. \quad (3.2.9)$$

Summing for  $j \in \{1, \dots, p\}$  this inequality applied with  $\psi$  equal to the  $j$ -th coordinate of  $\alpha$ , remarking that Proposition 3.2.1 and the Lipschitz continuity of  $\alpha$  ensure that  $\sup_{t \in [0, T]} \left| \mathbb{E}[\alpha(X_t^{1,N,h})] - \mathbb{E}[\alpha(X_t)] \right| < +\infty$  and applying Gronwall's lemma, we deduce that  $\sup_{t \in [0, T]} \left| \mathbb{E}[\alpha(X_t)] - \mathbb{E}[\alpha(X_t^{1,N,h})] \right| \leq C \left( \frac{1}{N} + h \right)$ . We conclude by combining this estimation and (3.2.9).  $\square$

**Remark 3.2.4.** • A careful look at the proof shows that the Lipschitz continuity of  $a$  and  $b$  in the time variable is not needed to obtain that  $\sup_{t \in [0, T]} \left| \mathbb{E}[\psi(X_t^{1,N})] - \mathbb{E}[\psi(X_t)] \right| \leq \frac{C}{N}$ . Indeed, this property is only used to estimate  $e_{jk}^{1,\psi}(s)$  which vanishes when  $h = 0$  since  $\tau_s^0 = s$ . Moreover, if  $a$  and  $b$  are only Hölder continuous with exponent  $\alpha \in (0, 1)$  in the time variable, then the above estimation deteriorates to  $\sup_{t \in [0, T]} \left| \mathbb{E}[\psi(X_t^{1,N,h})] - \mathbb{E}[\psi(X_t)] \right| \leq C \left( \frac{1}{N} + h^\alpha \right)$ .



- On the other hand, when there is no nonlinearity in the sense of McKean in the SDE (3.2.1) ( $p = 0$ ), we obtain that the weak order of convergence of the Euler scheme is 1 under mere global Lipschitz continuity of  $a, b$ , continuity of  $\sigma, b$  and  $C^2$  regularity in space of  $\sigma, b, \psi$  with bounded and Lipschitz derivatives. The key step is that the decomposition of  $e_{jk}^{2,\psi}(s) + e_{jk}^{3,\psi}(s)$  avoids to differentiate the solution to the Feynman-Kac partial differential equation more than two times in space and only requires Lipschitz continuity of the second order spatial derivatives.

### 3.3 Numerical Experiments

We conduct two types of numerical tests. First, we estimate the bias using regular Monte-Carlo for examples of one dimensional ( $n = 1$ ) mean-field SDEs taken from [55] to provide numerical evidence of the  $\mathcal{O}(N^{-1})$  behaviour of the bias for a fixed value of the time step  $h$ . Then we present the antithetic sampling results on these same examples.

The code for running these experiments has a number of iterations as an input parameter. This latter is useful to observe the behaviour of the bias when increasing the number of particles. Therefore, we give an initial number of particles that we multiply by two from an iteration to the other. Except for the polynomial drift and the plane rotator examples where it is respectively equal to 8 and 4, the number of iterations chosen is equal to five and the initial number of particles is twenty so that the final number of particles is 320. The simulation is done with  $5 \cdot 10^6$  runs except for the Plane rotator example where we push further the number of Monte Carlo runs up to  $4.9 \times 10^8$ .

We also define the precision as half the width of the 95% confidence interval of the empirical mean i.e.  $Precision = 1.96 \times \sqrt{\frac{Variance}{number\ of\ runs}}$  where  $Variance$  denotes the empirical variance over the runs of the empirical mean over the particles.

In order to measure the relevance of the antithetic sampling technique for variance reduction, we compute the variance of the difference between the empirical mean of the system with  $2N$  particles and the mean of the empirical means of the two independent systems with  $N$  particles:

$$\text{Var} \left[ \frac{1}{2N} \sum_{i=1}^{2N} \psi(X_T^{i,2N}) - \frac{1}{2N} \sum_{i=1}^N \left( \psi(X_T^{i,N}) + \psi(Y_T^{i,N}) \right) \right]. \quad (3.3.1)$$

Here  $(Y_t^{i,N})_{1 \leq i \leq N}$  is the system with  $N$  particles driven by  $(X_0^i, W^i)_{N+1 \leq i \leq 2N}$ .

#### 3.3.1 Generalised Ornstein-Uhlenbeck process

##### Model

We consider the following generalization of the Ornstein-Uhlenbeck SDE to a linear mean-field SDE:

$$dX_t = [\gamma X_t + \beta \mathbb{E}[X_t]] dt + v dW_t, \text{ with } X(0) = x_0$$

for parameters  $\gamma, \beta, v \in \mathbb{R}$  and initial data  $x_0 \in \mathbb{R}$ .

The functions  $\alpha(x) = x$ ,  $b(t, y, x) = \gamma x + \beta y$  and  $\sigma(t, y, x) = v$  therefore satisfy the hypotheses of regularity of Theorem 3.2.2.

The first and second moments of  $X_t$  are respectively given by:

- $\mathbb{E}[X_t] = x_0 \exp((\gamma + \beta)t)$ ,
- $\mathbb{E}[X_t^2] = x_0^2 \exp(2(\gamma + \beta)t) + \frac{v^2}{2\gamma} (\exp(2\gamma t) - 1)$ .



The associated particle approximation of the SDE is given by the following system:

$$dX_t^{i,N} = \left[ \gamma X_t^{i,N} + \frac{\beta}{N} \sum_{j=1}^N X_t^{j,N} \right] dt + v dW_t^i, \quad 1 \leq i \leq N \text{ with } X_0^{i,N} = x_0$$

where  $\left( (W_t^i)_{t \geq 0} \right)_{1 \leq i \leq N}$  are independent Brownian motions.

Because of the linearity of the drift coefficient, it is possible to obtain closed form expressions for  $\mathbb{E}[X_t^{1,N,h}]$  and  $\mathbb{E}[(X_t^{1,N,h})^2]$  and deduce the asymptotic behaviour of the biases of the first and second order moments as  $N \rightarrow \infty$  and  $h \rightarrow 0$ .

Let  $h = T/K$  for some  $K \in \mathbb{N}^*$ . For  $k \in \{0, \dots, K-1\}$  and  $i \in \{1, \dots, N\}$ , we have

$$X_{(k+1)h}^{i,N,h} = (1 + \gamma h) X_{kh}^{i,N,h} + \beta h \bar{X}_{kh}^{N,h} + v \left( W_{(k+1)h}^i - W_{kh}^i \right) \quad \text{with } \bar{X}_{kh}^{N,h} = \frac{1}{N} \sum_{j=1}^N X_{kh}^{j,N,h}.$$

Hence  $\mathbb{E} \left[ X_{kh}^{i,N,h} \right] = (1 + (\gamma + \beta) h)^k x_0$  so that

$$\mathbb{E} [X_T] - \mathbb{E} \left[ X_T^{i,N,h} \right] = \frac{1}{2} (\gamma + \beta)^2 T e^{(\gamma + \beta)T} h x_0 + \mathcal{O}(h^2) \quad \text{as } h \rightarrow 0.$$

To compute  $\mathbb{E} \left[ \left( X_{kh}^{1,N,h} \right)^2 \right]$ , we remark that for  $k \in \{0, \dots, K-1\}$

$$\begin{aligned} \mathbb{E} \left[ \left( X_{(k+1)h}^{1,N,h} \right)^2 \right] &= (1 + \gamma h)^2 \mathbb{E} \left[ \left( X_{kh}^{1,N,h} \right)^2 \right] \\ &\quad + (2(1 + \gamma h) + \beta h) \beta h \left( \frac{1}{N} \mathbb{E} \left[ \left( X_{kh}^{1,N,h} \right)^2 \right] + \frac{N-1}{N} \mathbb{E} \left[ X_{kh}^{1,N,h} X_{kh}^{2,N,h} \right] \right) + v^2 h, \end{aligned}$$

$$\begin{aligned} \mathbb{E} \left[ X_{(k+1)h}^{1,N,h} X_{(k+1)h}^{2,N,h} \right] &= (1 + \gamma h)^2 \mathbb{E} \left[ X_{kh}^{1,N,h} X_{kh}^{2,N,h} \right] \\ &\quad + (2(1 + \gamma h) + \beta h) \beta h \left( \frac{1}{N} \mathbb{E} \left[ \left( X_{kh}^{1,N,h} \right)^2 \right] + \frac{N-1}{N} \mathbb{E} \left[ X_{kh}^{1,N,h} X_{kh}^{2,N,h} \right] \right). \end{aligned}$$

Since  $X_0^{i,N,h} = x_0$  for all  $i \in \{1, \dots, N\}$ , subtracting the two last equations, we obtain

$$\forall k \in \{1, \dots, K\}, \quad \mathbb{E} \left[ \left( X_{kh}^{1,N,h} \right)^2 \right] - \mathbb{E} \left[ X_{kh}^{1,N,h} X_{kh}^{2,N,h} \right] = \frac{(1 + \gamma h)^{2k} - 1}{2\gamma + \gamma^2 h} \times v^2$$

and deduce that

$$\mathbb{E} \left[ \left( X_{(k+1)h}^{1,N,h} \right)^2 \right] = (1 + (\gamma + \beta) h)^2 \mathbb{E} \left[ \left( X_{kh}^{1,N,h} \right)^2 \right] + (2(1 + \gamma h) + \beta h) \beta \frac{1-N}{N} \times \frac{(1 + \gamma h)^{2k} - 1}{2\gamma + \gamma^2 h} \times v^2 h + v^2 h.$$

We conclude that:

$$\begin{aligned} \mathbb{E} \left[ \left( X_{kh}^{1,N,h} \right)^2 \right] &= (1 + (\gamma + \beta) h)^{2k} x_0^2 + \frac{(1 + (\gamma + \beta) h)^{2k} - 1}{2(\gamma + \beta) + (\gamma + \beta)^2 h} \left( 1 + \frac{N-1}{N} \times \frac{2\beta + (2\beta\gamma + \beta^2) h}{2\gamma + \gamma^2 h} \right) v^2 \\ &\quad + \frac{1-N}{N} \times \frac{(1 + (\gamma + \beta) h)^{2k} - (1 + \gamma h)^{2k}}{2\gamma + \gamma^2 h} \times v^2 \\ &= (1 + (\gamma + \beta) h)^{2k} x_0^2 + \frac{N-1}{N} \times \frac{(1 + \gamma h)^{2k} - 1}{2\gamma + \gamma^2 h} \times v^2 + \frac{1}{N} \times \frac{(1 + (\gamma + \beta) h)^{2k} - 1}{2(\gamma + \beta) + (\gamma + \beta)^2 h} \times v^2 \end{aligned}$$

so that

$$\begin{aligned} \mathbb{E} \left[ (X_T)^2 \right] - \mathbb{E} \left[ \left( X_T^{1,N,h} \right)^2 \right] &= \left( (\gamma + \beta)^2 T e^{2(\gamma + \beta)T} x_0^2 + \frac{e^{2\gamma T} - 1 + 2\gamma T e^{2\gamma T}}{4} v^2 \right) h \\ &\quad + \left( \frac{e^{2\gamma T} - 1}{2\gamma} + \frac{1 - e^{2(\gamma + \beta)T}}{2(\gamma + \beta)} \right) \frac{v^2}{N} + \mathcal{O} \left( h^2 + \frac{h}{N} + \frac{1}{N^2} \right) \end{aligned}$$

as  $h \rightarrow 0$  and  $N \rightarrow \infty$ . Moreover,

$$\mathbb{E} \left[ (X_T^h)^2 \right] - \mathbb{E} \left[ \left( X_T^{1,N,h} \right)^2 \right] = \frac{1}{N} \left( \frac{(1 + \gamma h)^{2k} - 1}{2\gamma + \gamma^2 h} + \frac{1 - (1 + (\gamma + \beta)h)^{2k}}{2(\gamma + \beta) + (\gamma + \beta)^2 h} \right) \times v^2.$$

The bias of the time discretized second order moment is exactly of order 1 in  $\frac{1}{N}$ .

### Results

In order to illustrate the behaviour of the first and second order moments, we compute the difference between the closed-form discretized moments and the estimated moments and expect the difference to be null up to the statistical error. The results are shown in Tables 3.1 and 3.2 below where we denote by "Difference" that entity. As a test case, we use this model with  $\gamma = \frac{1}{2}$ ,  $\beta = \frac{4}{5}$ ,  $v^2 = \frac{1}{2}$  and  $x_0 = 1$ .

Concerning the first order moment, we proved above that the bias does not depend on the number  $N$  of particles, which is what is observe in the first row of the table below.

| Nb. particles                    | 20      | 40      | 80       | 160      | 320      |
|----------------------------------|---------|---------|----------|----------|----------|
| Estimated 1 <sup>st</sup> moment | 1.34862 | 1.34861 | 1.34869  | 1.34866  | 1.34866  |
| Difference                       | 0.00003 | 0.00004 | -0.00004 | -0.00001 | -0.00001 |
| Precision                        | 0.00016 | 0.00011 | 0.00008  | 0.00006  | 0.00004  |

Table 3.1 – Generalised Ornstein-Uhlenbeck SDE: Comparison of the estimated first moments with the closed-form discretized value 1.34865 as well as the associated precision when increasing the number of particles for a number of  $5 \cdot 10^6$  runs, 50 time steps and a time horizon  $T=1$ .

As for the second order moment, from the third row we observe that the estimation fits the closed-form discretized value which confirms the behaviour of the bias of order 1 in  $\frac{1}{N}$ .

| Nb. particles                                  | 20      | 40       | 80       | 160      | 320      |
|------------------------------------------------|---------|----------|----------|----------|----------|
| Closed-form discretized 2 <sup>nd</sup> moment | 2.15552 | 2.14648  | 2.14195  | 2.13969  | 2.13856  |
| Estimated 2 <sup>nd</sup> moment               | 2.15531 | 2.14655  | 2.14205  | 2.13970  | 2.13859  |
| Difference                                     | 0.00021 | -0.00007 | -0.00010 | -0.00001 | -0.00003 |
| Precision                                      | 0.00045 | 0.00032  | 0.00022  | 0.00016  | 0.00011  |

Table 3.2 – Generalised Ornstein-Uhlenbeck SDE: Comparison of the estimated second moments with their closed-form discretized values as well as the associated precision when increasing the number of particles for a number of  $5 \cdot 10^6$  runs, 50 time steps and a time horizon  $T=1$ .

Concerning the antithetic sampling of the first order moment, using the linearity of the model, one exactly checks that for  $k \in \{0, \dots, K-1\}$ ,  $\bar{X}_{kh}^{2N,h} = \frac{1}{2} \left( \bar{X}_{kh}^{N,h} + \bar{Y}_{kh}^{N,h} \right)$  where  $\bar{Y}_{kh}^{N,h}$  denotes the empirical mean of the discretized system driven by  $(W^i)_{N+1 \leq i \leq 2N}$ . The first row of Table 3.3 confirms that the variance (3.3.1) for  $\psi(x) = x$  is null.

For the second order moment, we compute the variance (3.3.1) for  $\psi(x) = x^2$  and observe the ratio of decrease  $\frac{\text{Variance}(N/2)}{\text{Variance}(N)}$  when increasing the number of particles. The results are shown in the third and forth rows of Table 3.3.

| Nb. particles                           | 20          | 40          | 80          | 160         | 320         |
|-----------------------------------------|-------------|-------------|-------------|-------------|-------------|
| Variance for the 1 <sup>st</sup> moment | 7.82256e-32 | 9.52151e-32 | 1.52652e-31 | 2.79857e-31 | 5.39704e-31 |
| Precision                               | 9.88245e-35 | 1.18553e-34 | 1.89269e-34 | 3.47065e-34 | 6.69575e-34 |
| Variance for the 2 <sup>nd</sup> moment | 0.000766641 | 0.000191654 | 4.77727e-05 | 1.19847e-05 | 2.99537e-06 |
| Ratio of decrease                       | ×           | 4.00013     | 4.01179     | 3.98613     | 4.00109     |
| Precision                               | 2.4997e-06  | 6.27208e-07 | 1.55659e-07 | 3.9365e-08  | 9.75389e-09 |

Table 3.3 – Generalised Ornstein-Uhlenbeck SDE: Evolution of the antithetic variance for both  $\psi(x) = x$  and  $\psi(x) = x^2$  with its associated precision when increasing the number of particles for a number of  $5.10^6$  runs, 50 time steps and a time horizon of 1.

For the second order moment, the ratio of successive variances is around 4 which means that the variance of the antithetic estimator is roughly proportional to  $N^{-2}$ . The antithetic sampling technique therefore shows an important improvement for this diffusion.

### 3.3.2 Plane Rotator

#### Model

The following SDE refers to a model for coupled oscillators in the presence of noise, also known as the Kuramoto model:

$$\begin{aligned}
 dX_t &= \left[ K \int_{\mathbb{R}} \sin(y - X_t) dP_t^\mu(y) - \sin(X_t) \right] dt + \sqrt{2k_B T} dW_t \\
 &= \left[ K \cos(X_t) \int_{\mathbb{R}} \sin(y) dP_t^\mu - K \sin(X_t) \int_{\mathbb{R}} \cos(y) dP_t^\mu - \sin(X_t) \right] dt + \sqrt{2k_B T} dW_t
 \end{aligned}$$

where  $X_0$  is distributed according to  $\mu$ ,  $P_t^\mu$  denotes the distribution of  $X_t$ ,  $K > 0$  a coupling parameter,  $k_B$  the Boltzmann constant and  $T$  the temperature.

The functions  $\sigma(x) = \sqrt{2k_B T}$  and  $\alpha(x) = \begin{bmatrix} \sin(x) \\ \cos(x) \end{bmatrix}$  satisfy the hypotheses of Theorem 3.2.2. One may also find a function  $b(t, y, x)$  coinciding with  $K(\cos(x)y_1 - \sin(x)y_2) - \sin(x)$  on  $[0, T] \times [-1, 1]^2 \times \mathbb{R}$  which satisfies the hypotheses even if  $(x, y) \rightarrow \cos(x)y_1 - \sin(x)y_2$  is not Lipschitz continuous.

The particle system has the following dynamics:

$$\begin{aligned}
 dX_t^{i,N} &= \left[ K \left( \cos(X_t^{i,N}) \frac{1}{N} \sum_{j=1}^N \sin(X_t^{j,N}) - \sin(X_t^{i,N}) \frac{1}{N} \sum_{j=1}^N \cos(X_t^{j,N}) \right) - \sin(X_t^{i,N}) \right] dt \\
 &\quad + \sqrt{2k_B T} dW_t^i
 \end{aligned}$$

where  $((W_t^i)_{t \geq 0})_{1 \leq i \leq N}$  are independent Brownian motions.

#### Results

We use this model with  $K = 1$ ,  $k_B T = \frac{1}{8}$  and initial distribution  $\mu = \mathcal{N}(\frac{\pi}{4}, \frac{3\pi}{4})$  as a test case. The reference value was computed for  $2.5 \times 10^8$  Monte Carlo runs, 1000 particles and the same input parameters as for the general estimation of the bias (time horizon  $T = 1$  and time step  $h = 50$ ). The value obtained is 0.737576. The results are shown in Table 3.4.

| Nb. particles                | 20          | 40          | 80          | 160         |
|------------------------------|-------------|-------------|-------------|-------------|
| 1 <sup>st</sup> moment error | 0.000725    | 0.000355    | 0.000175    | 0.000067    |
| Ratio of decrease            | ×           | 2.04225     | 2.02857     | 2.61194     |
| Precision                    | 4.79156e-05 | 3.38946e-05 | 2.39723e-05 | 1.69533e-05 |

Table 3.4 – Plane Rotator SDE: Evolution of the first order moment errors as well as the associated precision when increasing the number of particles for a number of  $4.9 \times 10^8$  runs, 50 time steps and a time horizon  $T=1$ .

These results are consistent with a bias proportional to  $N^{-1}$ .

Table 3.5 exposes the results obtained for the antithetic variance.

| Nb. particles     | 20          | 40          | 80          | 160        |
|-------------------|-------------|-------------|-------------|------------|
| Variance          | 0.000119023 | 3.12526e-05 | 8.01086e-06 | 2.0277e-06 |
| Ratio of decrease | ×           | 3.80842     | 3.90128     | 3.95071    |
| Precision         | 2.51877e-08 | 6.75418e-09 | 1.75258e-09 | 4.4622e-10 |

Table 3.5 – Plane Rotator SDE: Evolution of the antithetic variance and its associated precision when increasing the number of particles for a number of  $4.9 \times 10^8$  runs, 50 time steps and a time horizon  $T=1$ .

The antithetic sampling method is therefore relevant in this example since the ratio of decrease is close to four which means that the variance of the antithetic estimation is roughly proportional to  $N^{-2}$ . This behaviour has also been observed in Section 3.2.2 [35].

### 3.3.3 Polynomial Drift

#### Model

Let us consider the following mean-field SDE:

$$dX_t = [\gamma X_t + \mathbb{E}[X_t] - X_t \mathbb{E}[X_t^2]] dt + X_t dW_t \quad \text{with } X(0) = x_0$$

for a certain parameter  $\gamma \in \mathbb{R}$  and initial data  $x_0 \in \mathbb{R}$ .

The function  $\sigma(t, y, x) = x$  satisfy the hypotheses of regularity. However, the functions  $b(t, y, x) = \gamma x + y_1 - xy_2$  and  $\alpha(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix}$  are not Lipschitz.

From the evolution of the Euler discretization

$$X_{(k+1)h}^h = X_{kh}^h \left( (1 + \gamma h - h \mathbb{E}[(X_{kh}^h)^2]) \right) + h \mathbb{E}[X_{kh}^h] + X_{kh}^h (W_{(k+1)h} - W_{kh}),$$

we deduce that:

$$\begin{aligned} \mathbb{E}[X_{(k+1)h}^h] &= (1 + (\gamma + 1)h - \mathbb{E}[(X_{kh}^h)^2]h) \mathbb{E}[X_{kh}^h], \\ \mathbb{E}[(X_{(k+1)h}^h)^2] &= (1 + \gamma h - \mathbb{E}[(X_{kh}^h)^2]h)^2 \mathbb{E}[(X_{kh}^h)^2] + 2(1 + \gamma h - \mathbb{E}[(X_{kh}^h)^2]h) h \mathbb{E}^2[X_{kh}^h] \\ &\quad + \mathbb{E}[(X_{kh}^h)^2]h + \mathbb{E}^2[X_{kh}^h]h^2. \end{aligned}$$

And we solve this system of inductive equations numerically to obtain reference values of the first and second order moments.

The idea here, once again, is to approximate  $\mathbb{E}[X]$  and  $\mathbb{E}[X^2]$  by their empirical means in order to define the dynamics of the particle system:

$$dX_t^{i,N} = \left[ \gamma X_t^{i,N} + \frac{1}{N} \sum_{j=1}^N X_t^{j,N} - X_t^{i,N} \frac{1}{N} \sum_{j=1}^N \left( X_t^{j,N} \right)^2 \right] dt + X_t^{i,N} dW_t^i$$

where  $\left( (W_t^i)_{t \geq 0} \right)_{1 \leq i \leq N}$  are independent Brownian motions.

## Results

We use this model with  $\gamma = 2$  and  $x = 1$  as a test case. The reference values obtained for the time discretized first and second moments are respectively 1.3845 and 3.13743 . For this example, we push the iterations further until 8 so that the final number of particles is 2560. We denote by "Ratio of decrease 1" and "Ratio of decrease 2" the respective entities  $\frac{\text{First moment error}(N/2)}{\text{First moment error}(N)}$  and  $\frac{\text{Second moment error}(N/2)}{\text{Second moment error}(N)}$ . The results are shown in Tables 3.6 and 3.7.

| Nb. particles                | 40        | 80        | 160       | 320       | 640       | 1280      | 2560      |
|------------------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1 <sup>st</sup> moment error | -0.02597  | -0.01561  | -0.00898  | -0.00492  | -0.00261  | -0.00138  | -0.00071  |
| Ratio of dec. 1              | ×         | 1.66368   | 1.73831   | 1.82520   | 1.88506   | 1.89130   | 1.94366   |
| Precision                    | 9.348e-05 | 7.402e-05 | 5.779e-05 | 4.431e-05 | 3.330e-05 | 2.459e-05 | 1.789e-05 |

Table 3.6 – Polynomial Drift SDE: Evolution of the first order moment errors as well as the associated precision when increasing the number of particles for a number of  $5 \cdot 10^6$  runs, 50 time steps and a time horizon  $T=1$

| Nb. particles                | 40        | 80        | 160       | 320       | 640       | 1280      | 2560      |
|------------------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 2 <sup>nd</sup> moment error | 0.06025   | 0.03575   | 0.02077   | 0.01157   | 0.00637   | 0.00333   | 0.00171   |
| Ratio of dec. 2              | ×         | 1.68531   | 1.72123   | 1.79516   | 1.81633   | 1.91291   | 1.94737   |
| Precision                    | 0.0005018 | 0.0003940 | 0.0003029 | 0.0002297 | 0.0001707 | 0.0001251 | 9.063e-05 |

Table 3.7 – Polynomial Drift SDE: Evolution of the second order moment errors as well as the associated precision when increasing the number of particles for a number of  $5 \cdot 10^6$  runs, 50 time steps and a time horizon  $T=1$

We observe that the ratios of decrease of both the first and the second order moment errors seem to grow as the number of particle increases. It confirms the behaviour of the bias in  $\mathcal{O}(\frac{1}{N})$  of Theorem 3.2.2 and even tends towards a behaviour of order 1 in  $\frac{1}{N}$  when the number of particle is large.

In order to measure the relevance of the antithetic sampling, we compute the variance (3.3.1) for the first and second order moments. We denote by "Ratio of decrease V1" and "Ratio of decrease V2" the respective entities  $\frac{\text{Variance for first moment error}(N/2)}{\text{Variance for first moment error}(N)}$  and  $\frac{\text{Variance for Second moment error}(N/2)}{\text{Variance for Second moment error}(N)}$ . The results are shown in Tables 3.8 and 3.9.

| Nb. particles                     | 40        | 80        | 160       | 320       | 640       | 1280      | 2560      |
|-----------------------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Variance for 1 <sup>st</sup> mom. | 0.0008910 | 0.0004406 | 0.0001930 | 7.474e-05 | 2.584e-05 | 8.000e-06 | 2.262e-06 |
| Ratio of dec. V1                  | ×         | 2.023     | 2.283     | 2.582     | 2.892     | 3.230     | 3.537     |
| Precision                         | 3.621e-06 | 2.260e-06 | 1.312e-06 | 6.650e-07 | 3.278e-07 | 1.294e-07 | 7.131e-08 |

Table 3.8 – Polynomial Drift SDE: Evolution of the antithetic variance for  $\psi(x) = x$  with its associated precision when increasing the number of particles for a number of  $5 \cdot 10^6$  runs, 50 time steps and a time horizon  $T=1$ .

| Nb. particles                     | 40       | 80        | 160       | 320       | 640       | 1280      | 2560      |
|-----------------------------------|----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Variance for 2 <sup>nd</sup> mom. | 0.02380  | 0.01078   | 0.00440   | 0.00170   | 0.00058   | 0.00018   | 5.282e-05 |
| Ratio of dec. V2                  | ×        | 2.2074    | 2.45094   | 2.60197   | 2.89779   | 3.20462   | 3.44723   |
| Precision                         | 0.000113 | 6.672e-05 | 3.597e-05 | 2.415e-05 | 1.125e-05 | 5.042e-06 | 2.862e-06 |

Table 3.9 – Polynomial Drift SDE: Evolution of the antithetic variance for  $\psi(x) = x^2$  with its associated precision when increasing the number of particles for a number of  $5 \cdot 10^6$  runs, 50 time steps and a time horizon  $T=1$ .

The results exposed in Tables 3.8 and 3.9 both show that when increasing the number of particles, the ratios of decrease grow gradually. Therefore, the antithetic sampling method may be relevant when simulating with a large number of particles.

### 3.3.4 Viscous Burgers equation

#### Model

Let us consider the following mean-field SDE for a parameter  $v > 0$ :

$$\begin{cases} dX_t = \bar{F}_t(X_t) dt + v dW_t, \\ \bar{F}_t(x) = \mathbb{P}(X_t \geq x), \quad \forall (t, x) \in [0, T] \times \mathbb{R}. \end{cases}$$

Using the Fokker-Planck equation satisfied by the density of  $X_t$ , we show that  $\bar{F}_t(x)$  is solution to the viscous Burgers equation:

$$\partial_t V(t, x) = \frac{v^2}{2} \partial_{xx} V(t, x) - V(t, x) \partial_x V(t, x).$$

We now suppose that the initial condition is  $X(0) = 0$  so that  $\bar{F}_0(x) = \mathbf{1}_{\{x \leq 0\}}$ . Then the Cole-Hopf transformation gives:

$$\bar{F}_t(x) = \frac{\mathcal{N}\left(\frac{t-x}{v\sqrt{t}}\right)}{\exp\left(\frac{2x-t}{2v^2}\right) \mathcal{N}\left(\frac{x}{v\sqrt{t}}\right) + \mathcal{N}\left(\frac{t-x}{v\sqrt{t}}\right)}$$

where  $\mathcal{N}(x) = \int_{-\infty}^x \exp\left(-\frac{y^2}{2}\right) \frac{dy}{\sqrt{2\pi}}$ .

The function  $\sigma(t, y, x) = v$  and  $b(t, y, x) = y$  are regular enough to satisfy the hypotheses of Theorem 3.2.2. However, this type of example is not interacting through moments but through a kernel and even a discontinuous one.

We approximate  $\bar{F}_t(x)$  by its associated empirical mean  $\bar{F}_t^N(x) := \frac{1}{N} \sum_{j=1}^N \mathbf{1}_{\{X_t^{j,N} \geq x\}}$  calculated upon  $N$  particles which leads to the following dynamics for the particle system:

$$dX_t^{i,N} = \frac{1}{N} \sum_{j=1}^N \mathbf{1}_{\{X_t^{j,N} \geq X_t^{i,N}\}} dt + v dW_t^i$$

where  $\left((W_t^i)_{t \geq 0}\right)_{1 \leq i \leq N}$  are independent Brownian motions.

### Results

We use this model with  $v = \frac{1}{4}$  and  $x_0 = 0$  as a test case. We estimate the solution  $\bar{F}_1(\frac{1}{2})$  of the viscous Burgers equation. Since for  $t = 1$  and  $x = \frac{1}{2}$ ,  $t - x = x$  and  $2x - t = 0$ , one has  $\bar{F}_1(\frac{1}{2}) = \frac{1}{2}$ . The results are shown in Table 3.10.

| Nb. particles     | 20          | 40          | 80          | 160         | 320         |
|-------------------|-------------|-------------|-------------|-------------|-------------|
| Solution Error    | 0.0141425   | 0.0070329   | 0.00352185  | 0.00174521  | 0.000870359 |
| Ratio of decrease | $\times$    | 2.01091     | 1.99693     | 2.01802     | 2.00515     |
| Precision         | 9.84069e-05 | 6.94588e-05 | 4.90673e-05 | 3.46603e-05 | 2.45057e-05 |

Table 3.10 – Viscous Burgers equation: Evolution of the solution errors as well as the associated precision when increasing the number of particles for a number of  $5 \cdot 10^6$  runs, 500 time steps and a time horizon  $T=1$ .

We observe that the ratio of decrease of the solution error defined by  $\frac{\text{Solution error}(N/2)}{\text{Solution error}(N)}$  is consistent with an error proportional to  $N^{-1}$  confirming a bias behaviour of order 1 in  $\frac{1}{N}$ .

As for the antithetic sampling, we compute the corresponding variance (3.3.1) as well as the related precision and obtain:

| Nb. particles            | 20          | 40          | 80          | 160         | 320         |
|--------------------------|-------------|-------------|-------------|-------------|-------------|
| Variance of the solution | 0.00286895  | 0.00109648  | 0.000413156 | 0.000155082 | 5.89916e-05 |
| Ratio of decrease        | $\times$    | 2.61651     | 2.65392     | 2.66411     | 2.62889     |
| Precision                | 3.98911e-06 | 1.49543e-06 | 5.52988e-07 | 2.04566e-07 | 7.65869e-08 |

Table 3.11 – Viscous Burgers Equation: Evolution of the antithetic variance and its associated precision when increasing the number of particles for a number of  $5 \cdot 10^6$  runs, 500 time steps and a time horizon  $T=1$ .

From Table 3.11, we observe that the ratio of decrease of the variance  $\frac{\text{Variance of solution}(N/2)}{\text{Variance of solution}(N)}$  is around 2.64 which is slightly greater than two. Therefore, there is a slight gain in using the antithetic sampling for this type of diffusion.





## CHAPTER 4

# WEAK AND STRONG ERROR ANALYSIS FOR MEAN FIELD RANK-BASED PARTICLE APPROXIMATIONS OF ONE DIMENSIONAL VISCOUS SCALAR CONSERVATION LAWS

*Ce chapitre est un article écrit avec Benjamin Jourdain. Une version raccourcie a été soumise à Annals of Applied Probability.*

### Abstract

In this paper, we analyse the rate of convergence of a system of  $N$  interacting particles with mean field rank-based interaction in the drift coefficient and constant diffusion coefficient. We first adapt arguments by Kolli and Shkolnikov [56] to check trajectorial propagation of chaos with optimal rate  $N^{-1/2}$  to the associated stochastic differential equations nonlinear in the sense of McKean. We next relax the assumptions needed by Bossy [8] to check the convergence in  $L^1(\mathbb{R})$  with rate  $\mathcal{O}\left(\frac{1}{\sqrt{N}} + h\right)$  of the empirical cumulative distribution function of the Euler discretization with step  $h$  of the particle system to the solution of a one dimensional viscous scalar conservation law. Last, we prove that the bias of this stochastic particle method behaves as  $\mathcal{O}\left(\frac{1}{N} + h\right)$ . We provide numerical results which confirm our theoretical estimates.

**Keywords:** Propagation of chaos, mean-field interaction, rank-based model, weak-error analysis.

## 4.1 Introduction

The order of weak convergence in terms of the number  $N$  of particles for the approximation of diffusions nonlinear in the sense of McKean solving

$$X_t = X_0 + \int_0^t \varsigma(s, X_s, \mu_s) dW_s + \int_0^t \vartheta(s, X_s, \mu_s) ds \text{ with } \mu_s \text{ denoting the probability distribution of } X_s,$$

by the systems of  $N$  interacting particles

$$\tilde{X}_t^{i,N} = \tilde{X}_0^i + \int_0^t \varsigma(s, \tilde{X}_s^{i,N}, \check{\mu}_s^N) dW_s^i + \int_0^t \vartheta(s, \tilde{X}_s^{i,N}, \check{\mu}_s^N) ds, \quad i \in \{1, \dots, N\} \text{ with } \check{\mu}_s^N = \frac{1}{N} \sum_{i=1}^N \delta_{\tilde{X}_s^{i,N}}, \quad (4.1.1)$$

has been recently investigated in several papers [67, 6, 13, 14]. Here  $(W_t)_{t \geq 0}$  is a  $d$ -dimensional Brownian motion independent from the initial  $\mathbb{R}^n$ -valued random vector  $X_0$ ,  $(W^i, \tilde{X}_0^i)_{i \geq 1}$  are i.i.d. copies of  $(W, X_0)$ ,  $\varsigma : [0, T] \times \mathbb{R}^n \times \mathcal{P}(\mathbb{R}^n) \rightarrow \mathbb{R}^{n \times d}$  and  $\vartheta : [0, T] \times \mathbb{R}^n \times \mathcal{P}(\mathbb{R}^n) \rightarrow \mathbb{R}^n$  with  $\mathcal{P}(\mathbb{R}^n)$  denoting the space of Borel probability distributions on  $\mathbb{R}^n$ . Typically, under some regularity assumptions, the bias is of order  $N^{-1}$  while it is well known since [84] that the strong error is of order  $N^{-1/2}$ . From a numerical perspective, this implies that simulating  $N$  independent copies of the system with  $N$  particles leads to a bias and a statistical error both of order  $N^{-1}$  which is also the order of the global error resulting from one single simulation of the system with  $N^2$  particles. When the computation time of the interaction is quadratic, then the cost of these  $N$  copies is of order  $N^3$  compared to the order  $N^4$  of the computation cost of the system with  $N^2$  particles.

In Theorem 6.1 [67], Mischler, Mouhot and Wennberg prove that for  $\varsigma$  uniformly elliptic and not depending on the time and measure arguments,  $\sup_{t \in [0, T]} \left| \mathbb{E} \left[ \varphi \left( \tilde{X}_t^{1,N} \right) \right] - \int_{\mathbb{R}^n} \varphi(x) \mu_t(dx) \right| \leq \frac{C}{N}$  when  $\varphi$  is Lipschitz and has some Sobolev regularity and  $\vartheta(t, x, \mu) = Ax + \int U(x - y) \mu(dy)$  for some constant matrix  $A$  and some function  $U$  with Sobolev regularity. In [6], we consider the case of interaction through moments:  $\begin{pmatrix} \varsigma \\ \vartheta \end{pmatrix} (s, x, \mu) = \begin{pmatrix} \sigma \\ b \end{pmatrix} \left( s, \int_{\mathbb{R}^n} \alpha(x) \mu(dx), x \right)$ . When  $\alpha : \mathbb{R}^n \rightarrow \mathbb{R}^p$ ,  $\sigma : [0, T] \times \mathbb{R}^p \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times d}$ ,  $b : [0, T] \times \mathbb{R}^p \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  and  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  are twice continuously differentiable with bounded derivatives and Lipschitz second order derivatives and  $\sigma \sigma^*$  is globally Lipschitz, we obtain

$$\exists C < \infty, \forall h \in [0, T], \forall N \in \mathbb{N}^*, \sup_{t \in [0, T]} \left| \mathbb{E} \left[ \varphi \left( \tilde{X}_t^{1,N,h} \right) \right] - \int_{\mathbb{R}^n} \varphi(x) \mu_t(dx) \right| \leq C \left( \frac{1}{N} + h \right)$$

where  $\tilde{X}_t^{i,N,0}$  denotes the particle system (4.1.1) and  $\tilde{X}_t^{i,N,h}$  its Euler discretization with step  $h$  when  $h > 0$ . When  $n = d$ , in Theorem 2.17 [13], Chassagneux, Szpruch and Tse prove the expansion of the bias

$$\mathbb{E} [\Phi(\check{\mu}_T^N)] - \Phi(\mu_T) = \sum_{j=1}^{k-1} \frac{C_j}{N^j} + \mathcal{O} \left( \frac{1}{N^k} \right),$$

for time-homogeneous coefficients  $\varsigma$  and  $\vartheta$ ,  $(2k+1)$ -times differentiable with respect to both the spatial coordinates and the probability measure argument (for the notion of lifted differentiability introduced by Lions in his lectures at the Collège de France) with  $\varsigma$  bounded and  $X_0$  admitting a finite moment of order  $(2k+1)$ . They assume the same regularity on the test function  $\Phi$  on the space of probability measures on  $\mathbb{R}^d$  which is possibly nonlinear:  $\Phi(\mu)$  is not necessarily of the form  $\int_{\mathbb{R}^d} \varphi(x) \mu(dx)$ . In Theorem 3.6 [14], under uniform ellipticity, Chaudru de Raynal and Frikha prove  $|\mathbb{E} [\Phi(\check{\mu}_T^N)] - \Phi(\mu_T)| \leq \frac{C}{N}$  when  $\Phi$  has two bounded and Hölder continuous linear functional derivatives and  $\varsigma \varsigma^*$  and  $\vartheta$  are bounded and globally Hölder continuous with respect to the spatial variables and have two bounded and Hölder continuous linear functional derivatives with respect to the measure argument. Notice that the existence of a linear functional derivative requires less regularity than the Fréchet differentiability of the lift since the lifted derivative is the gradient of the linear functional derivative with respect to the spatial variables.

Our aim in the present paper is to check that the  $\mathcal{O} \left( \frac{1}{N} + h \right)$  behaviour of the weak error for the Euler discretization with step  $h$  of the system with  $N$  particles generalizes to a stochastic differential equation with an even discontinuous drift coefficient. This SDE is one-dimensional ( $n = d = 1$ ) and has a constant

diffusion coefficient  $\varsigma(s, x, \mu) = \sigma$  for  $\sigma > 0$ . The drift coefficient writes  $\vartheta(s, x, \mu) = \lambda(\mu((-\infty, x]))$  where  $\mathbb{R} \times \mathcal{P}(\mathbb{R}) \ni (x, \mu) \mapsto \mu((-\infty, x])$  is not even continuous and  $\lambda$  is the derivative of a  $C^1$  function  $\Lambda : [0, 1] \rightarrow \mathbb{R}$ :

$$\begin{cases} X_t = X_0 + \sigma W_t + \int_0^t \lambda(F(s, X_s)) ds, & t \in [0, T] \\ F(s, x) = \mathbb{P}(X_s \leq x), \quad \forall (s, x) \in [0, T] \times \mathbb{R}. \end{cases} \quad (4.1.2)$$

We denote by  $m$  the probability distribution of  $X_0$  and by  $F_0$  its cumulative distribution function. According to Section 2 in the paper [9] specialized to the case  $\Lambda(u) = u^2/2$  and Proposition 1.2 and Theorem 2.1 [42] for a general function  $\Lambda$ , weak existence and uniqueness hold for the SDE (4.1.2). By [91], it actually admits a unique strong solution. This result of strong uniqueness was later extended to the multi-dimensional case by [87]. For  $t > 0$ , by the Girsanov theorem, the law  $\mu_t$  of  $X_t$  admits a density  $p(t, x)$  with respect to the Lebesgue measure (see Lemma 4.3.1 below). The function  $p(t, x)$  is a weak solution to the Fokker-Planck equation  $\partial_t p(t, x) + \partial_x(\lambda(F(t, x))p(t, x)) = \frac{\sigma^2}{2} \partial_{xx} p(t, x)$ . By integration with respect to the spatial variable  $x$ , we deduce that  $F(t, x)$  is a weak solution to the following viscous conservation law:

$$\begin{cases} \partial_t F(t, x) + \partial_x(\Lambda(F(t, x))) = \frac{\sigma^2}{2} \partial_{xx} F(t, x), \\ F_0(x) = m((-\infty, x]). \end{cases} \quad (4.1.3)$$

The corresponding particle dynamics is

$$\check{X}_t^{i,N} = X_0^i + \sigma W_t^i + \int_0^t \lambda \left( \frac{1}{N} \sum_{j=1}^N \mathbf{1}_{\{\check{X}_s^{j,N} \leq \check{X}_s^{i,N}\}} \right) ds, \quad 1 \leq i \leq N, \quad t \in [0, T]. \quad (4.1.4)$$

As for the initial positions  $(X_0^i)_{i \geq 1}$ , we will consider both cases of the random initialization  $((X_0^i)_{i \geq 1})$  i.i.d. according to  $m$ ) and an optimal deterministic initialization which will be made precise in Section 4.2.

In fact, for  $1 \leq i \leq N$ , the coefficient  $\lambda(i/N)$  is close to

$$\lambda^N(i) = N \left( \Lambda \left( \frac{i}{N} \right) - \Lambda \left( \frac{i-1}{N} \right) \right) \quad (4.1.5)$$

so that the dynamics is close (see Corollary 4.2.2 for a precise statement) to the one introduced in [46] :

$$X_t^{i,N} = X_0^i + \sigma W_t^i + \int_0^t \lambda^N \left( \sum_{j=1}^N \mathbf{1}_{\{X_s^{j,N} \leq X_s^{i,N}\}} \right) ds, \quad 1 \leq i \leq N, \quad t \in [0, T]. \quad (4.1.6)$$

We denote by  $\mu_t^N = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i,N}}$  the empirical measure and by  $F^N(t, x) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{X_t^{i,N} \leq x\}}$  the empirical cumulative distribution function at time  $t$  of this second particle system. Both dynamics are so-called rank-based models since the drift (and the diffusion) coefficient only depend on the rank of the  $i$ -th particle in the system. We call them mean-field rank-based since the interaction between the particles is also of mean-field type. The ability of rank-based models to reproduce stylized empirical properties observed on stock markets [20], has motivated their mathematical study [4]. By the Girsanov theorem, the stochastic differential equations (4.1.4) and (4.1.6) admit a unique weak solution and, according to [87], they actually admit a unique strong solution. Under concavity of  $\Lambda$ , Jourdain and Malrieu [46] prove propagation of chaos with optimal rate  $N^{-1/2}$  and study the long-time behaviour of the particle system (4.1.6) and its mean-field limit (4.1.2). For the particle system (4.1.4), this study is extended by Jourdain and Reygner [50] when the diffusion coefficient is no longer constant but also of mean-field rank-based type. For this more general model and without the concavity assumption, Kolli and Shkolnikov [56] recently proved propagation of chaos with optimal rate  $N^{-1/2}$  and convergence of the associated fluctuations when the initial probability measure  $m$  admits a bounded density with respect to the Lebesgue measure. We choose to focus on the modified dynamics (4.1.6) because when  $y^1 < y^2 < \dots < y^N$ , then the distribution derivative of  $x \mapsto \Lambda \left( \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{y^i \leq x\}} \right)$  is  $x \mapsto \frac{1}{N} \sum_{i=1}^N \lambda^N(i) \mathbf{1}_{\{y^i \leq x\}}$  and not (when  $\Lambda$  is not

affine)  $x \mapsto \frac{1}{N} \sum_{i=1}^N \lambda(i/N) \mathbf{1}_{\{y^i \leq x\}}$ . For this reason, it is more closely connected to the PDE (4.1.3). As our error analysis is based on a comparison of the mild formulation of the PDE (4.1.3) and the perturbed mild formulation satisfied by empirical cumulative distribution function of the Euler discretization of the particle system, we concentrate on (4.1.6), for which no extra error term appears in this perturbed version. But we will also explain how our results extend to (4.1.4). Let us also introduce the Euler discretization with time-step  $h \in (0, T]$  of (4.1.6) :

$$X_t^{i,N,h} = X_0^i + \sigma W_t^i + \int_0^t \lambda^N \left( \sum_{j=1}^N \mathbf{1}_{\{X_{\tau_s^h}^{j,N,h} \leq X_{\tau_s^h}^{i,N,h}\}} \right) ds, \quad 1 \leq i \leq N, \quad t \in [0, T] \text{ where } \tau_s^h = \lfloor s/h \rfloor h. \quad (4.1.7)$$

The empirical cumulative distribution function of  $\mu_t^{N,h} = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i,N,h}}$  is  $F^{N,h}(t, x) := \frac{1}{N} \sum_{j=1}^N \mathbf{1}_{\{X_t^{j,N,h} \leq x\}}$ .

It is natural and convenient to consider that  $\tau_s^0 = s$  and  $(X_t^{i,N,0})_{t \in [0, T], 1 \leq i \leq N} = (X_t^{i,N})_{t \in [0, T], 1 \leq i \leq N}$ .

Using these notations, we then have by convention that  $F^{N,0}(t, x) = F^N(t, x)$ . Moreover, we will refer to the empirical cumulative distribution function  $F_0^{N,h}$  at initialization by  $\tilde{F}_0^N$  when choosing positions that are i.i.d. according to  $m$  and by  $\tilde{F}_0^N$  when choosing optimal deterministic initial positions. Finally let us define  $(\check{X}_t^{i,N,h})_{t \in [0, T], 1 \leq i \leq N}$  like  $(\check{X}_t^{i,N})_{t \in [0, T], 1 \leq i \leq N}$  by replacing  $\lambda^N(k)$  by  $\lambda(k/N)$  in (4.1.7) and set  $\check{\mu}_t^{N,h} = \frac{1}{N} \sum_{i=1}^N \delta_{\check{X}_t^{i,N,h}}$  and  $\check{F}^{N,h}(t, x) := \frac{1}{N} \sum_{j=1}^N \mathbf{1}_{\{\check{X}_t^{j,N,h} \leq x\}}$ .

The paper is organized as follows. In Section 2, we state our results. Taking advantage of the constant diffusion coefficient to adapt arguments given in [56], we obtain propagation of chaos with optimal rate  $N^{-1/2}$  for the non time-discretized particle systems (4.1.6) and (4.1.4) without any assumption on the initial probability measure  $m$ . Then we state that the strong rate of convergence of  $\mu_t^{N,h}$  to  $\mu_t$  for the Wasserstein distance with index one (or equivalently of  $F^{N,h}(t, \cdot)$  to  $F(t, \cdot)$  for the  $L^1$  norm) is  $\mathcal{O}\left(\frac{1}{\sqrt{N}} + h\right)$ , a result already obtained long ago by Bossy [8] under more regularity assumptions on the initial probability measure  $m$  and the function  $\Lambda$ . Our main result is that the weak rate of convergence is  $\mathcal{O}\left(\frac{1}{N} + h\right)$ . In Section 3, we introduce the reordered particle system and establish the mild formulation of the PDE (4.1.3) satisfied by  $F(t, x)$  and the perturbed version satisfied by  $F^{N,h}(t, x)$ . Section 4 is dedicated to the proofs of the results in Section 2. In Section 5, we study the initial error for both the random and the optimal deterministic initializations. We finally provide numerical experiments in Section 6 to illustrate our results. Beforehand, we introduce some additional notation.

## Notation:

- We denote by  $L_\Lambda = \sup_{u \in [0, 1]} |\lambda(u)|$  the Lipchitz constant of  $\Lambda$ . When  $\lambda$  is also assumed to be Lipschitz continuous, we denote similarly its Lipschitz constant by  $L_\lambda$ .
- For  $1 \leq p < \infty$ , we denote by  $L^p(\mathbb{R})$  the space of measurable real valued functions which are  $L^p$ -integrable for the Lebesgue measure i.e.  $f \in L^p$  if  $\|f\|_{L^p} = \left( \int_{\mathbb{R}} |f(x)|^p dx \right)^{\frac{1}{p}} < \infty$ .
- The space  $L^\infty(\mathbb{R})$  refers to the space of almost everywhere bounded measurable real valued functions endowed with the norm  $\|f\|_{L^\infty} = \inf\{C \geq 0 : |f(x)| \leq C \text{ for almost every } x \in \mathbb{R}\}$ .
- We denote the positive part of  $y \in \mathbb{R}$  by  $y^+ = \max(y, 0)$ .
- We denote by  $\Gamma$  the Gamma function defined by  $\Gamma(x) = \int_0^{+\infty} y^{x-1} \exp(-y) dy$  for  $x \in (0, +\infty)$ .
- For notational simplicity, when a function  $g$  defined on  $[0, T] \times \mathbb{R}$  and  $x \in \mathbb{R}$ , we may use sometimes the notation  $g_0(x) := g(0, x)$ .

## 4.2 Main results

We recall that  $\lambda$  is the derivative of the  $C^1$  function  $\Lambda : [0, 1] \rightarrow \mathbb{R}$ .

Kolli and Shkolnikov [56] prove a quantitative propagation of chaos result at optimal rate  $N^{-1/2}$  and convergence of the associated fluctuations for the particle system without time-discretization in the much more general and difficult case when the diffusion coefficient is also mean-field rank based. Taking advantage of the constancy of the diffusion coefficient, we are going to relax their assumptions on  $\lambda$  and  $m$  to prove the following result.

**Theorem 4.2.1.** *Let the initial positions  $X_0^i$  be i.i.d. according to  $m$  and  $(X_t^i)_{t \geq 0}$  denote the solution to the stochastic differential equation nonlinear in the sense of McKean (4.1.2) starting from  $X_0^i$  and driven by  $(W_t^i)_{t \geq 0}$ . If  $\lambda$  is Lipschitz continuous, then*

$$\forall \rho, T > 0, \exists C < \infty, \forall N \in \mathbb{N}^*, \forall i \in \{1, \dots, N\}, \mathbb{E} \left[ \sup_{t \in [0, T]} |X_t^i - X_t^{i, N}|^\rho + \sup_{t \in [0, T]} |X_t^i - \check{X}_t^{i, N}|^\rho \right] \leq CN^{-\rho/2}.$$

The estimation  $\mathbb{E} \left[ \sup_{t \in [0, T]} |X_t^i - \check{X}_t^{i, N}|^\rho \right] \leq CN^{-\rho/2}$  follows from Theorem 1.6 [56] when  $\lambda$  is differentiable with an Hölder continuous derivative and  $m$  has a bounded density w.r.t. the Lebesgue measure and a finite moment of order  $2 + \varepsilon$  for some  $\varepsilon > 0$ . An immediate consequence of Theorem 4.2.1 is to quantify the proximity of the two particles dynamics (4.1.4) and (4.1.6).

**Corollary 4.2.2.** *Assume that the initial positions  $X_0^i$  are i.i.d. according to  $m$  and that  $\lambda$  is Lipschitz continuous. Then:*

$$\forall \rho, T > 0, \exists C < \infty, \forall N \in \mathbb{N}^*, \forall i \in \{1, \dots, N\}, \mathbb{E} \left[ \sup_{t \in [0, T]} |X_t^{i, N} - \check{X}_t^{i, N}|^\rho \right] \leq CN^{-\rho/2}.$$

**Remark 4.2.3.** *By Theorem 2.1 in [40], propagation of chaos also holds at optimal rate  $N^{-1/2}$  in total variation distance.*

In the remaining of this section, we give the main results concerning the convergence of the empirical cumulative distribution function  $F^{N, h}$  of the Euler discretization with time-step  $h$  of the system with  $N$  interacting particles towards its limit  $F$ . We will make an intensive use of the interpretation of the  $L^1$ -norm of their difference as the Wasserstein distance with index 1 between  $\mu_t^{N, h} = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i, N, h}}$  and the law  $\mu_t$  of  $X_t$ .

The Wasserstein distance of index  $\rho \geq 1$  between two probability measures  $\mu$  and  $\nu$  on  $\mathbb{R}^d$  is defined by

$$\mathcal{W}_\rho^\rho(\mu, \nu) = \inf \{ \mathbb{E} [|X - Y|^\rho] ; \text{Law}(X) = \mu, \text{Law}(Y) = \nu \}.$$

In dimension  $d = 1$ , the Hoeffding-Fréchet or comonotone coupling given by the inverse transform sampling is optimal:

$$\mathcal{W}_\rho^\rho(\mu, \nu) = \int_0^1 |F_\mu^{-1}(u) - F_\nu^{-1}(u)|^\rho du \quad (4.2.1)$$

where  $F_\eta(x) = \eta((-\infty; x])$  and  $F_\eta^{-1}(u) = \inf \{x \in \mathbb{R} : F_\eta(x) \geq u\}$  respectively denote the cumulative distribution function and the quantile function of a probability measure  $\eta$  on  $\mathbb{R}$ . Since  $\int_0^1 |F_\mu^{-1}(u) - F_\nu^{-1}(u)| du = \int_{\mathbb{R}} |F_\mu(x) - F_\nu(x)| dx$ , the  $\mathcal{W}_1$  distance between two probability measures  $\mu$  and  $\nu$  on the real line is equal to the  $L^1$ -norm of the difference between the cumulative distribution functions of  $\mu$  and  $\nu$ :

$$\mathcal{W}_1(\mu, \nu) = \int_{\mathbb{R}} |F_\mu(x) - F_\nu(x)| dx. \quad (4.2.2)$$

We will also take advantage of the dual formulation of the  $\mathcal{W}_1$  distance which holds whatever  $d \in \mathbb{N}^*$ :

$$\mathcal{W}_1(\mu, \nu) = \sup_{\varphi \in \mathcal{L}} \left( \int_{\mathbb{R}^d} \varphi(x) \mu(dx) - \int_{\mathbb{R}^d} \varphi(x) \nu(dx) \right) \quad (4.2.3)$$

where  $\mathcal{L}$  denotes the set of all 1-Lipschitz function  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ .

The initial positions  $(X_0^i)_{i \geq 1}$  of the particles are either deterministic or random variables.

- When choosing a random initialization, we denote by  $\hat{F}_0^N(x) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{X_0^i \leq x\}}$  and  $\hat{\mu}_0^N = \frac{1}{N} \sum_{i=1}^N \delta_{X_0^i}$  the empirical cumulative distribution function and the empirical measure of the  $N$  first random variables in the sequence  $(X_0^i)_{i \geq 1}$  i.i.d. according to  $m$ .
- When choosing a deterministic initialization, we seek to construct a family  $x_1^N \leq x_2^N \leq \dots \leq x_N^N$  of initial positions minimizing the  $L^1$  norm of the difference between the piecewise constant function  $\tilde{F}_0^N(x) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{x_i^N \leq x\}}$  and  $F_0$ . According to (4.2.1),

$$\int_{\mathbb{R}} |\tilde{F}_0^N(x) - F_0(x)| dx = \sum_{i=1}^N \int_{\frac{i-1}{N}}^{\frac{i}{N}} |x_i^N - F_0^{-1}(u)| du.$$

Since, as remarked in [49], for  $i \in \llbracket 1, N \rrbracket$ ,  $y \mapsto N \int_{\frac{i-1}{N}}^{\frac{i}{N}} |y - F_0^{-1}(u)| du$  is minimal for  $y$  equal to the median  $F_0^{-1}(\frac{2i-1}{2N})$  of the image of the uniform law on  $[\frac{i-1}{N}, \frac{i}{N}]$  by  $F_0^{-1}$ , we choose  $x_i^N = F_0^{-1}(\frac{2i-1}{2N})$ . We denote by  $\tilde{\mu}_0^N = \frac{1}{N} \sum_{i=1}^N \delta_{F_0^{-1}(\frac{2i-1}{2N})}$  the associated empirical measure.

The next proposition, proved in Section 4.5.2, discusses assumptions under which the  $L^1$ -norm of the difference between  $F_0$  and  $\hat{F}_0^N$  or  $\tilde{F}_0^N$  is of order  $N^{-1/2}$ .

**Proposition 4.2.4.** *We denote for simplicity  $\int_{\mathbb{R}} |x|^{2+} m(dx) < \infty$  the existence of  $\varepsilon > 0$  such that  $\int_{\mathbb{R}} |x|^{2+\varepsilon} m(dx) < \infty$  and  $\int_{\mathbb{R}} |x|^{2-} m(dx) < \infty$  the fact that  $\int_{\mathbb{R}} |x|^{2-\varepsilon} m(dx) < \infty$  for each  $\varepsilon \in (0, 2]$ . We have the following results concerning the  $\mathcal{O}(N^{-1/2})$  behaviour of the errors:*

$$\begin{aligned} \sup_{N \geq 1} \sqrt{N} \mathbb{E} [\mathcal{W}_1(\hat{\mu}_0^N, m)] &< \infty \\ \Updownarrow \\ \int_{\mathbb{R}} |x|^{2+} m(dx) < \infty &\Rightarrow \int_{\mathbb{R}} \sqrt{F_0(x)(1-F_0(x))} dx < \infty \Rightarrow \int_{\mathbb{R}} |x|^2 m(dx) < \infty \\ \\ \int_{\mathbb{R}} |x|^2 m(dx) < \infty &\Rightarrow \sup_{x \geq 1} x \int_x^{+\infty} (F_0(-y) + 1 - F_0(y)) dy < \infty \Rightarrow \int_{\mathbb{R}} |x|^{2-} m(dx) < \infty \\ \Updownarrow \\ \sup_{N \geq 1} \sqrt{N} \mathcal{W}_1(\tilde{\mu}_0^N, m) &< \infty \end{aligned}$$

Concerning the weak error, since the empirical cumulative distribution function of i.i.d. samples is unbiased,  $\mathbb{E} [\hat{F}_0^N(x)] = F_0(x)$  for all  $N \geq 1$  and  $x \in \mathbb{R}$  and

$$\int_{\mathbb{R}} |\mathbb{E} [\hat{F}_0^N(x)] - F_0(x)| dx = 0. \quad (4.2.4)$$

As for the deterministic initialization, we have that:

$$\int_{\mathbb{R}} |\tilde{F}_0^N(x) - F_0(x)| dx = \int_{-\infty}^{F_0^{-1}(\frac{1}{2N})} F_0(x) dx + \sum_{i=1}^{N-1} \int_{F_0^{-1}(\frac{2i-1}{2N})}^{F_0^{-1}(\frac{2i+1}{2N})} \left| F_0(x) - \frac{i}{N} \right| dx + \int_{F_0^{-1}(\frac{2N-1}{2N})}^{+\infty} (1 - F_0(x)) dx \quad (4.2.5)$$

where the integrand is not greater than  $1/2N$ . When  $\exists -\infty < c \leq d < \infty$  such that  $m([c, d]) = 1$ , the integrand vanishes outside the interval  $[c, d]$ . One then easily deduces the next proposition proved in [49] by using the alternative formulation:

$$\int_{\mathbb{R}} \left| \tilde{F}_0^N(x) - F_0(x) \right| dx = \int_0^1 \left| \left( \tilde{F}_0^N \right)^{-1}(u) - F_0^{-1}(u) \right| du = \sum_{i=1}^N \int_{\frac{2i-1}{2N}}^{\frac{i}{N}} \left( F_0^{-1}(u) - F_0^{-1}\left(u - \frac{1}{2N}\right) \right) du. \quad (4.2.6)$$

**Proposition 4.2.5.** *When  $m$  is compactly supported i.e.  $\exists -\infty < c \leq d < \infty$  such that  $m([c, d]) = 1$ , then*

$$\mathcal{W}_1(\tilde{\mu}_0^N, m) \leq \frac{d-c}{2N}.$$

Let us now state our estimation of the strong error which is proved in Section 4.4.2.

**Theorem 4.2.6.** *Assume that for some  $\rho > 1$ ,  $\int_{\mathbb{R}} |x|^\rho m(dx) < \infty$  and assume either that the initial positions are optimal deterministic or the initial positions are i.i.d. according to  $m$ . Then*

$$\exists C < \infty, \forall N \in \mathbb{N}^*, \quad \sup_{t \leq T} \mathbb{E} \left[ \mathcal{W}_1 \left( \mu_t^{N,0}, \mu_t \right) \right] \leq C \left( \mathbb{E} [\mathcal{W}_1(\mu_0^N, m)] + \frac{1}{\sqrt{N}} \right).$$

Moreover, if  $\lambda$  is Lipschitz continuous then:

$$\exists C < \infty, \forall N \in \mathbb{N}^*, \forall h \in (0, T], \quad \sup_{t \leq T} \mathbb{E} \left[ \mathcal{W}_1 \left( \mu_t^{N,h}, \mu_t \right) \right] \leq C \left( \mathbb{E} [\mathcal{W}_1(\mu_0^N, m)] + \frac{1}{\sqrt{N}} + h \right).$$

Combining the theorem with Proposition 4.2.4, we have the following corollary:

**Corollary 4.2.7.** *Assume that the initial positions are*

- *either i.i.d. according to  $m$  and  $\int_{\mathbb{R}} \sqrt{F_0(x)(1-F_0(x))} dx < \infty$ ,*
- *or optimal deterministic and  $\sup_{x \geq 1} x \int_x^{+\infty} (F_0(-y) + 1 - F_0(y)) dy < \infty$ .*

Then:

$$\exists C < \infty, \forall N \in \mathbb{N}^*, \quad \sup_{t \leq T} \mathbb{E} \left[ \mathcal{W}_1 \left( \mu_t^{N,0}, \mu_t \right) \right] \leq \frac{C}{\sqrt{N}}.$$

Moreover, if  $\lambda$  is Lipschitz continuous then:

$$\exists C < \infty, \forall N \in \mathbb{N}^*, \forall h \in (0, T], \quad \sup_{t \leq T} \mathbb{E} \left[ \mathcal{W}_1 \left( \mu_t^{N,h}, \mu_t \right) \right] \leq C \left( \frac{1}{\sqrt{N}} + h \right).$$

Let us now state our main result, proved in Section 4.4.3, concerning the weak error: the  $L^1$ -weak error between the empirical cumulative distribution function  $F^{N,h}$  of the Euler discretization with time-step  $h$  of the system with  $N$  interacting particles and its limit  $F$  is  $\mathcal{O}(\frac{1}{N} + h)$ . We denote by  $\mathbb{E}[\mu_t^{N,h}]$  the probability measure on  $\mathbb{R}$  defined by

$$\int_{\mathbb{R}} \varphi(x) \mathbb{E}[\mu_t^{N,h}](dx) = \mathbb{E} \left[ \int_{\mathbb{R}} \varphi(x) \mu_t^{N,h}(dx) \right] = \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^N \varphi(X_t^{i,N,h}) \right]$$

for each  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  measurable and bounded. The cumulative distribution function of  $\mathbb{E}[\mu_t^{N,h}]$  is equal to  $\mathbb{E}[F^{N,h}(t, x)]$  and  $\mathcal{W}_1(\mathbb{E}[\mu_t^{N,h}], \mu_t) = \int_{\mathbb{R}} |\mathbb{E}[F^{N,h}(t, x)] - F(t, x)| dx$ .

**Theorem 4.2.8.** *Assume that  $\lambda$  is Lipschitz continuous and the initial positions are*

- *either i.i.d. according to  $m$  and  $\int_{\mathbb{R}} \sqrt{F_0(x)(1-F_0(x))} dx < \infty$ ,*



- or optimal deterministic and  $\sup_{x \geq 1} x \int_x^{+\infty} (F_0(-y) + 1 - F_0(y)) dy < \infty$ .

Then:

$$\exists C_b < \infty, \forall N \in \mathbb{N}^*, \forall h \in [0, T], \quad \sup_{t \leq T} \mathcal{W}_1 \left( \mathbb{E} [\mu_t^{N,h}], \mu_t \right) \leq C_b \left( \mathcal{W}_1 (\mathbb{E} [\mu_0^N], m) + \left( \frac{1}{N} + h \right) \right).$$

Combining the theorem with (4.2.4) and Proposition 4.2.5, we obtain the following corollary:

**Corollary 4.2.9.** *Assume that  $\lambda$  is Lipschitz continuous and the initial positions are*

- either i.i.d. according to  $m$  and  $\int_{\mathbb{R}} \sqrt{F_0(x)(1 - F_0(x))} dx < \infty$ ,
- or optimal deterministic with  $m$  compactly supported.

Then:

$$\exists C_b < \infty, \forall N \in \mathbb{N}^*, \forall h \in [0, T], \quad \sup_{t \leq T} \mathcal{W}_1 \left( \mathbb{E} [\mu_t^{N,h}], \mu_t \right) \leq C_b \left( \frac{1}{N} + h \right).$$

Using the dual formulation (4.2.3) of the Wasserstein distance, we deduce that if  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  is Lipschitz continuous with constant  $\text{Lip}(\varphi)$  then

$$\forall N \in \mathbb{N}^*, \forall t, h \in [0, T], \quad \left| \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^N \varphi \left( X_t^{i,N,h} \right) \right] - \mathbb{E} [\varphi(X_t)] \right| \leq C_b \text{Lip}(\varphi) \left( \frac{1}{N} + h \right).$$

**Remark 4.2.10.** *For the dynamics (4.1.4) with initial positions deterministic and given by  $x_i^N = F_0^{-1}(\frac{i}{N})$  when  $i = 1, \dots, N-1$  and  $x_N^N = F_0^{-1}(1 - \frac{1}{2N})$ , Bossy [8] proved an estimation also dealing with the supremum of the expected error between  $\check{F}^{N,h}(t, x)$  and  $F(t, x)$  similar to the last statement in Corollary 4.2.7:*

$$\exists C < \infty, \forall N \in \mathbb{N}^*, \forall h \in (0, T], \quad \sup_{t \leq T} \mathbb{E} \left[ \mathcal{W}_1 \left( \check{\mu}_t^{N,h}, \mu_t \right) \right] + \sup_{(t,x) \in [0,T] \times \mathbb{R}} \mathbb{E} \left[ \left| \check{F}^{N,h}(t, x) - F(t, x) \right| \right] \leq C \left( \frac{1}{\sqrt{N}} + h \right).$$

She assumes additional regularity on the coefficient  $\Lambda$ , namely that  $\Lambda$  is  $C^3$ , and on the initial measure  $m$ , namely that  $F_0$  is  $C^2$  bounded with bounded first and second order derivatives in  $x$  and that  $\exists M, \beta > 0, \alpha \geq 0$  such that  $|\partial_x F_0(x)| \leq \alpha \exp(-\beta x^2/2)$  when  $|x| > M$ . Her proof is based on the regularity of the backward Kolmogorov PDE associated with the generator of the diffusion (4.1.2). By contrast, our approach is based on a comparison of the mild formulation of the forward in time PDE (4.1.3) satisfied by  $F(t, x)$  and the perturbed mild formulation satisfied by  $F^{N,h}(t, x)$ . In fact, all the above results hold with  $\mu_t^{N,h}$  replaced by  $\check{\mu}_t^{N,h}$ . For those concerning  $\sup_{t \leq T} \mathbb{E} [\mathcal{W}_1(\mu_t^N, \mu_t)]$ , we just need to add the assumption that

$\lambda$  is Hölder continuous with exponent  $1/2$  to ensure that  $\sup_{1 \leq i \leq N} \sqrt{N} |\lambda^N(i) - \lambda(i/N)| < \infty$ . Notice that under Lipschitz continuity of  $\lambda$ , we even get  $\sup_{1 \leq i \leq N} N |\lambda^N(i) - \lambda(i/N)| < \infty$ . See Remark 4.3.4 below, where we outline how to adapt the proofs.

### 4.3 Dynamics of the reordered particle system and mild formulations

The reordering of mean-field rank based particle systems without time discretization has been first introduced in [44] and has proved to be a very useful tool in the study of the limit  $N \rightarrow \infty$  with vanishing viscosity (the parameter  $\sigma$  depends on  $N$  and tends to 0 as  $N \rightarrow \infty$ ) [45, 48] (the latter when the driving Brownian motions are replaced by symmetric  $\alpha$ -stable Lévy processes with  $\alpha > 1$ ), the long time behaviour of both the particle system and its mean-field limit [50] and the small noise limit  $\sigma \rightarrow 0$  of the particle system [51]. Before deriving the dynamics of the reordering of the Euler discretization (4.1.7), let us check the existence of the density  $p(t, x)$  of  $X_t$  for  $t > 0$ , which guarantees that, in the sense of distributions,  $\partial_x \Lambda(F(t, x)) = \lambda(F(t, x))p(t, x)$  so that  $F(t, x)$  is a weak solution of the viscous scalar conservation law (4.1.3).



**Lemma 4.3.1.** *For  $t > 0$ ,  $X_t$  admits a density  $p(t, x)$  with respect to the Lebesgue measure.*

*Proof.* We recall (4.1.2):  $X_t = X_0 + \sigma W_t + \int_0^t \lambda(F(s, X_s)) ds$ . The Brownian Motion  $(W_t)_{t \geq 0}$  is defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and let  $\mathcal{F}_t = \sigma(X_0, (W_s)_{s \leq t})$  denote the smallest filtration containing the natural filtration of the Brownian motion  $(W_t)_{t \geq 0}$  and  $\sigma(X_0)$ . Also, let  $\mathbb{Q}$  be the probability measure equivalent to  $\mathbb{P}$  defined, using the boundedness of  $\lambda$ , by:

$$\frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_t} = \exp \left\{ -\frac{1}{\sigma} \int_0^t \lambda(F(s, X_s)) dW_s - \frac{1}{2\sigma^2} \int_0^t \lambda^2(F(s, X_s)) ds \right\}.$$

Then by Girsanov's Theorem,  $(\frac{1}{\sigma}(X_t - X_0))_{t \geq 0}$  is a  $\mathbb{Q}$ -Brownian motion independent of  $X_0$ . This means that for any measurable and bounded function  $g$ , we have:

$$\mathbb{E}[g(X_t)] = \mathbb{E} \left[ g(X_0 + \sigma W_t) \exp \left\{ \frac{1}{\sigma} \int_0^t \lambda(F(s, X_0 + \sigma W_s)) dW_s - \frac{1}{2\sigma^2} \int_0^t \lambda^2(F(s, X_0 + \sigma W_s)) ds \right\} \right].$$

Let  $A$  be a Borel set of null Lebesgue measure. We choose  $g \equiv \mathbf{1}_A$  and  $t > 0$ . With the above equation and since  $X_0 + \sigma W_t$  has a density w.r.t. the Lebesgue measure, we have  $\mathbb{P}(X_t \in A) = 0$ . Therefore,  $X_t$  admits a density w.r.t. the Lebesgue measure.  $\square$

Let for each  $t \geq 0$ ,  $\eta_t$  be a permutation of  $\{1, \dots, N\}$  such that  $X_t^{\eta_t(1), N, h} \leq X_t^{\eta_t(2), N, h} \leq \dots \leq X_t^{\eta_t(N), N, h}$  and  $(Y_t^{i, N, h} = X_t^{\eta_t(i), N, h})_{i \in \llbracket 1, N \rrbracket}$  denote the increasing reordering also called order statistics of  $(X_t^{i, N, h})_{i \in \llbracket 1, N \rrbracket}$ . Even if the empirical measures of the reordered and original positions do not coincide

in general at the level of sample-paths, one has, for each  $t \geq 0$ ,  $\frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i, N, h}} = \frac{1}{N} \sum_{i=1}^N \delta_{Y_t^{i, N, h}}$  and therefore

$$F^{N, h}(t, x) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{Y_t^{i, N, h} \leq x\}}. \text{ By the Girsanov Theorem, reasoning like in the proof of Lemma 4.3.1,}$$

we show that for  $t > 0$ , the vector  $(X_t^{1, N, h}, X_t^{2, N, h}, \dots, X_t^{N, N, h})$  admits a density with respect to the Lebesgue measure on  $\mathbb{R}^N$  and therefore

$$\forall t > 0, a.s., \text{ the original (resp. reordered) particles have distinct positions.} \quad (4.3.1)$$

We are going to check that the function  $F(t, x)$  solves a mild formulation of the PDE (4.1.3) and  $F^{N, h}(t, x)$  solves a perturbed version of this mild formulation. To do so, it is convenient to obtain the dynamics of the reordered positions  $Y_t^{i, N, h}$ . Let  $\bar{\tau}_s^h = \lceil s/h \rceil h$  denote the discretization time right after  $s$ . We recall that  $\tau_s^h = \lfloor s/h \rfloor h$  denotes the discretization time right before  $s$ . We set  $t_k = kh$  for  $k \in \mathbb{N}$ . For  $s \in [t_k, t_{k+1})$ ,  $\tau_s^h = t_k$  and, for  $s \in (t_k, t_{k+1}]$ ,  $\bar{\tau}_s^h = t_{k+1}$ . For  $t > 0$ , let  $\eta_t^{-1}$  denote the inverse of the permutation  $\eta_t$ . By (4.3.1), *a.s.*, for each  $k \in \mathbb{N}^*$ , the positions  $(X_{t_k}^{i, N, h})_{1 \leq i \leq N}$  are distinct and for  $t$  in the time-interval  $[t_k, t_{k+1})$ ,  $X_t^{i, N, h}$  evolves with the drift coefficient  $\lambda^N(\eta_{t_k}^{-1}(i)) = \lambda^N(\eta_{\tau_t^h}^{-1}(i))$ . To obtain the same expression of the drift coefficient on the first time interval  $[0, t_1)$  we will use from now on the convention

$$\eta_0^{-1}(i) = \sum_{j=1}^N \mathbf{1}_{\{X_0^j \leq X_0^i\}} \quad \text{for } 1 \leq i \leq N. \quad (4.3.2)$$

With this convention, which is consistent with the usual definition of the inverse of a permutation only if the initial positions are distinct, we have

$$dX_t^{i, N, h} = \sigma dW_t^i + \lambda^N(\eta_{\tau_t^h}^{-1}(i)) dt, \quad 1 \leq i \leq N.$$

By Girsanov's theorem, we may define a new probability measure equivalent to the original one on each finite time horizon under which the processes  $(X_t^{i, N, h} - X_0^i = W_t^i + \int_0^t \lambda^N(\eta_{\tau_s^h}^{-1}(i)) ds)_{t \geq 0, 1 \leq i \leq N}$

are independent Brownian motions. Applying Lemma 3.7 [84], which states that under this probability measure, the reordered positions evolve as a  $N$ -dimensional Brownian motion normally reflected at the boundary of the simplex, we deduce that

$$dY_t^{i,N,h} = \sum_{j=1}^N \mathbf{1}_{\{Y_t^{i,N,h} = X_t^{j,N,h}\}} \left( \sigma dW_t^j + \lambda^N \left( \eta_{\tau_t^h}^{-1}(j) \right) dt \right) + (\gamma_t^i - \gamma_t^{i+1}) d|K|_t, \quad 1 \leq i \leq N$$

where the process  $K$  with coordinates  $K_t^i = \int_0^t (\gamma_s^i - \gamma_s^{i+1}) d|K|_s$  is an  $\mathbb{R}^N$ -valued continuous process with finite variation  $|K|$  such that:

$$d|K|_t \text{ a.e.}, \gamma_t^1 = \gamma_t^{N+1} = 0 \text{ and for } 2 \leq i \leq N, \gamma_t^i \geq 0 \text{ and } \gamma_t^i \left( Y_t^{i,N,h} - Y_t^{i-1,N,h} \right) = 0. \quad (4.3.3)$$

Defining a  $N$ -dimensional Brownian motion  $(\beta^1, \dots, \beta^N)$  by  $\beta_t^i = \sum_{j=1}^N \int_0^t \mathbf{1}_{\{Y_s^{i,N,h} = X_s^{j,N,h}\}} dW_s^j$  and using the definition of  $\eta_t$  and (4.3.1), we have

$$dY_t^{i,N,h} = \sigma d\beta_t^i + \lambda^N \left( \eta_{\tau_t^h}^{-1}(\eta_t(i)) \right) dt + (\gamma_t^i - \gamma_t^{i+1}) d|K|_t, \quad 1 \leq i \leq N. \quad (4.3.4)$$

Denoting by  $G_t(x) = \exp(-\frac{x^2}{2\sigma^2 t}) / \sqrt{2\pi\sigma^2 t}$  the probability density function of the normal law  $\mathcal{N}(0, \sigma^2 t)$ , we are now ready to state the mild formulation of the PDE (4.1.3) satisfied by  $F(t, x)$  and the perturbed version satisfied by  $F^{N,h}(t, x)$ .

**Proposition 4.3.2.** *For each  $t \geq 0$  and each  $h \in [0, T]$ , we have  $dx$  almost everywhere:*

$$F(t, x) = G_t * F_0(x) - \int_0^t \partial_x G_{t-s} * \Lambda(F(s, \cdot))(x) ds, \quad (4.3.5)$$

$$\begin{aligned} a.s. \quad F^{N,h}(t, x) = & G_t * F_0^{N,h}(x) - \int_0^t \partial_x G_{t-s} * \Lambda(F^{N,h}(s, \cdot))(x) ds - \frac{\sigma}{N} \sum_{i=1}^N \int_0^t G_{t-s}(X_s^{i,N,h} - x) dW_s^i \\ & + \frac{1}{N} \sum_{i=1}^N \int_0^t G_{t-s}(Y_s^{i,N,h} - x) \left[ \lambda^N(i) - \lambda^N \left( \eta_{\tau_s^h}^{-1}(\eta_s(i)) \right) \right] ds. \end{aligned} \quad (4.3.6)$$

**Remark 4.3.3.** *When  $h = 0$ , one should notice that the fourth term in  $F^{N,h}(t, x)$  is null so that:*

$$F^{N,0}(t, x) = G_t * F_0^{N,0}(x) - \int_0^t \partial_x G_{t-s} * \Lambda(F^{N,0}(s, \cdot))(x) ds - \frac{\sigma}{N} \sum_{i=1}^N \int_0^t G_{t-s}(X_s^{i,N,0} - x) dW_s^i.$$

**Remark 4.3.4.** *Let similarly for each  $t \geq 0$ ,  $\check{\eta}_t$  be a permutation of  $\{1, \dots, N\}$  such that  $\check{X}_t^{\check{\eta}_t(1),N,h} \leq \check{X}_t^{\check{\eta}_t(2),N,h} \leq \dots \leq \check{X}_t^{\check{\eta}_t(N),N,h}$  and  $(\check{Y}_t^{i,N,h} = \check{X}_t^{\check{\eta}_t(i),N,h})_{i \in [1,N]}$ . Let also  $\check{\eta}_t^{-1}$  denote the inverse of the permutation  $\check{\eta}_t$  for  $t > 0$  and  $\check{\eta}_0^{-1} = \eta_0^{-1}$ . Reasoning like in the proof of Proposition 4.3.2, we may derive the perturbed mild equation satisfied by the associated empirical cumulative distribution function  $\check{F}^{N,h}(t, x)$ :*

$$\begin{aligned} \check{F}^{N,h}(t, x) = & G_t * F_0^{N,h}(x) - \int_0^t \partial_x G_{t-s} * \Lambda(\check{F}^{N,h}(s, \cdot))(x) ds - \frac{\sigma}{N} \sum_{i=1}^N \int_0^t G_{t-s}(\check{X}_s^{i,N,h} - x) dW_s^i \\ & + \frac{1}{N} \sum_{i=1}^N \int_0^t G_{t-s}(\check{Y}_s^{i,N,h} - x) \left[ \lambda^N(i) - \lambda \left( \check{\eta}_{\tau_s^h}^{-1}(\check{\eta}_s(i))/N \right) \right] ds. \end{aligned}$$

Using the estimation

$$\int_{\mathbb{R}} \left| \frac{1}{N} \sum_{i=1}^N \int_0^t G_{t-s}(\check{Y}_s^{i,N,h} - x) \left[ \lambda^N(i) - \lambda \left( \check{\eta}_{\tau_s^h}^{-1}(\check{\eta}_s(i)) \right) - \lambda \left( \check{\eta}_{\tau_s^h}^{-1}(\check{\eta}_s(i))/N \right) \right] ds \right| dx \leq t \max_{1 \leq j \leq N} |\lambda^N(j) - \lambda(j/N)|$$

of the additional error term in comparison with (4.3.6), we may adapt all proofs to check the statements at the end of Remark 4.2.10.

*Proof.* Let  $t > 0$ ,  $f$  be a  $C^1$  and compactly supported function on  $\mathbb{R}$  and  $\varphi(s, x) = \int_{\mathbb{R}} \mathbf{1}_{\{x \leq y\}} G_{t-s} * f(y) dy$  be the convolution of  $G_{t-s}$  with  $x \mapsto \int_x^{+\infty} f(y) dy$  for  $(s, x) \in [0, t] \times \mathbb{R}$  and  $\varphi(t, x) = \int_{\mathbb{R}} \mathbf{1}_{\{x \leq y\}} f(y) dy$ . The function  $\varphi(s, x)$  is continuously differentiable w.r.t. to  $s$  and twice continuously differentiable w.r.t. to  $x$  on  $[0, t] \times \mathbb{R}$  and solves

$$\partial_s \varphi(s, x) + \frac{\sigma^2}{2} \partial_{xx} \varphi(s, x) = 0 \text{ for } (s, x) \in [0, t] \times \mathbb{R}. \quad (4.3.7)$$

Computing  $\varphi(t, X_t)$  where  $(X_s)_{s \geq 0}$  solves (4.1.2) and using (4.3.7), we obtain that:

$$\varphi(t, X_t) = \varphi(0, X_0) - \sigma \int_0^t G_{t-s} * f(X_s) dW_s - \int_0^t \lambda(F(s, X_s)) G_{t-s} * f(X_s) ds.$$

Since, on  $[0, t] \times \mathbb{R}$ ,  $G_{t-s} * f(x)$  is bounded by the supremum of  $|f|$ , the expectation of the stochastic integral is zero. By Fubini's theorem and since  $G_t$  is even, the expectations of  $\varphi(t, X_t)$  and  $\varphi(0, X_0)$  are respectively equal to  $\int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_{\{x \leq y\}} \mu_t(dx) f(y) dy = \int_{\mathbb{R}} F(t, y) f(y) dy$  and  $\int_{\mathbb{R}} F_0(y) G_t * f(y) dy = \int_{\mathbb{R}} G_t * F_0(y) f(y) dy$ . Using Fubini's theorem, the equality  $G_{t-s} * f(x) = - \int_{\mathbb{R}} \mathbf{1}_{\{x \leq y\}} \partial_y G_{t-s} * f(y) dy$ , the fact that, by the chain rule for continuous functions with finite variation,  $\int_{\mathbb{R}} \mathbf{1}_{\{x \leq y\}} \lambda(F(s, x)) p(s, x) dx = \Lambda(F(s, y)) - \Lambda(0)$ , the equality  $\int_{\mathbb{R}} \partial_y G_{t-s} * f(y) dy = 0$  and the oddness of  $\partial_y G_{t-s}$ , we obtain that the expectation of the last term in the right-hand side is equal to

$$\begin{aligned} \int_0^t \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_{\{x \leq y\}} \lambda(F(s, x)) p(s, x) dx \partial_y G_{t-s} * f(y) dy ds &= \int_0^t \int_{\mathbb{R}} (\Lambda(F(s, y)) - \Lambda(0)) \partial_y G_{t-s} * f(y) dy ds \\ &= - \int_0^t \int_{\mathbb{R}} \partial_y G_{t-s} * \Lambda(F(s, \cdot))(y) f(y) dy ds. \end{aligned}$$

Exchanging the time and space integrals by Fubini's theorem, we deduce that

$$\int_{\mathbb{R}} F(t, x) f(x) dx = \int_{\mathbb{R}} G_t * F_0(x) f(x) dx - \int_{\mathbb{R}} f(x) \int_0^t \partial_x G_{t-s} * \Lambda(F(s, \cdot))(x) ds dx.$$

Since  $f$  is arbitrary, we conclude that  $F$  satisfies the mild formulation (4.3.5).

Let us now establish that  $F^{N,h}$  satisfies a perturbed version of this equation. By computing  $\varphi(t, Y_t^{i,N,h})$  by Itô's formula, using (4.3.7) and summing over  $i \in \{1, \dots, N\}$ , we obtain

$$\begin{aligned} & \int_{\mathbb{R}} \sum_{i=1}^N \mathbf{1}_{\{Y_t^{i,N,h} \leq y\}} f(y) dy \\ &= \int_{\mathbb{R}} \sum_{i=1}^N \mathbf{1}_{\{Y_0^{i,N,h} \leq y\}} G_t * f(y) dy - \sum_{i=1}^N \int_0^t G_{t-s} * f(Y_s^{i,N,h}) \left( \sigma d\beta_s^i + \lambda^N \left( \eta_{\tau_s^h}^{-1}(\eta_s(i)) \right) ds \right) \\ & \quad + \sum_{i=1}^N \int_0^t \partial_x \varphi(s, Y_s^{i,N,h}) (\gamma_s^i - \gamma_s^{i+1}) d|K|_s. \end{aligned} \quad (4.3.8)$$

By summation by parts and (4.3.3),

$$\sum_{i=1}^N \int_0^t \partial_x \varphi(s, Y_s^{i,N,h}) (\gamma_s^i - \gamma_s^{i+1}) d|K|_s = \sum_{i=2}^N \int_0^t (\partial_x \varphi(s, Y_s^{i,N,h}) - \partial_x \varphi(s, Y_s^{i-1,N,h})) \gamma_s^i d|K|_s = 0.$$

Since the empirical cumulative distribution functions of the original and the reordered systems at time  $t$  coincide and the function  $G_t$  is even, the left-hand side and the first term in the right-hand side

are respectively equal to  $N \int_{\mathbb{R}} F^{N,h}(t, y) f(y) dy$  and  $N \int_{\mathbb{R}} G_t * F^{N,h}(0, y) f(y) dy$ . The definition of the Brownian motion  $\beta$  and (4.3.1) imply that

$$\sum_{i=1}^N \int_0^t G_{t-s} * f(Y_s^{i,N,h}) d\beta_s^i = \sum_{j=1}^N \int_0^t G_{t-s} * f(X_s^{j,N,h}) dW_s^j.$$

Dividing (4.3.8) by  $N$ , we deduce that

$$\begin{aligned} \int_{\mathbb{R}} F^{N,h}(t, y) f(y) dy &= \int_{\mathbb{R}} G_t * F^{N,h}(0, y) f(y) dy - \frac{\sigma}{N} \sum_{i=1}^N \int_0^t G_{t-s} * f(X_s^{i,N,h}) dW_s^i \\ &\quad - \frac{1}{N} \sum_{i=1}^N \int_0^t G_{t-s} * f(Y_s^{i,N,h}) \lambda^N \left( \eta_{\tau_s}^{-1}(\eta_s(i)) \right) ds. \end{aligned}$$

We are going to add and substract

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N G_{t-s} * f(Y_s^{i,N,h}) \lambda^N(i) &= - \int_{\mathbb{R}} \partial_y G_{t-s} * f(y) \sum_{i=1}^N \mathbf{1}_{\{Y_s^{i,N,h} \leq y\}} (\Lambda(i/N) - \Lambda((i-1)/N)) dy \\ &= - \int_{\mathbb{R}} \partial_y G_{t-s} * f(y) (\Lambda(F^{N,h}(s, y)) - \Lambda(0)) dy \\ &= \int_{\mathbb{R}} f(y) \partial_y G_{t-s} * \Lambda(F^{N,h}(s, \cdot))(y) dy. \end{aligned}$$

On the other hand, since  $f$  is square integrable and with the use of Young's inequality for the product and the estimate (B.2.7) from Lemma B.2.2, we have that:

$$\begin{aligned} \int_{\mathbb{R}} \left\{ \int_0^t |G_{t-s}(X_s^{i,N,h} - x) f(x)|^2 ds \right\}^{1/2} dx &= \int_{\mathbb{R}} |f(x)| \left\{ \int_0^t |G_{t-s}(X_s^{i,N,h} - x)|^2 ds \right\}^{1/2} dx \\ &\leq \frac{1}{2} \int_{\mathbb{R}} f^2(x) dx + \frac{1}{2} \int_{\mathbb{R}} \int_0^t G_{t-s}^2(X_s^{i,N,h} - x) ds dx \\ &= \frac{1}{2} \int_{\mathbb{R}} f^2(x) dx + \frac{1}{2\sigma} \sqrt{\frac{t}{\pi}} < \infty. \end{aligned}$$

For that reason, we can use a stochastic Fubini theorem stated by Veraar [86] and recalled in Lemma B.2.1 to deduce that  $\frac{\sigma}{N} \sum_{i=1}^N \int_0^t G_{t-s} * f(X_s^{i,N,h}) dW_s^i = \frac{\sigma}{N} \sum_{i=1}^N \int_{\mathbb{R}} f(x) \left\{ \int_0^t G_{t-s}(X_s^{i,N,h} - x) dW_s^i \right\} dx$ .

Therefore

$$\begin{aligned} \int_{\mathbb{R}} F^{N,h}(t, x) f(x) dx &= \int_{\mathbb{R}} G_t * F^{N,h}(0, x) f(x) dx - \frac{\sigma}{N} \sum_{i=1}^N \int_{\mathbb{R}} f(x) \left\{ \int_0^t G_{t-s}(X_s^{i,N,h} - x) dW_s^i \right\} dx \\ &\quad - \int_0^t \int_{\mathbb{R}} f(x) \partial_x G_{t-s} * \Lambda(F^{N,h}(s, \cdot))(x) dx ds \\ &\quad + \frac{1}{N} \sum_{i=1}^N \int_0^t \int_{\mathbb{R}} G_{t-s}(Y_s^{i,N,h} - x) f(x) dx \left\{ \lambda^N(i) - \lambda^N \left( \eta_{\tau_s}^{-1}(\eta_s(i)) \right) \right\} ds. \end{aligned}$$

Since  $f$  is bounded and  $\Lambda$  is bounded on the interval  $[0, 1]$ , using (B.2.4), we check that we can apply Fubini's theorem to interchange the space and time integrals in the two last terms of the right-hand side. Since  $f$  is arbitrary, we conclude that (4.3.6) holds a.s.  $dx$  a.e..  $\square$

## 4.4 Proofs of Section 2

### 4.4.1 Quantitative propagation of chaos result

The proof of Theorem 4.2.1 relies on the following Lemma which estimates for  $t > 0$  the  $L^\infty$ -norm of the density  $p(t, x)$  of  $X_t$  solution to (4.1.2) which exists according to Lemma 4.3.1.

**Lemma 4.4.1.**

$$\forall T \in (0, +\infty), \exists C_{\infty, T} < \infty, \forall t \in (0, T], \quad \|p(t, \cdot)\|_{L^\infty} \leq C_{\infty, T} t^{-1/2}.$$

*Proof.* Reasoning like at the beginning of the proof of Proposition 4.3.2 but with the function  $\varphi(s, x)$  equal to  $G_{t-s} * f$  and not its antiderivative, we easily check that  $p(t, x)$  satisfies the mild formulation :

$$\forall t > 0, \text{ dx a.e.}, \quad p(t, x) = G_t * m(x) - \int_0^t \partial_x G_{t-s} * (\lambda(F(s, \cdot))p(s, \cdot))(x) ds. \quad (4.4.1)$$

Since for  $t > 0$ ,  $\|G_t * m\|_{L^\infty} \leq \|G_t\|_{L^\infty} = (2\pi\sigma^2 t)^{-1/2}$ , it is enough to check that the estimation holds for the time integral in the mild formulation. By Jensen's inequality, Minkowski's inequality, Young's inequality then (B.2.3) and (B.2.6), (4.4.1) implies that, for  $t > 0$ ,

$$\begin{aligned} \|p(t, \cdot)\|_{L^2} &\leq \|G_t\|_{L^2} + \int_0^t \|\partial_x G_{t-s}\|_{L^2} \|\lambda(F(s, \cdot))p(s, \cdot)\|_{L^1} ds \\ &\leq \frac{1}{\sqrt{2\sigma(\pi t)^{1/4}}} + L_\Lambda \int_0^t \frac{ds}{2\sigma^{3/2}\pi^{1/4}(t-s)^{3/4}} = \frac{1}{\sqrt{2\sigma(\pi t)^{1/4}}} + \frac{2L_\Lambda t^{1/4}}{\sigma^{3/2}\pi^{1/4}} \end{aligned}$$

where the right-hand side is not greater than  $\left(\frac{1}{\sqrt{2\sigma(\pi)^{1/4}}} + \frac{2L_\Lambda T^{1/2}}{\sigma^{3/2}\pi^{1/4}}\right) t^{-1/4}$  for  $t \in (0, T]$ . With the boundedness of  $\lambda$  (we recall that  $\lambda$  is the derivative of the  $C^1$  function  $\Lambda : [0, 1] \rightarrow \mathbb{R}$ ) and Young's inequality for convolutions, we deduce that for  $t \in (0, T]$ ,

$$\begin{aligned} \|p(t, \cdot)\|_{L^\infty} &\leq \|G_t\|_{L^\infty} + L_\Lambda \int_0^t \|\partial_x G_{t-s}\|_{L^2} \|\lambda(F(s, \cdot))p(s, \cdot)\|_{L^2} ds \\ &\leq (2\pi\sigma^2 t)^{-1/2} + \frac{L_\Lambda}{2\sigma^{3/2}\pi^{1/4}} \left( \frac{1}{\sqrt{2\sigma(\pi)^{1/4}}} + \frac{2L_\Lambda T^{1/2}}{\sigma^{3/2}\pi^{1/4}} \right) \int_0^t \frac{ds}{(t-s)^{3/4}s^{1/4}}. \end{aligned}$$

Since  $\int_0^t \frac{ds}{(t-s)^{3/4}s^{1/4}} ds = \int_0^1 \frac{du}{(1-u)^{3/4}u^{1/4}} \leq \int_0^1 \frac{du}{(1-u)^{3/4}u^{1/4}} T^{1/2} t^{-1/2}$ , we easily conclude.  $\square$

We are now ready to prove Theorem 4.2.1 by adapting the proof of Theorem 1.6 [56]. Since, by Jensen's inequality, the conclusion with  $\rho = 1$  implies the conclusion with  $\rho \in (0, 1)$ , we suppose without loss of generality that  $\rho \geq 1$ . Lemma 4.4.1 implies the following estimation of the Lipschitz constant of  $x \mapsto \lambda(F(t, x))$ :

$$\forall t \in (0, T], \quad L_{\lambda(F(t, \cdot))} \leq C_{\infty, T} L_\Lambda t^{-1/2}. \quad (4.4.2)$$

We deduce that for a finite constant  $C$  which may change from line to line and depends on  $T$  but not on  $N$ :

$$\begin{aligned} \sup_{s \in [0, t]} |X_s^i - \check{X}_s^{i, N}|^\rho &\leq \left( \int_0^t \left| \lambda(F(u, X_u^i)) - \lambda(\check{F}^N(u, \check{X}_u^{i, N})) \right| du \right)^\rho \\ &= \left( \int_0^t u^{-\frac{\rho-1}{2\rho}} \times u^{\frac{\rho-1}{2\rho}} \left| \lambda(F(u, X_u^i)) - \lambda(\check{F}^N(u, \check{X}_u^{i, N})) \right| du \right)^\rho \\ &\leq \left( \int_0^t u^{-1/2} du \right)^{\rho-1} \int_0^t u^{(\rho-1)/2} \left( \left| \lambda(F(u, X_u^i)) - \lambda(F(u, \check{X}_u^{i, N})) \right| + \left| \lambda(F(u, \check{X}_u^{i, N})) - \lambda(\check{F}^N(u, \check{X}_u^{i, N})) \right| \right)^\rho du \\ &\leq C \int_0^t \left( u^{-1/2} |X_u^i - \check{X}_u^{i, N}|^\rho + u^{(\rho-1)/2} \left| \lambda(F(u, \check{X}_u^{i, N})) - \lambda(\check{F}^N(u, \check{X}_u^{i, N})) \right|^\rho \right) du, \end{aligned}$$

where we used Hölder's inequality for the second inequality. Using exchangeability of  $(\check{X}^{1, N}, \dots, \check{X}^{N, N})$ , denoting by  $\check{Y}_u^{1, N} \leq \check{Y}_u^{2, N} \leq \dots \leq \check{Y}_u^{N, N}$  (resp.  $Y_u^1 \leq Y_u^2 \leq \dots \leq Y_u^N$ ) the increasing reordering of  $(\check{X}_u^{1, N}, \dots, \check{X}_u^{N, N})$  (resp.  $(X_u^1, \dots, X_u^N)$ ) and using that (4.3.1) and its proof generalizes to the particle

system (4.1.4) then (4.4.2), we obtain that

$$\begin{aligned} \mathbb{E} \left[ \left| \lambda \left( F \left( u, \check{X}_u^{i,N} \right) \right) - \lambda \left( \check{F}^N \left( u, \check{X}_u^{i,N} \right) \right) \right|^\rho \right] &= \mathbb{E} \left[ \left| \frac{1}{N} \sum_{j=1}^N \left| \lambda \left( F \left( u, \check{X}_u^{j,N} \right) \right) - \lambda \left( \check{F}^N \left( u, \check{X}_u^{j,N} \right) \right) \right|^\rho \right] \\ &= \mathbb{E} \left[ \left| \frac{1}{N} \sum_{j=1}^N \left| \lambda \left( F \left( u, \check{Y}_u^{j,N} \right) \right) - \lambda \left( \check{F}^N \left( u, \check{Y}_u^{j,N} \right) \right) \right|^\rho \right] \\ &= \mathbb{E} \left[ \left| \frac{1}{N} \sum_{j=1}^N \left| \lambda(F(u, \check{Y}_u^{j,N})) - \lambda(F(u, Y_u^j)) + \lambda(F(u, Y_u^j)) - \lambda \left( \frac{j}{N} \right) \right|^\rho \right] \\ &\leq C \left( u^{-\rho/2} \mathbb{E} \left[ \left| \frac{1}{N} \sum_{j=1}^N \left| \check{Y}_u^{j,N} - Y_u^j \right|^\rho \right] + \mathbb{E} \left[ \left| \frac{1}{N} \sum_{j=1}^N \left| F(u, Y_u^j) - \frac{j}{N} \right|^\rho \right] \right). \end{aligned}$$

Since  $F(u, Y_u^1) \leq F(u, Y_u^2) \leq \dots \leq F(u, Y_u^N)$  is the increasing reordering of the random variables  $(F(u, X_u^i))_{1 \leq i \leq N}$  which are i.i.d. according to the uniform law on  $[0, 1]$ , according to the proof of Theorem 1.6 [56], the second expectation in the right-hand side is bounded from above by  $CN^{-\rho/2}$ . On the other hand, by (4.2.1),  $\frac{1}{N} \sum_{j=1}^N \left| \check{Y}_u^{j,N} - Y_u^j \right|^\rho = \mathcal{W}_\rho^\rho \left( \frac{1}{N} \sum_{i=1}^N \delta_{\check{Y}_u^i}, \frac{1}{N} \sum_{i=1}^N \delta_{X_u^i} \right) \leq \frac{1}{N} \sum_{j=1}^N \left| \check{X}_u^{j,N} - X_u^j \right|^\rho$ . We deduce that for all  $t \in [0, T]$ ,

$$\mathbb{E} \left[ \sup_{s \in [0, t]} \left| X_s^i - \check{X}_s^{i,N} \right|^\rho \right] \leq C \left( N^{-\rho/2} + \int_0^t u^{-1/2} \mathbb{E} \left[ \left| X_u^i - \check{X}_u^{i,N} \right|^\rho \right] du \right).$$

Performing the change of variable  $v = \sqrt{u}$  in the integral and setting  $f(t) = \mathbb{E} \left[ \sup_{s \in [0, t^2]} \left| X_s^i - \check{X}_s^{i,N} \right|^\rho \right]$ , we deduce that  $\forall t \in [0, \sqrt{T}]$ ,  $f(t) \leq C \left( N^{-\rho/2} + \int_0^t f(v) dv \right)$ . Since, by boundedness of  $\lambda$ , the function  $f$  is locally bounded, we conclude using Gronwall's lemma that  $\mathbb{E} \left[ \sup_{s \in [0, T]} \left| X_s^i - \check{X}_s^{i,N} \right|^\rho \right] \leq CN^{-\rho/2}$ .

Remarking that  $\left| X_t^{i,N} - X_0^i - \sigma W_t^i - \int_0^t \lambda \left( F^N(s, X_s^{i,N}) \right) ds \right| \leq \frac{L\lambda t}{2N}$ , we may adapt the arguments to deal with the particle system (4.1.6).

#### 4.4.2 Rate of convergence of the strong $L^1$ -error

To prove Theorem 4.2.6, we need the following lemmas.

**Lemma 4.4.2.**

$$\forall 0 \leq s \leq t \leq T, \forall \rho \geq 1, h \in [0, T], N \in \mathbb{N}^* \quad \sum_{j=1}^N \left| Y_t^{j,N,h} - Y_s^{j,N,h} \right|^\rho \leq \sum_{j=1}^N \left| X_t^{j,N,h} - X_s^{j,N,h} \right|^\rho.$$

*Proof.* By (4.2.1), we have  $\mathcal{W}_\rho^\rho \left( \mu_s^{N,h}, \mu_t^{N,h} \right) = \frac{1}{N} \sum_{j=1}^N \left| Y_t^{j,N,h} - Y_s^{j,N,h} \right|^\rho$ . Since  $\frac{1}{N} \sum_{j=1}^N \delta_{(X_s^{j,N,h}, X_t^{j,N,h})}$  is a coupling measure on  $\mathbb{R}^2$  with first marginal  $\mu_s^{N,h}$  and second marginal  $\mu_t^{N,h}$ , we conclude that

$$\frac{1}{N} \sum_{j=1}^N \left| X_t^{j,N,h} - X_s^{j,N,h} \right|^\rho \geq \mathcal{W}_\rho^\rho \left( \mu_s^{N,h}, \mu_t^{N,h} \right) = \frac{1}{N} \sum_{j=1}^N \left| Y_t^{j,N,h} - Y_s^{j,N,h} \right|^\rho.$$

□

This second lemma ensures the local integrability of  $t \mapsto \mathbb{E} \left[ \mathcal{W}_1 \left( \mu_t^{N,h}, \mu_t \right) \right]$ .

**Lemma 4.4.3.**

$$\forall t, h \in [0, T], \forall N \in \mathbb{N}^*, \quad \mathbb{E} \left[ \mathcal{W}_1 \left( \mu_t^{N,h}, \mu_t \right) \right] \leq 2\sigma \sqrt{\frac{2t}{\pi}} + 2L_\Lambda t + \mathbb{E} \left[ \mathcal{W}_1(\mu_0^{N,h}, m) \right].$$

If  $\int_{\mathbb{R}} |x|m(dx) < \infty$ , then for each  $N \in \mathbb{N}^*$  and each  $h \in [0, T]$ ,  $t \mapsto \mathbb{E}[\mathcal{W}_1(\mu_t^{N,h}, \mu_t)]$  is locally integrable on  $\mathbb{R}_+$ .

*Proof.* Using the triangle inequality, we have:

$$\mathbb{E} \left[ \mathcal{W}_1(\mu_t^{N,h}, \mu_t) \right] \leq \mathbb{E} \left[ \mathcal{W}_1(\mu_t^{N,h}, \mu_0^{N,h}) \right] + \mathbb{E} \left[ \mathcal{W}_1(\mu_0^{N,h}, m) \right] + \mathcal{W}_1(m, \mu_t). \quad (4.4.3)$$

Since

$$\begin{cases} \mathcal{W}_1(\mu_t, m) \leq \mathbb{E} [|X_t - X_0|] \leq \mathbb{E} [|\sigma W_t|] + \mathbb{E} \left[ \left| \int_0^t \lambda(F(s, X_s)) ds \right| \right] = \sigma \sqrt{\frac{2t}{\pi}} + L_\Lambda t, \\ \mathbb{E} \left[ \mathcal{W}_1(\mu_t^{N,h}, \mu_0^{N,h}) \right] \leq \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[ \left| \sigma W_t^i + \int_0^t \lambda^N \left( \mathbf{1}_{\{X_s^{j,N,h} \leq X_s^{i,N,h}\}} \right) ds \right| \right] \leq \sigma \sqrt{\frac{2t}{\pi}} + L_\Lambda t. \end{cases} \quad (4.4.4)$$

then by injecting (4.4.4) in (4.4.3), we obtain the upper-bound of  $\mathbb{E} \left[ \mathcal{W}_1(\mu_t^{N,h}, \mu_t) \right]$ . Moreover, by Lemma 4.5.1 below, the finiteness of the first order moment implies the finiteness of  $\mathbb{E} \left[ \mathcal{W}_1(\mu_0^{N,h}, m) \right]$  and therefore the local integrability of  $t \mapsto \mathbb{E} \left[ \mathcal{W}_1(\mu_t^{N,h}, \mu_t) \right]$ .  $\square$

The third lemma gives a control of the moments of order  $\rho \geq 1$  of  $X_t^{i,N,h}$ ,  $\forall i \in \llbracket 1, N \rrbracket$ .

**Lemma 4.4.4.** If  $\int_{\mathbb{R}} |x|^\rho m(dx) < \infty$  for some  $\rho \geq 1$ , then  $\forall N \in \mathbb{N}^*, \forall h \in [0, T]$ ,

$$\sup_{t \leq T} \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^N |X_t^{i,N,h}|^\rho \right] \leq M := \left( 2 \left( \int_{\mathbb{R}} |x|^\rho m(dx) \right)^{1/\rho} + \sqrt{2\sigma^2 T} \left( \frac{1}{\sqrt{\pi}} \Gamma \left( \frac{\rho+1}{2} \right) \right)^{1/\rho} + L_\Lambda T \right)^\rho.$$

*Proof.* By Minkowski's inequality,

$$\mathbb{E} \left[ |X_t^{i,N,h}|^\rho \right] \leq \left( \mathbb{E} [|X_0^i|^\rho]^{1/\rho} + \sigma \mathbb{E} [|W_t^i|^\rho]^{1/\rho} + \mathbb{E} \left[ \left| \int_0^t \lambda^N \left( \sum_{j=1}^N \mathbf{1}_{\{X_{\tau_s^h}^{j,N,h} \leq X_{\tau_s^h}^{i,N,h}\}} \right) ds \right|^\rho \right]^{1/\rho} \right)^\rho.$$

Since  $\mathbb{E} [|W_t^i|^\rho]^{1/\rho} = \sqrt{2t} \left( \frac{1}{\sqrt{\pi}} \Gamma \left( \frac{\rho+1}{2} \right) \right)^{1/\rho}$ , one easily concludes when the initial conditions are i.i.d. according to  $m$ . When they are optimal deterministic, we sum over  $i \in \{1, \dots, N\}$ , divide by  $N$  and use the second assertion in Lemma 4.5.16.  $\square$

**Proof of Theorem 4.2.6**

Defining for all  $t, h \in [0, T], N \in \mathbb{N}^*$ ,

$$R^{N,h}(t, x) = -\frac{\sigma}{N} \sum_{i=1}^N \int_0^t G_{t-s}(X_s^{i,N,h} - x) dW_s^i, \quad (4.4.5)$$

$$E^{N,h}(t, x) = \frac{1}{N} \sum_{i=1}^N \int_0^t G_{t-s}(Y_s^{i,N,h} - x) \left[ \lambda^N(i) - \lambda^N \left( \eta_{\tau_s^h}^{-1}(\eta_s(i)) \right) \right] ds, \quad (4.4.6)$$

we deduce from Proposition 4.3.2 that:

$$\begin{aligned} F^{N,h}(t, x) - F(t, x) &= G_t * \left( F_0^{N,h} - F_0 \right)(x) - \int_0^t \partial_x G_{t-s} * \left( \Lambda(F^{N,h}(s, \cdot)) - \Lambda(F(s, \cdot)) \right)(x) ds \\ &\quad + R^{N,h}(t, x) + E^{N,h}(t, x). \end{aligned} \quad (4.4.7)$$

Using the triangle inequality and taking expectations, we deduce that:

$$\begin{aligned} \mathbb{E} [\|F^{N,h}(t, \cdot) - F(t, \cdot)\|_{L^1}] &\leq \mathbb{E} [\|G_t * (F_0^{N,h} - F_0)\|_{L^1}] + \mathbb{E} [\|R^{N,h}(t, \cdot)\|_{L^1}] + \mathbb{E} [\|E^{N,h}(t, \cdot)\|_{L^1}] \\ &\quad + \mathbb{E} \left[ \left\| \int_0^t \partial_x G_{t-s} * (\Lambda(F^{N,h}(s, \cdot)) - \Lambda(F(s, \cdot))) ds \right\|_{L^1} \right]. \end{aligned}$$

Using the estimate (B.2.4) from Lemma B.2.2 and setting  $A = \frac{L_\Lambda}{\sigma} \sqrt{\frac{2}{\pi}}$ , we obtain:

$$\begin{aligned} \mathbb{E} \left[ \left\| \int_0^t \partial_x G_{t-s} * (\Lambda(F^{N,h}(s, \cdot)) - \Lambda(F(s, \cdot))) ds \right\|_{L^1} \right] &\leq \int_0^t \|\partial_x G_{t-s}\|_{L^1} L_\Lambda \mathbb{E} [\|F^{N,h}(s, \cdot) - F(s, \cdot)\|_{L^1}] ds \\ &= A \int_0^t \frac{1}{\sqrt{t-s}} \mathbb{E} [\|F^{N,h}(s, \cdot) - F(s, \cdot)\|_{L^1}] ds. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbb{E} [\|F^{N,h}(t, \cdot) - F(t, \cdot)\|_{L^1}] &\leq \mathbb{E} [\|F_0^{N,h} - F_0\|_{L^1}] + \mathbb{E} [\|R^{N,h}(t, \cdot)\|_{L^1}] + \mathbb{E} [\|E^{N,h}(t, \cdot)\|_{L^1}] \\ &\quad + A \int_0^t \frac{1}{\sqrt{t-s}} \mathbb{E} [\|F^{N,h}(s, \cdot) - F(s, \cdot)\|_{L^1}] ds. \end{aligned} \quad (4.4.8)$$

The next lemma states that the random variable  $R^{N,h}(t, x)$  is centered and provides an upper-bound for  $\mathbb{E} [\|R^{N,h}(t, \cdot)\|_{L^1}]$ .

**Lemma 4.4.5.** *We have  $\forall N \in \mathbb{N}^*, \forall h, t \in [0, T]$ ,  $\|\mathbb{E} [R^{N,h}(t, \cdot)]\|_{L^1} = 0$ . Moreover, if for some  $\rho > 1$ ,  $\int_{\mathbb{R}} |x|^\rho m(dx) < \infty$ , then:*

$$\exists R < \infty, \forall N \in \mathbb{N}^*, \forall h \in [0, T], \sup_{t \leq T} \mathbb{E} [\|R^{N,h}(t, \cdot)\|_{L^1}] \leq \frac{R}{\sqrt{N}}.$$

*Proof.* We have that  $\int_{\mathbb{R}} \mathbb{E} \left[ \int_0^t G_{t-s}^2 (X_s^{i,N,h} - x) ds \right] dx = \mathbb{E} \left[ \int_{\mathbb{R}} \int_0^t G_{t-s}^2 (X_s^{i,N,h} - x) ds dx \right] \leq \frac{1}{\sigma} \sqrt{\frac{t}{\pi}} < \infty$  according to the estimate (B.2.7) from Lemma B.2.2. Therefore,  $\mathbb{E} [R^{N,h}(t, x)] = 0$  dx a.e.. Moreover, denoting  $I_\rho = \int_{\mathbb{R}} \frac{dx}{1 + |x|^\rho}$  and using the Itô isometry for the first equality then Young's inequality for the second inequality and last the estimate (B.2.7) from Lemma B.2.2, we obtain:



$$\begin{aligned}
& \mathbb{E} [\|R^{N,h}(t, \cdot)\|_{L^1}] \\
& \leq \int_{\mathbb{R}} \mathbb{E}^{1/2} \left[ \left( \frac{\sigma}{N} \sum_{i=1}^N \int_0^t G_{t-s}(X_s^{i,N,h} - x) dW_s^i \right)^2 \right] dx \\
& = \frac{\sigma}{\sqrt{N}} \int_{\mathbb{R}} \mathbb{E}^{1/2} \left[ \frac{1}{N} \sum_{i=1}^N \int_0^t G_{t-s}^2(X_s^{i,N,h} - x) ds \right] dx \\
& = \frac{\sigma}{\sqrt{N}} \int_{\mathbb{R}} \mathbb{E}^{1/2} \left[ \frac{1}{N} \sum_{i=1}^N \int_0^t G_{t-s}^2(X_s^{i,N,h} - x) ds (1 + |x|^\rho) \right] \frac{dx}{\sqrt{1 + |x|^\rho}} \\
& \leq \frac{\sigma}{2\sqrt{N}} \int_{\mathbb{R}} \left( \frac{1}{1 + |x|^\rho} + \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^N \int_0^t G_{t-s}^2(X_s^{i,N,h} - x) ds (1 + |x|^\rho) \right] \right) dx \\
& = \frac{\sigma I_\rho}{2\sqrt{N}} + \frac{\sigma}{2\sqrt{N}} \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^N \int_0^t \left( \int_{\mathbb{R}} G_{t-s}^2(X_s^{i,N,h} - x) dx + \int_{\mathbb{R}} |x|^\rho G_{t-s}^2(X_s^{i,N,h} - x) dx \right) ds \right] \\
& = \frac{\sigma I_\rho}{2\sqrt{N}} + \frac{\sigma}{2\sqrt{N}} \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^N \frac{1}{\sigma} \sqrt{\frac{t}{\pi}} + \int_0^t \int_{\mathbb{R}} \frac{|X_s^{i,N,h} - y|^\rho}{2\sigma \sqrt{\pi(t-s)}} G_{(t-s)/2}(y) dy ds \right] \\
& \leq \frac{\sigma I_\rho}{2\sqrt{N}} + \frac{1}{2} \sqrt{\frac{t}{N\pi}} + \frac{2^{\rho-1}\sigma}{2\sqrt{N}} \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^N \int_0^t \frac{|X_s^{i,N,h}|^\rho}{2\sigma \sqrt{\pi(t-s)}} ds + \frac{\sigma^{\rho-1}}{2\pi} \Gamma\left(\frac{\rho+1}{2}\right) \int_0^t (t-s)^{(\rho-1)/2} ds \right] \\
& = \frac{1}{2\sqrt{N}} \left( \sigma I_\rho + \sqrt{\frac{t}{\pi}} \left( 1 + \frac{2^{\rho-1}\sigma^\rho}{\rho+1} \Gamma\left(\frac{\rho+1}{2}\right) \sqrt{\frac{t^\rho}{\pi}} \right) + \frac{2^{\rho-1}}{\sqrt{\pi}} \int_0^t \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^N |X_s^{i,N,h}|^\rho \right] \frac{ds}{2\sqrt{(t-s)}} \right).
\end{aligned}$$

With the use of Lemma 4.4.4, we conclude by setting

$$R = \frac{1}{2} \left( \sigma I_\rho + \sqrt{\frac{T}{\pi}} \left( 1 + 2^{\rho-1} M + \frac{2^{\rho-1}\sigma^\rho}{\rho+1} \Gamma\left(\frac{\rho+1}{2}\right) \sqrt{\frac{T^\rho}{\pi}} \right) \right).$$

□

Therefore, Inequality (4.4.8) becomes:

$$\begin{aligned}
\mathbb{E} [\|F^{N,h}(t, \cdot) - F(t, \cdot)\|_{L^1}] & \leq \mathbb{E} [\|F_0^{N,h} - F_0\|_{L^1}] + \frac{R}{\sqrt{N}} + \mathbb{E} [\|E^{N,h}(t, \cdot)\|_{L^1}] \\
& \quad + A \int_0^t \frac{1}{\sqrt{t-s}} \mathbb{E} [\|F^{N,h}(s, \cdot) - F(s, \cdot)\|_{L^1}] ds.
\end{aligned} \tag{4.4.9}$$

• One should notice that for  $h = 0$ ,  $E^{N,0}(t, x) = 0 \quad \forall t \in [0, T], N \in \mathbb{N}^*, x \in \mathbb{R}$ . Therefore, to control the term  $\mathbb{E} [\mathcal{W}_1(\mu_t^{N,0}, \mu_t)]$ , we iterate Inequality (4.4.9) and obtain:

$$\begin{aligned}
\mathbb{E} [\|F^{N,0}(t, \cdot) - F(t, \cdot)\|_{L^1}] & \leq (2A\sqrt{t} + 1) \left( \mathbb{E} [\|F_0^{N,0} - F_0\|_{L^1}] + \frac{R}{\sqrt{N}} \right) \\
& \quad + A^2 \int_0^t \mathbb{E} [\|F^{N,0}(r, \cdot) - F(r, \cdot)\|_{L^1}] \int_r^t \frac{ds}{\sqrt{t-s}\sqrt{s-r}} dr.
\end{aligned}$$

Since  $\int_r^t \frac{ds}{\sqrt{t-s}\sqrt{s-r}} = \pi$  and with the use of Lemma 4.4.3, we can apply Gronwall's lemma to deduce that:

$$\forall N \in \mathbb{N}^*, \sup_{t \leq T} \mathbb{E} [\mathcal{W}_1(\mu_t^{N,0}, \mu_t)] \leq C \left( \mathbb{E} [\mathcal{W}_1(\mu_0^N, m)] + \frac{1}{\sqrt{N}} \right)$$

where  $C = \max(1, R) (2A\sqrt{T} + 1) \exp(A^2\pi T)$ . This concludes the proof of Theorem 4.2.6 when  $h = 0$ .

• When  $h > 0$ , we need to estimate  $\mathbb{E} [\|E^{N,h}(t, \cdot)\|_{L^1}], h \in (0, T]$ .

**Proposition 4.4.6.** *We assume that for some  $\rho > 1$ ,  $\int_{\mathbb{R}} |x|^\rho m(dx) < \infty$  and that the function  $\lambda$  is Lipschitz continuous. Then  $\exists Z < \infty, \forall N \in \mathbb{N}^*, \forall h \in (0, T], \forall t \in [0, T]$ ,*

$$\mathbb{E} [\|E^{N,h}(t, \cdot)\|_{L^1}] \leq Z \left( \frac{1}{\sqrt{N}} + h + \sqrt{h} \mathbb{E} [\|F^{N,h}(t, \cdot) - F(t, \cdot)\|_{L^1}] \right. \\ \left. + \int_0^t \frac{1}{2\sqrt{t-s}} \mathbb{E} [\|F(s, \cdot) - F^{N,h}(s, \cdot)\|_{L^1}] ds \right).$$

Proposition 4.4.6 will be proved in Section 4.4.2.

From Equation (4.4.9) and Proposition 4.4.6, we have that:

$$(1 - Z\sqrt{h}) \mathbb{E} [\|F^{N,h}(t, \cdot) - F(t, \cdot)\|_{L^1}] \\ \leq \mathbb{E} [\|F_0^{N,h} - F_0\|_{L^1}] + \frac{Z+R}{\sqrt{N}} + Zh + \left(A + \frac{Z}{2}\right) \int_0^t \frac{1}{\sqrt{t-s}} \mathbb{E} [\|F^{N,h}(s, \cdot) - F(s, \cdot)\|_{L^1}] ds.$$

Hence, if we denote  $J = 2(Z+R)$  and  $K = 2A + Z$  then:

$$2(1 - Z\sqrt{h}) \mathbb{E} [\|F^{N,h}(t, \cdot) - F(t, \cdot)\|_{L^1}] \\ \leq 2\mathbb{E} [\|F_0^{N,h} - F_0\|_{L^1}] + J \left( \frac{1}{\sqrt{N}} + h \right) + \int_0^t \frac{K}{\sqrt{t-s}} \mathbb{E} [\|F^{N,h}(s, \cdot) - F(s, \cdot)\|_{L^1}] ds. \quad (4.4.10)$$

- When  $h \leq \frac{1}{4Z^2}$ , Equation (4.4.10) implies:

$$\mathbb{E} [\|F^{N,h}(t, \cdot) - F(t, \cdot)\|_{L^1}] \\ \leq 2\mathbb{E} [\|F_0^{N,h} - F_0\|_{L^1}] + J \left( \frac{1}{\sqrt{N}} + h \right) + \int_0^t \frac{K}{\sqrt{t-s}} \mathbb{E} [\|F^{N,h}(s, \cdot) - F(s, \cdot)\|_{L^1}] ds.$$

We iterate this inequality to obtain:

$$\mathbb{E} [\|F^{N,h}(t, \cdot) - F(t, \cdot)\|_{L^1}] \leq (1 + 2K\sqrt{t}) \left( 2\mathbb{E} [\|F_0^{N,h} - F_0\|_{L^1}] + J \left( \frac{1}{\sqrt{N}} + h \right) \right) \\ + K^2 \pi \int_0^t \mathbb{E} [\|F^{N,h}(r, \cdot) - F(r, \cdot)\|_{L^1}] dr.$$

With the use of Lemma 4.4.3, we can apply Gronwall's Lemma and deduce that for all  $t \in [0, T]$ :

$$\mathbb{E} [\|F^{N,h}(t, \cdot) - F(t, \cdot)\|_{L^1}] \leq (1 + 2K\sqrt{t}) \exp(K^2 \pi t) \left( 2\mathbb{E} [\|F_0^{N,h} - F_0\|_{L^1}] + J \left( \frac{1}{\sqrt{N}} + h \right) \right).$$

- When  $h > \frac{1}{4Z^2}$ , by Lemma 4.4.3 and (4.2.2),

$$\mathbb{E} [\|F^{N,h}(t, \cdot) - F(t, \cdot)\|_{L^1}] \leq 2\sigma \sqrt{\frac{2t}{\pi}} + 2L_\Lambda t + \mathbb{E} [\|F_0^{N,h} - F_0\|_{L^1}] \\ \leq 4Z^2 h \left( 2\sigma \sqrt{\frac{2t}{\pi}} + 2L_\Lambda t \right) + \mathbb{E} [\|F_0^{N,h} - F_0\|_{L^1}].$$

We choose  $C = \max \left( \max(2, J) (1 + 2K\sqrt{T}) \exp(K^2 \pi T), 4Z^2 (2\sigma \sqrt{\frac{2T}{\pi}} + 2L_\Lambda T) \right)$  and conclude that:

$$\forall N \in \mathbb{N}^*, \forall h \in (0, T], \forall t \in [0, T], \quad \sup_{t \leq T} \mathbb{E} [\mathcal{W}_1(\mu_t^{N,h}, \mu_t)] \leq C \left( \mathbb{E} [\mathcal{W}_1(\mu_0^N, m)] + \left( \frac{1}{\sqrt{N}} + h \right) \right).$$

Let us now prove Proposition 4.4.6 in the following section.

**Proof of Proposition 4.4.6**

We recall the expression of  $E^{N,h}(t, x)$ :

$$E^{N,h}(t, x) = \frac{1}{N} \sum_{i=1}^N \int_0^t G_{t-s}(Y_s^{i,N,h} - x) \left[ \lambda^N(i) - \lambda^N(\eta_{\tau_s^h}^{-1}(\eta_s(i))) \right] ds.$$

We do not know how to estimate the difference of values of  $\lambda^N$  between the brackets. For  $s > 0$ , we are going to take advantage of the permutation  $\eta_s^{-1} \circ \eta_{\tau_s^h}$  (because of the convention (4.3.2), this is not necessarily a permutation for  $s = 0$  and  $\eta_{\tau_s^h}^{-1} \circ \eta_{\tau_s^h}$  is equal to the identity permutation for  $s \geq h$  but not necessarily for  $s \in [0, h)$ ) to change indices and obtain the same value multiplied by a difference of values of the smooth function  $G_{t-s}$ . Using this permutation for the first equality then that  $Y_s^{\eta_s^{-1}(j), N, h} = X_s^{j, N, h}$  for  $s > 0$  and  $1 \leq j \leq N$  for the second one, we obtain that

$$\begin{aligned} E^{N,h}(t, x) &= \frac{1}{N} \sum_{i=1}^N \int_0^t G_{t-s}(Y_s^{i,N,h} - x) \lambda^N(i) - G_{t-s}\left(Y_s^{\eta_s^{-1}(\eta_{\tau_s^h}^i), N, h} - x\right) \lambda^N(\eta_{\tau_s^h}^{-1}(\eta_{\tau_s^h}^i)) ds \\ &= \frac{1}{N} \sum_{i=1}^N \int_0^t \left\{ \left( G_{t-s}(Y_s^{i,N,h} - x) - G_{t-s}\left(X_s^{\eta_{\tau_s^h}^i, N, h} - x\right) \right) \right\} \lambda^N(i) ds, \\ &\quad + \frac{1}{N} \sum_{i=1}^N \int_0^{t \wedge h} G_{t-s}(X_s^{\eta_0(i), N, h} - x) (\lambda^N(i) - \lambda^N(\eta_0^{-1}(\eta_0(i)))) ds. \end{aligned} \quad (4.4.11)$$

Subtracting  $G_{t-\tau_s^h}(Y_{\tau_s^h}^{i,N,h} - x) - G_{t-\tau_s^h}(X_{\tau_s^h}^{\eta_{\tau_s^h}^i, N, h} - x) = 0$  in the brace in the first term of the right-hand side makes apparent that this term is not too large since  $\tau_s^h$  is close to  $s$ . Computing  $G_{t-s}(Y_s^{i,N,h} - x) - G_{t-\tau_s^h}(Y_{\tau_s^h}^{i,N,h} - x)$  and  $G_{t-s}(X_s^{\eta_{\tau_s^h}^i, N, h} - x) - G_{t-\tau_s^h}(X_{\tau_s^h}^{\eta_{\tau_s^h}^i, N, h} - x)$  by Itô's formula, we obtain the following new expression of  $E^{N,h}(t, x)$ :

**Lemma 4.4.7.** *The process  $\tilde{\beta} = (\tilde{\beta}^1, \dots, \tilde{\beta}^N)$  where  $\tilde{\beta}_t^i = \sum_{j=1}^N \int_0^t \mathbf{1}_{\{\eta_{\tau_s^h}^i = j\}} dW_s^j$  is a  $N$ -dimensional Brownian motion and we can express  $E^{N,h}(t, x)$  as  $E^{N,h}(t, x) = \sum_{p=0}^5 e_p^{N,h}(t, x)$  where:*

- $e_0^{N,h}(t, x) = \frac{1}{N} \sum_{i=1}^N \int_0^{t \wedge h} G_{t-s}(X_s^{\eta_0(i), N, h} - x) (\lambda^N(i) - \lambda^N(\eta_0^{-1}(\eta_0(i)))) ds,$
- $e_1^{N,h}(t, x) = \frac{1}{N} \sum_{i=2}^N \int_0^t (t \wedge \bar{\tau}_s^h - s) (\lambda^N(i) - \lambda^N(i-1)) \partial_x G_{t-s}(Y_s^{i,N,h} - x) \gamma_s^i d|K|_s,$
- $e_2^{N,h}(t, x) = \frac{\sigma}{N} \sum_{i=1}^N \int_0^t (t \wedge \bar{\tau}_s^h - s) \lambda^N(i) \partial_x G_{t-s}(Y_s^{i,N,h} - x) d\beta_s^i,$
- $e_3^{N,h}(t, x) = -\frac{\sigma}{N} \sum_{i=1}^N \int_0^t (t \wedge \bar{\tau}_s^h - s) \lambda^N(i) \partial_x G_{t-s}(X_s^{\eta_{\tau_s^h}^i, N, h} - x) d\tilde{\beta}_s^i,$
- $e_4^{N,h}(t, x) = \frac{1}{N} \sum_{i=1}^N \int_0^t (t \wedge \bar{\tau}_s^h - s) \lambda^N(i) \lambda^N(\eta_{\tau_s^h}^{-1}(\eta_s(i))) \partial_x G_{t-s}(Y_s^{i,N,h} - x) ds,$
- $e_5^{N,h}(t, x) = -\frac{1}{N} \sum_{i=1}^N \int_0^t (t \wedge \bar{\tau}_s^h - s) \lambda^N(i) \lambda^N(\eta_{\tau_s^h}^{-1}(\eta_{\tau_s^h}^i)) \partial_x G_{t-s}(X_s^{\eta_{\tau_s^h}^i, N, h} - x) ds.$

Notice that in the definition of  $e_s^{N,h}(t, x)$ ,  $\lambda^N \left( \eta_{\tau_s^h}^{-1}(\eta_{\tau_s^h}(i)) \right) = \lambda^N(i)$  for  $s \geq h$ , but because of the convention (4.3.2), this equality does not necessarily hold for  $s \in [0, h)$ .

*Proof.* For  $1 \leq i, k \leq N$  and  $t \geq 0$ , one has

$$\langle \tilde{\beta}^i, \tilde{\beta}^k \rangle_t = \sum_{j=1}^N \int_0^t \mathbf{1}_{\{\eta_{\tau_s^h}(i)=j\}} \mathbf{1}_{\{\eta_{\tau_s^h}(k)=j\}} ds = \int_0^t \mathbf{1}_{\{\eta_{\tau_s^h}(i)=\eta_{\tau_s^h}(k)\}} ds = \mathbf{1}_{\{i=k\}} t,$$

since  $\eta_{\tau_s^h}$  is a permutation for each  $s \geq 0$ . One deduces that  $\tilde{\beta}$  is a Brownian motion by applying Lévy's characterization. By (4.4.11) and the equality  $G_{t-\tau_s^h} \left( Y_{\tau_s^h}^{i,N,h} - x \right) - G_{t-\tau_s^h} \left( X_{\tau_s^h}^{\eta_{\tau_s^h}(i),N,h} - x \right) = 0$ , it is enough to check that

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \int_0^t & \left\{ \left( G_{t-s} \left( Y_s^{i,N,h} - x \right) - G_{t-\tau_s^h} \left( Y_{\tau_s^h}^{i,N,h} - x \right) \right) \right. \\ & \left. - \left( G_{t-s} \left( X_s^{\eta_{\tau_s^h}(i),N,h} - x \right) - G_{t-\tau_s^h} \left( X_{\tau_s^h}^{\eta_{\tau_s^h}(i),N,h} - x \right) \right) \right\} \lambda^N(i) ds = \sum_{p=1}^5 e_p^{N,h}(t, x). \end{aligned}$$

We are going to compute the two differences in the right-hand side by applying Itô's formula. To do so, let us recall the dynamics of  $X_u^{\eta_{\tau_s^h}(i),N,h}$  for  $u \in [\tau_s^h, \bar{\tau}_s^h)$ :

$$dX_u^{\eta_{\tau_s^h}(i),N,h} = \sigma d\tilde{\beta}_u^i + \lambda^N \left( \eta_{\tau_s^h}^{-1}(\eta_{\tau_s^h}(i)) \right) du = \sigma d\tilde{\beta}_u^i + \lambda^N \left( \eta_{\tau_u^h}^{-1}(\eta_{\tau_u^h}(i)) \right) du.$$

We then have:

$$\begin{aligned} G_{t-s} \left( X_s^{\eta_{\tau_s^h}(i),N,h} - x \right) &= G_{t-\tau_s^h} \left( X_{\tau_s^h}^{\eta_{\tau_s^h}(i),N,h} - x \right) + \sigma \int_{\tau_s^h}^s \partial_x G_{t-u} \left( X_u^{\eta_{\tau_s^h}(i),N,h} - x \right) d\tilde{\beta}_u^i \\ &+ \int_{\tau_s^h}^s \left( \partial_u G_{t-u} \left( X_u^{\eta_{\tau_s^h}(i),N,h} - x \right) + \frac{\sigma^2}{2} \partial_{xx} G_{t-u} \left( X_u^{\eta_{\tau_s^h}(i),N,h} - x \right) \right. \\ &\left. + \lambda^N \left( \eta_{\tau_u^h}^{-1}(\eta_{\tau_u^h}(i)) \right) \partial_x G_{t-u} \left( X_u^{\eta_{\tau_s^h}(i),N,h} - x \right) \right) du. \end{aligned}$$

Since  $\partial_u G_{t-u} = -\partial_t G_{t-u}$ , by the heat equation (B.2.1) from Lemma B.2.2, we have:

$$\begin{aligned} & \int_0^t \left( G_{t-s} \left( X_s^{\eta_{\tau_s^h}(i),N,h} - x \right) - G_{t-\tau_s^h} \left( X_{\tau_s^h}^{\eta_{\tau_s^h}(i),N,h} - x \right) \right) ds \\ &= \sigma \int_0^t \int_{\tau_s^h}^s \partial_x G_{t-u} \left( X_u^{\eta_{\tau_s^h}(i),N,h} - x \right) d\tilde{\beta}_u^i ds + \int_0^t \int_{\tau_s^h}^s \lambda^N \left( \eta_{\tau_u^h}^{-1}(\eta_{\tau_u^h}(i)) \right) \partial_x G_{t-u} \left( X_u^{\eta_{\tau_s^h}(i),N,h} - x \right) du ds. \end{aligned} \quad (4.4.12)$$

Let us suppose that  $t > 0$  and treat each term of the right-hand side of the above equation. For  $x \in \mathbb{R}$ , The function  $u \mapsto \partial_x G_{t-u} \left( X_u^{\eta_{\tau_s^h}(i),N,h} - x \right)$  is continuous on  $[0, t)$ . Since  $\left( X_t^{1,N,h}, \dots, X_t^{N,N,h} \right)$  admits a density, as stated after the proof of Lemma 4.3.1,  $\mathbb{P} \left( X_t^{\eta_{\tau_s^h}(i),N,h} = x \right) \leq \sum_{j=1}^N \mathbb{P} \left( X_t^{j,N,h} = x \right) = 0$  a.s.. Therefore, a.s. the previous function has a vanishing limit as  $u \rightarrow t$  and is therefore bounded on the interval  $[0, t]$ . We can then apply Fubini's theorem to obtain:

$$\begin{aligned} & \int_0^t \int_{\tau_s^h}^s \lambda^N \left( \eta_{\tau_u^h}^{-1}(\eta_{\tau_u^h}(i)) \right) \partial_x G_{t-u} \left( X_u^{\eta_{\tau_s^h}(i),N,h} - x \right) du ds \\ &= \int_0^t (t \wedge \bar{\tau}_u^h - u) \lambda^N \left( \eta_{\tau_u^h}^{-1}(\eta_{\tau_u^h}(i)) \right) \partial_x G_{t-u} \left( X_u^{\eta_{\tau_u^h}(i),N,h} - x \right) du \quad \text{a.s..} \end{aligned}$$

Secondly, with the use of Young's inequality and the same arguments of density of  $(X_t^{1,N,h}, \dots, X_t^{N,N,h})$ , we get:

$$\begin{aligned} & \int_0^t \left( \int_{\tau_s^h}^s \left| \partial_x G_{t-u} \left( X_u^{\eta_{\tau_s^h}(i),N,h} - x \right) \right|^2 du \right)^{1/2} ds \\ &= \int_0^t \left( \int_{\tau_s^h}^s \frac{\left( X_u^{\eta_{\tau_s^h}(i),N,h} - x \right)^2}{2\sigma^5 \sqrt{\pi}(t-u)^{5/2}} G_{(t-u)/2} \left( X_u^{\eta_{\tau_s^h}(i),N,h} - x \right) du \right)^{1/2} ds \\ &\leq \frac{t}{2} + \frac{1}{2} \int_0^t \int_{\tau_s^h}^s \frac{\left( X_u^{\eta_{\tau_s^h}(i),N,h} - x \right)^2}{2\sigma^5 \sqrt{\pi}(t-u)^{5/2}} G_{(t-u)/2} \left( X_u^{\eta_{\tau_s^h}(i),N,h} - x \right) du ds < \infty \quad a.s.. \end{aligned}$$

Therefore, we can apply the stochastic Fubini Lemma B.2.1 and obtain:

$$\int_0^t \int_{\tau_s^h}^s \partial_x G_{t-u} \left( X_u^{\eta_{\tau_s^h}(i),N,h} - x \right) d\tilde{\beta}_u^i ds = \int_0^t (t \wedge \bar{\tau}_s^h - u) \partial_x G_{t-u} \left( X_s^{\eta_{\tau_s^h}(i),N,h} - x \right) d\tilde{\beta}_u^i.$$

Equation (4.4.12) becomes:

$$\begin{aligned} & \int_0^t \left( G_{t-s} \left( X_s^{\eta_{\tau_s^h}(i),N,h} - x \right) - G_{t-\tau_s^h} \left( X_{\tau_s^h}^{\eta_{\tau_s^h}(i),N,h} - x \right) \right) ds = \sigma \int_0^t (t \wedge \bar{\tau}_s^h - s) \partial_x G_{t-s} \left( X_s^{\eta_{\tau_s^h}(i),N,h} - x \right) d\tilde{\beta}_s^i \\ & + \int_0^t (t \wedge \bar{\tau}_s^h - s) \lambda^N \left( \eta_{\tau_s^h}^{-1}(\eta_{\tau_s^h}(i)) \right) \partial_x G_{t-s} \left( X_s^{\eta_{\tau_s^h}(i),N,h} - x \right) ds. \end{aligned} \quad (4.4.13)$$

Now, let us apply Itô's formula to  $G_{t-s}(Y_s^{i,N,h} - x)$ . Using once again the property (B.2.1) of the kernel  $G_t(x)$  from Lemma B.2.2 and the dynamics of  $Y_s^{i,N,h}$  given by (4.3.4), we have:

$$\begin{aligned} & G_{t-s}(Y_s^{i,N,h} - x) - G_{t-\tau_s^h}(Y_{\tau_s^h}^{i,N,h} - x) = \sigma \int_{\tau_s^h}^s \partial_x G_{t-u}(Y_u^{i,N,h} - x) d\beta_u^i \\ & + \int_{\tau_s^h}^s \partial_x G_{t-u}(Y_u^{i,N,h} - x) \lambda^N \left( \eta_{\tau_u^h}^{-1}(\eta_u(i)) \right) du + \int_{\tau_s^h}^s \partial_x G_{t-u}(Y_u^{i,N,h} - x) (\gamma_u^i - \gamma_u^{i+1}) d|K|_u. \end{aligned}$$

We use the same reasoning as for  $X_s^{\eta_{\tau_s^h}(i),N,h}$  to treat the integrals from 0 to  $t$  of the first two terms :

$$\begin{aligned} & \int_0^t \int_{\tau_s^h}^s \partial_x G_{t-u}(Y_u^{i,N,h} - x) d\beta_u^i ds = \int_0^t (t \wedge \bar{\tau}_s^h - s) \partial_x G_{t-s}(Y_s^{i,N,h} - x) d\beta_s^i, \\ & \int_0^t \int_{\tau_s^h}^s \partial_x G_{t-u}(Y_u^{i,N,h} - x) \lambda^N \left( \eta_{\tau_u^h}^{-1}(\eta_u(i)) \right) du ds = \int_0^t (t \wedge \bar{\tau}_s^h - s) \partial_x G_{t-s}(Y_s^{i,N,h} - x) \lambda^N \left( \eta_{\tau_s^h}^{-1}(\eta_s(i)) \right) ds. \end{aligned}$$

As for the last term  $\int_0^t \int_{\tau_s^h}^s \partial_x G_{t-u}(Y_u^{i,N,h} - x) (\gamma_u^i - \gamma_u^{i+1}) d|K|_u ds$ , we sum over  $i \in \llbracket 1, N \rrbracket$  after multiplying by  $\lambda^N(i)$  then we apply Fubini's theorem. Using the property (4.3.3), we finally obtain:

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N \lambda^N(i) \int_0^t \int_{\tau_s^h}^s \partial_x G_{t-u}(Y_u^{i,N,h} - x) (\gamma_u^i - \gamma_u^{i+1}) d|K|_u ds \\ &= \frac{1}{N} \sum_{i=2}^N \int_0^t (t \wedge \bar{\tau}_s^h - s) (\lambda^N(i) \partial_x G_{t-s}(Y_s^{i,N,h} - x) - \lambda^N(i-1) \partial_x G_{t-s}(Y_s^{i-1,N,h} - x)) \gamma_s^i d|K|_s \\ &= \frac{1}{N} \sum_{i=2}^N \int_0^t (t \wedge \bar{\tau}_s^h - s) (\lambda^N(i) - \lambda^N(i-1)) \partial_x G_{t-s}(Y_s^{i,N,h} - x) \gamma_s^i d|K|_s. \end{aligned}$$

Therefore,

$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N \int_0^t \lambda^N(i) \left( G_{t-s} (Y_s^{i,N,h} - x) - G_{t-\tau_s^h} (Y_{\tau_s^h}^{i,N,h} - x) \right) ds \\ &= \frac{\sigma}{N} \sum_{i=1}^N \lambda^N(i) \left\{ \int_0^t (t \wedge \bar{\tau}_s^h - s) \partial_x G_{t-s} (Y_s^{i,N,h} - x) d\beta_s^i + \int_0^t (t \wedge \bar{\tau}_s^h - s) \lambda^N \left( \eta_{\tau_s^h}^{-1}(\eta_s(i)) \right) \partial_x G_{t-s} (Y_s^{i,N,h} - x) ds \right\} \\ &+ \frac{1}{N} \sum_{i=2}^N \int_0^t (t \wedge \bar{\tau}_s^h - s) (\lambda^N(i) - \lambda^N(i-1)) \partial_x G_{t-s} (Y_s^{i,N,h} - x) \gamma_s^i d|K|_s. \end{aligned}$$

We conclude by combining this equality and the sum over  $i \in \llbracket 1, N \rrbracket$  of (4.4.13) multiplied by  $\lambda^N(i)/N$ .  $\square$

Now that we got rid of the difference of  $\lambda^N$  in the term  $E^{N,h}(t, x)$ , we can control the mean of the  $L^1$ -norm of this term. We present a succession of lemmas that will estimate each  $\mathbb{E} \left[ \|e_p^{N,h}(t, \cdot)\|_{L^1} \right]$  for  $p \in \llbracket 0, 5 \rrbracket$ .

Since  $G_{t-s}$  is a probability density,  $\mathbb{E} \left[ \|e_0^{N,h}(t, \cdot)\|_{L^1} \right] \leq \frac{1}{N} \sum_{i=1}^N \int_0^{t \wedge h} |\lambda^N(i) - \lambda^N(\eta_0^{-1}(\eta_0(i)))| ds$ .

Therefore, we obtain the following result concerning the term  $e_0^{N,h}(t, x)$ :

**Lemma 4.4.8.**

$$\forall N \in \mathbb{N}^*, \forall h \in (0, T], \quad \sup_{t \leq T} \left\| \mathbb{E} \left[ e_0^{N,h}(t, \cdot) \right] \right\|_{L^1} \leq \sup_{t \leq T} \mathbb{E} \left[ \|e_0^{N,h}(t, \cdot)\|_{L^1} \right] \leq 2L_\Lambda h.$$

We remark that the terms  $e_4^{N,h}(t, x)$  and  $e_5^{N,h}(t, x)$  are of the same nature.

**Lemma 4.4.9.** For  $r \in \{4, 5\}$ :

$$\exists C_{4,5} < \infty, \forall N \in \mathbb{N}^*, \forall h \in (0, T], \quad \sup_{t \leq T} \left\| \mathbb{E} \left[ e_r^{N,h}(t, \cdot) \right] \right\|_{L^1} \leq \sup_{t \leq T} \mathbb{E} \left[ \|e_r^{N,h}(t, \cdot)\|_{L^1} \right] \leq C_{4,5} h.$$

*Proof.* Let us treat the term  $e_5^{N,h}(t, x)$ .

We have, using the estimate (B.2.4) from Lemma B.2.2 for the second inequality, then the opposite monotonicities of the functions  $s \mapsto \frac{1}{\sqrt{t-s}}$  and  $s \mapsto (t \wedge t_{k+1} - s)$  on the time interval  $[t \wedge t_k, t \wedge t_{k+1}]$  for the third inequality that:

$$\begin{aligned} \left\| \mathbb{E} \left[ e_5^{N,h}(t, \cdot) \right] \right\|_{L^1} &\leq \mathbb{E} \left[ \|e_5^{N,h}(t, \cdot)\|_{L^1} \right] \leq \frac{1}{N} \sum_{i=1}^N \int_0^t (t \wedge \bar{\tau}_s^h - s) |\lambda^N(i)| \left| \lambda^N \left( \eta_{\tau_s^h}^{-1}(\eta_{\tau_s^h}(i)) \right) \right| \|\partial_x G_{t-s}\|_{L^1} ds \\ &\leq \frac{L_\Lambda^2}{\sigma} \sqrt{\frac{2}{\pi}} \sum_{k \in \mathbb{N}: t_k < t} \int_{t_k}^{t \wedge t_{k+1}} \frac{(t \wedge t_{k+1} - s)}{\sqrt{t-s}} ds \\ &\leq \frac{L_\Lambda^2}{\sigma} \sqrt{\frac{2}{\pi}} \left\{ \sum_{k \in \mathbb{N}: t_k < t} \frac{1}{t \wedge t_{k+1} - t_k} \int_{t_k}^{t \wedge t_{k+1}} (t \wedge t_{k+1} - s) ds \int_{t_k}^{t \wedge t_{k+1}} \frac{ds}{\sqrt{t-s}} \right\} \\ &\leq \frac{L_\Lambda^2}{\sigma} \sqrt{\frac{2}{\pi}} \left\{ \frac{h}{2} \sum_{k \in \mathbb{N}: t_k < t} \int_{t_k}^{t \wedge t_{k+1}} \frac{ds}{\sqrt{t-s}} \right\} = \frac{L_\Lambda^2}{\sigma} \sqrt{\frac{2t}{\pi}} h. \end{aligned}$$

The term  $e_4^{N,h}(t, x)$  can be estimated in the same way and the conclusion holds with  $C_{4,5} = \frac{L_\Lambda^2}{\sigma} \sqrt{\frac{2T}{\pi}}$ .  $\square$

We remark that the terms  $e_2^{N,h}(t, x)$  and  $e_3^{N,h}(t, x)$  are of the same nature as well.

**Lemma 4.4.10.** *For  $r \in \{2, 3\}$ ,  $\forall N \in \mathbb{N}^*$ ,  $\forall h \in (0, T]$ ,  $\|\mathbb{E}[e_r^{N,h}(t, \cdot)]\|_{L^1} = 0$ . Moreover, if  $\int_{\mathbb{R}} |x|^\rho m(dx) < \infty$  for some  $\rho > 1$ , then:*

$$\exists C_{2,3} < \infty, \quad \sup_{t \leq T} \mathbb{E} [\|e_r^{N,h}(t, \cdot)\|_{L^1}] \leq \frac{C_{2,3}}{\sqrt{N}}.$$

*Proof.* Let us treat the term  $e_3^{N,h}(t, x)$ .

Using the estimate (B.2.2) from Lemma A.1.2, we obtain that

$$\int_{\mathbb{R}} \mathbb{E} \left[ \int_0^t (t \wedge \bar{\tau}_s^h - s)^2 (\lambda^N(i))^2 (\partial_x G_{t-s})^2 \left( X_s^{\eta_{\tau_s^h(i), N, h}} - x \right) ds \right] dx \leq \int_0^t \frac{(t-s)^2 L_\Lambda^2}{4\sigma^3 (t-s)^{3/2} \sqrt{\pi}} ds < \infty.$$

Therefore,  $\mathbb{E}[e_3^{N,h}(t, x)] = 0$  dx a.e.. Moreover, using the estimate (B.2.3) from Lemma B.2.2, the Itô isometry for the first equality, Young's inequality for the second inequality and last Fubini's theorem, we obtain:

$$\begin{aligned} \mathbb{E} [\|e_3^{N,h}(t, \cdot)\|_{L^1}] &\leq \int_{\mathbb{R}} \mathbb{E}^{1/2} \left[ \left( \frac{\sigma}{N} \sum_{i=1}^N \int_0^t (t \wedge \bar{\tau}_s^h - s) \lambda^N(i) \partial_x G_{t-s} \left( X_s^{\eta_{\tau_s^h(i), N, h}} - x \right) d\tilde{\beta}_s^i \right)^2 \right] dx \\ &= \frac{\sigma}{\sqrt{N}} \int_{\mathbb{R}} \mathbb{E}^{1/2} \left[ \frac{1}{N} \sum_{i=1}^N \int_0^t (t \wedge \bar{\tau}_s^h - s)^2 |\lambda^N(i)|^2 (\partial_x G_{t-s})^2 \left( X_s^{\eta_{\tau_s^h(i), N, h}} - x \right) ds \right] dx \\ &= \frac{\sigma}{\sqrt{N}} \int_{\mathbb{R}} \mathbb{E}^{1/2} \left[ \frac{1}{N} \sum_{i=1}^N \int_0^t (t \wedge \bar{\tau}_s^h - s)^2 |\lambda^N(i)|^2 (\partial_x G_{t-s})^2 \left( X_s^{\eta_{\tau_s^h(i), N, h}} - x \right) ds (1 + |x|^\rho) \right] \frac{dx}{\sqrt{1 + |x|^\rho}} \\ &\leq \frac{\sigma}{2\sqrt{N}} \int_{\mathbb{R}} \left( \frac{1}{1 + |x|^\rho} + L_\Lambda^2 \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^N \int_0^t (t \wedge \bar{\tau}_s^h - s)^2 (\partial_x G_{t-s})^2 \left( X_s^{\eta_{\tau_s^h(i), N, h}} - x \right) ds (1 + |x|^\rho) \right] \right) dx \\ &= \frac{\sigma I_\rho}{2\sqrt{N}} + \frac{L_\Lambda^2}{4\sigma^4 \sqrt{N} \pi} \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^N \int_0^t \frac{(t \wedge \bar{\tau}_s^h - s)^2}{(t-s)^{5/2}} \left\{ \int_{\mathbb{R}} \left( X_s^{\eta_{\tau_s^h(i), N, h}} - x \right)^2 G_{(t-s)/2} \left( X_s^{\eta_{\tau_s^h(i), N, h}} - x \right) dx \right. \right. \\ &\quad \left. \left. + \int_{\mathbb{R}} |x|^\rho \left( X_s^{\eta_{\tau_s^h(i), N, h}} - x \right)^2 G_{(t-s)/2} \left( X_s^{\eta_{\tau_s^h(i), N, h}} - x \right) dx \right\} ds \right] \\ &= \frac{\sigma I_\rho}{2\sqrt{N}} + \frac{L_\Lambda^2}{4\sigma^4 \sqrt{N} \pi} \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^N \int_0^t \frac{(t \wedge \bar{\tau}_s^h - s)^2}{(t-s)^{5/2}} \left\{ \int_{\mathbb{R}} y^2 G_{(t-s)/2}(y) dy \right. \right. \\ &\quad \left. \left. + \int_{\mathbb{R}} \left| X_s^{\eta_{\tau_s^h(i), N, h}} - y \right|^\rho y^2 G_{(t-s)/2}(y) dy \right\} ds \right] \\ &\leq \frac{\sigma I_\rho}{2\sqrt{N}} + \frac{L_\Lambda^2}{4\sigma^4 \sqrt{N} \pi} \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^N \int_0^t \frac{(t \wedge \bar{\tau}_s^h - s)^2}{(t-s)^{5/2}} \left\{ \left( 1 + 2^{\rho-1} \left| X_s^{\eta_{\tau_s^h(i), N, h}} \right|^\rho \right) \frac{\sigma^2 (t-s)}{2} \right. \right. \\ &\quad \left. \left. + 2^{\rho-1} \frac{(\sigma \sqrt{t-s})^{2+\rho}}{\sqrt{\pi}} \Gamma \left( \frac{\rho+3}{2} \right) \right\} ds \right] \\ &= \frac{1}{2\sqrt{N}} \left( \sigma I_\rho + \frac{L_\Lambda^2}{4\sigma^2 \sqrt{\pi}} \left\{ \int_0^t \frac{(t \wedge \bar{\tau}_s^h - s)^2}{(t-s)^{3/2}} ds + 2^{\rho-1} \int_0^t \frac{(t \wedge \bar{\tau}_s^h - s)^2}{(t-s)^{3/2}} \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^N \left| X_s^{\eta_{\tau_s^h(i), N, h}} \right|^\rho \right] ds \right. \right. \\ &\quad \left. \left. + \frac{2^{\rho-2} \sigma^\rho}{\sqrt{\pi}} \Gamma \left( \frac{\rho+3}{2} \right) \int_0^t \frac{(t \wedge \bar{\tau}_s^h - s)^2}{(t-s)^{\frac{3-\rho}{2}}} ds \right\} \right). \end{aligned}$$

With the use of Lemma 4.4.4, we obtain:

$$\mathbb{E} \left[ \left\| e_3^{N,h}(t, \cdot) \right\|_{L^1} \right] \leq \frac{1}{2\sqrt{N}} \left( \sigma I_\rho + \frac{L_\Lambda^2}{4\sigma^2\sqrt{\pi}} \left\{ (1 + 2^{\rho-1}M) \int_0^t \frac{(t \wedge \bar{\tau}_s^h - s)^2}{(t-s)^{3/2}} ds \right. \right. \\ \left. \left. + \frac{2^{\rho-2}\sigma^\rho}{\sqrt{\pi}} \Gamma\left(\frac{\rho+3}{2}\right) \int_0^t \frac{(t \wedge \bar{\tau}_s^h - s)^2}{(t-s)^{\frac{3-\rho}{2}}} ds \right\} \right).$$

When  $t \geq h$ ,  $\int_0^t \frac{(t \wedge \bar{\tau}_s^h - s)^2}{(t-s)^{3/2}} ds \leq h^2 \int_0^{t-h} \frac{ds}{(t-s)^{3/2}} + \int_{t-h}^t \sqrt{t-s} ds \leq \frac{8}{3}h^{3/2}$  and  $\int_0^t \frac{(t \wedge \bar{\tau}_s^h - s)^2}{(t-s)^{\frac{3-\rho}{2}}} ds \leq \frac{2}{\rho-1}h^2t^{\frac{\rho-1}{2}} + \frac{2}{\rho+3}h^{\frac{\rho+3}{2}} \leq \frac{2(\rho+2)}{(\rho-1)(\rho+3)}h^{\frac{\rho+3}{2}}$ . When  $t \leq h$ ,  $\int_0^t \frac{(t \wedge \bar{\tau}_s^h - s)^2}{(t-s)^{3/2}} ds = \int_0^t \sqrt{t-s} ds = \frac{2}{3}t^{3/2} \leq \frac{2}{3}h^{3/2}$  and  $\int_0^t \frac{(t \wedge \bar{\tau}_s^h - s)^2}{(t-s)^{\frac{3-\rho}{2}}} ds \leq \frac{2}{\rho+3}h^{\frac{\rho+3}{2}}$ . The term  $e_2^{N,h}(t, x)$  can be estimated in the same way and the conclusion holds with

$$C_{2,3} = \frac{1}{2} \left( \sigma I_\rho + \frac{L_\Lambda^2 T^{3/2}}{4\sigma^2\sqrt{\pi}} \left( \frac{8}{3} (1 + 2^{\rho-1}M) + \frac{2^{\rho-1}\sigma^\rho}{\sqrt{\pi}} \times \frac{2(\rho+2)}{(\rho-1)(\rho+3)} \Gamma\left(\frac{\rho+3}{2}\right) T^{\frac{\rho-1}{2}} \right) \right).$$

□

Now, we finally treat the term  $e_1^{N,h}(t, x)$  in the lemma below:

**Lemma 4.4.11.** *If  $\lambda$  is Lipschitz continuous with constant  $L_\lambda$  then  $\exists C_1 < \infty, \forall N \in \mathbb{N}^*, \forall h \in (0, T], \forall t \in [0, T]$ ,*

$$\mathbb{E} \left[ \left\| e_1^{N,h}(t, \cdot) \right\|_{L^1} \right] \leq C_1 \left( h + \sqrt{h} \mathbb{E} \left[ \left\| F^{N,h}(t, \cdot) - F(t, \cdot) \right\|_{L^1} \right] + \mathbf{1}_{\{t \geq h\}} \int_0^{t-h} \frac{h}{2(t-s)^{3/2}} \mathbb{E} \left[ \left\| F^{N,h}(s, \cdot) - F(s, \cdot) \right\|_{L^1} \right] ds \right).$$

The proof of this assertion relies on the following results.

**Lemma 4.4.12.** *We have:*

$$\exists Q < \infty, \forall 0 \leq s \leq t \leq T, \quad \left\| F(t, \cdot) - F(s, \cdot) \right\|_{L^1} \leq Q \left( \sqrt{t} - \sqrt{s} \right) - L_\Lambda(t-s) \ln(t-s).$$

*Proof.* Let  $0 \leq s \leq t \leq T$ . We recall that:

$$F(t, x) - F(s, x) = (G_t - G_s) * F_0(x) - \int_0^t \partial_x G_{t-u} * \Lambda(F(u, \cdot))(x) du + \int_0^s \partial_x G_{s-u} * \Lambda(F(u, \cdot))(x) du.$$

Using Equality (B.2.1) and the estimates (B.2.4) and (B.2.5) from Lemma B.2.2, as well as the fact that



$(t-s) \leq 2\sqrt{T}(\sqrt{t}-\sqrt{s})$  and  $(1+\ln(x)) \leq x$ , we obtain:

$$\begin{aligned}
& \|F(t, \cdot) - F(s, \cdot)\|_{L^1} \\
& \leq \|(G_t - G_s) * F_0\|_{L^1} + \left\| \int_0^s (G_{t-u} - G_{s-u}) * \partial_x \Lambda(F(u, \cdot)) du \right\|_{L^1} + \left\| \int_s^t G_{t-u} * \partial_x \Lambda(F(u, \cdot)) du \right\|_{L^1} \\
& \leq \left\| \int_s^t (\partial_u G_u * F_0) du \right\|_{L^1} + \left\| \int_0^s \int_{s-u}^{t-u} \partial_r G_r * \partial_x \Lambda(F(u, \cdot)) dr du \right\|_{L^1} + \int_s^t \|G_{t-u} * (\lambda(F(u, \cdot))p(u, \cdot))\|_{L^1} du \\
& \leq \frac{\sigma^2}{2} \int_s^t \|(\partial_x G_u * m)\|_{L^1} du + \frac{\sigma^2}{2} \int_0^s \int_{s-u}^{t-u} \|\partial_{xx} G_r * (\lambda(F(u, \cdot))p(u, \cdot))\|_{L^1} dr du + L_\Lambda(t-s) \\
& \leq \frac{\sigma^2}{2} \int_s^t \sqrt{\frac{2}{\pi\sigma^2 u}} du + L_\Lambda \int_0^s \ln\left(\frac{t-u}{s-u}\right) du + L_\Lambda(t-s) \\
& = \sigma\sqrt{\frac{2}{\pi}}(\sqrt{t}-\sqrt{s}) + L_\Lambda(t \ln(t) - s \ln(s) - (t-s) \ln(t-s)) + L_\Lambda(t-s) \\
& \leq \left(\sigma\sqrt{\frac{2}{\pi}} + 2L_\Lambda\sqrt{T}\right)(\sqrt{t}-\sqrt{s}) + L_\Lambda \int_s^t (1+\ln(x)) dx - L_\Lambda(t-s) \ln(t-s) \\
& \leq \left(\sigma\sqrt{\frac{2}{\pi}} + 2L_\Lambda\sqrt{T}\right)(\sqrt{t}-\sqrt{s}) + \frac{L_\Lambda}{2}(t-s)(t+s) - L_\Lambda(t-s) \ln(t-s) \\
& \leq \left(\sigma\sqrt{\frac{2}{\pi}} + 2L_\Lambda\sqrt{T}(1+T)\right)(\sqrt{t}-\sqrt{s}) - L_\Lambda(t-s) \ln(t-s).
\end{aligned}$$

The conclusion holds with  $Q = \sigma\sqrt{2/\pi} + 2L_\Lambda\sqrt{T}(1+T)$ .  $\square$

The next lemma provides two different estimations of the term  $\mathbb{E} \left[ \int_s^t \gamma_u^i d|K|_u \right]$ . They are both useful to prove Lemma 4.4.11.

**Lemma 4.4.13.**  $\forall N \in \mathbb{N}^*, \forall i \in \llbracket 2, N \rrbracket, \forall h \in (0, T], \forall 0 \leq s \leq t \leq T,$

$$\mathbb{E} \left[ \left( \int_s^t \gamma_u^i d|K|_u \right)^2 \right] \leq 9N^2 (\sigma^2 + L_\Lambda^2 T) (t-s), \quad (4.4.14)$$

$$\text{and } \mathbb{E} \left[ \int_s^t \gamma_u^i d|K|_u \right] \leq N (\mathbb{E} [\|F^{N,h}(t, \cdot) - F^{N,h}(s, \cdot)\|_{L^1}] + L_\Lambda(t-s)). \quad (4.4.15)$$

*Proof.* Let  $2 \leq i \leq N$ . Since  $\gamma_u^{N+1} = 0$ , we have  $\int_s^t \gamma_u^i d|K|_u = \int_s^t \sum_{j=i}^N (\gamma_u^j - \gamma_u^{j+1}) d|K|_u$  and with the dynamics (4.3.4) of  $Y^{j,N,h}$ , we deduce that

$$\int_s^t \gamma_u^i d|K|_u = \int_s^t \sum_{j=i}^N (\gamma_u^j - \gamma_u^{j+1}) d|K|_u = \sum_{j=i}^N \left\{ \left( Y_t^{j,N,h} - Y_s^{j,N,h} \right) - \sigma \left( \beta_t^j - \beta_s^j \right) - \int_s^t \lambda^N \left( \sigma_{\tau_u^h}^{-1}(\eta_u(j)) \right) du \right\}.$$

Let us start by proving the estimation of  $\mathbb{E} \left[ \left( \int_s^t \gamma_u^i d|K|_u \right)^2 \right]$ . With the use of Jensen's inequality and

Lemma 4.4.2 for  $\rho = 2$ , we obtain:

$$\begin{aligned}
& \mathbb{E} \left[ \left( \int_s^t \gamma_u^i d|K|_u \right)^2 \right] \\
& \leq 3N \left( \sum_{j=1}^N \mathbb{E} \left[ \left| Y_t^{j,N,h} - Y_s^{j,N,h} \right|^2 \right] + \sum_{j=1}^N \left( \int_s^t \left| \lambda^N \left( \sigma_{\tau_u^h}^{-1}(\eta_u(j)) \right) \right| du \right)^2 + \sigma^2 \sum_{j=1}^N \mathbb{E} \left[ \left| \beta_t^j - \beta_s^j \right|^2 \right] \right) \\
& \leq 3N \left( \sum_{j=1}^N \mathbb{E} \left[ \left| X_t^{j,N,h} - X_s^{j,N,h} \right|^2 \right] + NL_\Lambda^2(t-s)^2 + N\sigma^2(t-s) \right) \\
& \leq 3N \left( \sum_{j=1}^N 2\mathbb{E} \left[ \sigma^2 |W_t^j - W_s^j|^2 + L_\Lambda^2(t-s)^2 \right] + NL_\Lambda^2(t-s)^2 + N\sigma^2(t-s) \right) \\
& \leq 9N^2 (\sigma^2 + L_\Lambda^2 T) (t-s).
\end{aligned}$$

Notice that because of the latter contribution of  $\mathbb{E} \left[ |W_t^j - W_s^j|^2 \right]$ , it was not useful to take advantage of the independence of the Brownian motions  $\beta^j$  which ensures  $\mathbb{E} \left[ \left| \sum_{j=i}^N (\beta_t^j - \beta_s^j) \right|^2 \right] = (N+1-i)(t-s)$ . Let us now prove the second estimation of  $\mathbb{E} \left[ \int_s^t \gamma_u^i d|K|_u \right]$ . To do so, we use that, according to (4.2.1) and (4.2.2),  $\frac{1}{N} \sum_{i=1}^N \left| Y_t^{i,N,h} - Y_s^{i,N,h} \right| = \mathcal{W}_1 \left( \mu_t^{N,h}, \mu_s^{N,h} \right) = \int_{\mathbb{R}} |F^{N,h}(t, x) - F^{N,h}(s, x)| dx = \|F^{N,h}(t, \cdot) - F^{N,h}(s, \cdot)\|_{L^1}$  to obtain:

$$\begin{aligned}
\mathbb{E} \left[ \int_s^t \gamma_u^i d|K|_u \right] & \leq \sum_{j=1}^N \mathbb{E} \left[ \left| Y_t^{j,N,h} - Y_s^{j,N,h} \right| \right] + \sum_{j=1}^N \int_s^t \left| \lambda^N \left( \sigma_{\tau_u^h}^{-1}(\eta_u(j)) \right) \right| du \\
& \leq N\mathbb{E} \left[ \|F^{N,h}(t, \cdot) - F^{N,h}(s, \cdot)\|_{L^1} \right] + NL_\Lambda(t-s).
\end{aligned}$$

□

Let us now prove Lemma 4.4.11.

*Proof.* We recall that  $e_1^{N,h}(t, x) = \frac{1}{N} \sum_{i=2}^N \int_0^t (t \wedge \bar{\tau}_s^h - s) (\lambda^N(i) - \lambda^N(i-1)) \partial_x G_{t-s} (Y_s^{i,N,h} - x) \gamma_s^i d|K|_s$ .

For  $i \in \llbracket 2, N \rrbracket$ , we have  $|\lambda^N(i) - \lambda^N(i-1)| = \left| N \int_{\frac{i-1}{N}}^{\frac{i}{N}} \left( \lambda(u) - \lambda \left( u - \frac{1}{N} \right) \right) du \right| \leq \frac{L_\lambda}{N}$ . Using the estimate (B.2.4) from Lemma B.2.2 and the property (4.3.3), we have:

$$\mathbb{E} \left[ \|e_1^{N,h}(t, \cdot)\|_{L^1} \right] \leq \frac{L_\lambda}{\sigma N} \sqrt{\frac{2}{\pi}} \left\{ \frac{1}{N} \sum_{i=2}^N \mathbb{E} \left[ \int_0^t \frac{(t \wedge \bar{\tau}_s^h - s)}{\sqrt{t-s}} \gamma_s^i d|K|_s \right] \right\}. \quad (4.4.16)$$

- For  $t \leq h$ , since  $\frac{(t \wedge \bar{\tau}_s^h - s)}{\sqrt{t-s}} = \sqrt{t-s} \leq \sqrt{h}$ , we deduce from (4.4.14) that:

$$\mathbb{E} \left[ \|e_1^{N,h}(t, \cdot)\|_{L^1} \right] \leq \frac{3L_\lambda}{\sigma} \sqrt{\frac{2(\sigma^2 + L_\Lambda^2 T)}{\pi}} h.$$

- For  $t \geq h$ , we decompose the right-hand side of inequality (4.4.16) onto the sub-intervals  $[0, t-h]$  and  $[t-h, t]$  for a better control. Therefore,

$$\mathbb{E} \left[ \|e_1^{N,h}(t, \cdot)\|_{L^1} \right] \leq \frac{L_\lambda h}{\sigma N^2} \sqrt{\frac{2}{\pi}} \sum_{i=2}^N \left( \mathbb{E} \left[ \int_0^{t-h} \frac{1}{\sqrt{t-s}} \gamma_s^i d|K|_s \right] + \frac{1}{\sqrt{h}} \mathbb{E} \left[ \int_{t-h}^t \gamma_s^i d|K|_s \right] \right).$$

As for the first term of the right-hand side of the above inequality, we introduce  $A_s = -\int_s^t \gamma_u^i d|K|_u$  and apply Fubini's theorem to obtain:

$$\begin{aligned} \int_0^{t-h} A_s \frac{ds}{2(t-s)^{3/2}} &= \int_0^{t-h} \left( A_0 + \int_0^s dA_r \right) \frac{ds}{2(t-s)^{3/2}} = A_0 \left( \frac{1}{\sqrt{h}} - \frac{1}{\sqrt{t}} \right) + \int_0^{t-h} \int_0^s \frac{dA_r}{2(t-s)^{3/2}} ds \\ &= A_0 \left( \frac{1}{\sqrt{h}} - \frac{1}{\sqrt{t}} \right) + \int_0^{t-h} \int_r^{t-h} \frac{ds}{2(t-s)^{3/2}} dA_r \\ &= A_0 \left( \frac{1}{\sqrt{h}} - \frac{1}{\sqrt{t}} \right) + \int_0^{t-h} \left( \frac{1}{\sqrt{h}} - \frac{1}{\sqrt{t-r}} \right) dA_r \\ &= -\frac{1}{\sqrt{t}} A_0 + \frac{1}{\sqrt{h}} A_{t-h} - \int_0^{t-h} \frac{1}{\sqrt{t-r}} dA_r. \end{aligned}$$

Consequently, we obtain that:

$$\begin{aligned} \mathbb{E} \left[ \int_0^{t-h} \frac{1}{\sqrt{t-s}} \gamma_s^i d|K|_s \right] &+ \frac{1}{\sqrt{h}} \mathbb{E} \left[ \int_{t-h}^t \gamma_u^i d|K|_u \right] \\ &= \frac{1}{\sqrt{t}} \mathbb{E} \left[ \int_0^t \gamma_u^i d|K|_u \right] + \mathbb{E} \left[ \int_0^{t-h} \frac{1}{2(t-s)^{3/2}} \int_s^t \gamma_u^i d|K|_u ds \right]. \end{aligned}$$

We shall use the estimate (4.4.14) and the estimate (4.4.15) from Lemma 4.4.13 for respectively the first term and the second term of the right-hand side of the following inequality:

$$\begin{aligned} \mathbb{E} \left[ \|e_1^{N,h}(t, \cdot)\|_{L^1} \right] &\leq \frac{L_\lambda h}{\sigma N^2} \sqrt{\frac{2}{\pi}} \sum_{i=2}^N \left\{ \frac{1}{\sqrt{t}} \mathbb{E} \left[ \int_0^t \gamma_u^i d|K|_u \right] + \int_0^{t-h} \frac{1}{2(t-s)^{3/2}} \mathbb{E} \left[ \int_s^t \gamma_u^i d|K|_u \right] ds \right\} \\ &\leq \frac{L_\lambda h}{\sigma N^2} \sqrt{\frac{2}{\pi}} \sum_{i=2}^N \left\{ 3N \sqrt{(\sigma^2 + L_\Lambda^2 T)} \right. \\ &\quad \left. + N \int_0^{t-h} \frac{L_\Lambda}{2\sqrt{t-s}} ds + N \int_0^{t-h} \frac{1}{2(t-s)^{3/2}} \mathbb{E} [\|F^{N,h}(t, \cdot) - F^{N,h}(s, \cdot)\|_{L^1}] ds \right\} \\ &= \frac{L_\lambda}{\sigma} \sqrt{\frac{2}{\pi}} \left( 3\sqrt{(\sigma^2 + L_\Lambda^2 T)} + L_\Lambda (\sqrt{t} - \sqrt{h}) \right) h \\ &\quad + \frac{L_\lambda}{\sigma} \sqrt{\frac{2}{\pi}} \int_0^{t-h} \frac{h}{2(t-s)^{3/2}} \mathbb{E} [\|F^{N,h}(t, \cdot) - F^{N,h}(s, \cdot)\|_{L^1}] ds \\ &\leq \frac{L_\lambda}{\sigma} \sqrt{\frac{2}{\pi}} \left( 3\sqrt{(\sigma^2 + L_\Lambda^2 T)} + L_\Lambda \sqrt{T} \right) h + \frac{L_\lambda}{\sigma} \sqrt{\frac{2}{\pi}} \int_0^{t-h} \frac{h}{2(t-s)^{3/2}} \mathbb{E} [\|F^{N,h}(t, \cdot) - F^{N,h}(s, \cdot)\|_{L^1}] ds. \end{aligned}$$

Since

$$\begin{aligned} &\mathbb{E} [\|F^{N,h}(t, \cdot) - F^{N,h}(s, \cdot)\|_{L^1}] \\ &\leq \mathbb{E} [\|F^{N,h}(t, \cdot) - F(t, \cdot)\|_{L^1}] + \|F(t, \cdot) - F(s, \cdot)\|_{L^1} + \mathbb{E} [\|F(s, \cdot) - F^{N,h}(s, \cdot)\|_{L^1}], \end{aligned}$$

using Lemma 4.4.12, we obtain:

$$\begin{aligned} \int_0^{t-h} \frac{h}{2(t-s)^{3/2}} \mathbb{E} [\|F^{N,h}(t, \cdot) - F^{N,h}(s, \cdot)\|_{L^1}] ds &\leq \sqrt{h} \mathbb{E} [\|F^{N,h}(t, \cdot) - F(t, \cdot)\|_{L^1}] \\ &\quad + h \int_0^{t-h} \frac{\mathbb{E} [\|F(s, \cdot) - F^{N,h}(s, \cdot)\|_{L^1}]}{2(t-s)^{3/2}} ds + h \int_0^t \left( Q \frac{\sqrt{t} - \sqrt{s}}{2(t-s)^{3/2}} - \frac{L_\Lambda \ln(t-s)}{2\sqrt{t-s}} \right) ds. \end{aligned}$$

To treat the last term of the right-hand side of the above inequality, we will use the fact that

$$\sup_{x>0} \{\sqrt{x}(2 - \ln(x))\} = 2.$$

$$\begin{aligned} \int_0^t \left( Q \frac{\sqrt{t} - \sqrt{s}}{2(t-s)^{3/2}} - \frac{L_\Lambda}{2} \frac{\ln(t-s)}{\sqrt{t-s}} \right) ds &= Q \int_0^1 \frac{1 - \sqrt{x}}{2(1-x)^{3/2}} dx - L_\Lambda \left( \sqrt{t} \ln(t) - 2\sqrt{t} \right) \\ &= Q \left( \left[ (1-x)^{-1/2} (1 - \sqrt{x}) \right]_0^1 + \int_0^1 \frac{dx}{2\sqrt{x}\sqrt{1-x}} \right) + L_\Lambda \sqrt{t} (2 - \ln(t)) \\ &= Q \left( \frac{\pi}{2} - 1 \right) + L_\Lambda \sqrt{t} (2 - \ln(t)) \\ &\leq Q \left( \frac{\pi}{2} - 1 \right) + 2L_\Lambda. \end{aligned}$$

Therefore,

$$\begin{aligned} &\mathbb{E} \left[ \left\| e_1^{N,h}(t, \cdot) \right\|_{L^1} \right] \\ &\leq C_1 \left( h + \sqrt{h} \mathbb{E} \left[ \left\| F^{N,h}(t, \cdot) - F(t, \cdot) \right\|_{L^1} \right] + \int_0^{t-h} \frac{h}{2(t-s)^{3/2}} \mathbb{E} \left[ \left\| F(s, \cdot) - F^{N,h}(s, \cdot) \right\|_{L^1} \right] ds \right) \end{aligned}$$

$$\text{where } C_1 = \frac{L_\Lambda}{\sigma} \sqrt{\frac{2}{\pi}} \left[ 1 \vee \left( 3\sqrt{(\sigma^2 + L_\Lambda^2 T)} + L_\Lambda(2 + \sqrt{T}) + Q \left( \frac{\pi}{2} - 1 \right) \right) \right].$$

□

Using Lemmas 4.4.7, 4.4.8, 4.4.9, 4.4.10 and 4.4.11 and the fact that for  $s \in [0, t-h]$ ,  $\frac{h}{2(t-s)^{3/2}} \leq \frac{1}{2\sqrt{t-s}}$ , we conclude the proof of Proposition 4.4.6 for the choice  $Z = 2 \max(L_\Lambda + C_1/2 + C_{4,5}, C_{2,3})$ .

**Remark 4.4.14.** In Lemma 4.4.13, we provide two estimations of  $\mathbb{E} \left[ \int_s^t \gamma_u^i d|K|_u \right]$ . If we only use the first estimation (4.4.14) in the proof of Lemma 4.4.11, we obtain, using a decomposition that we will detail in Section 4.4.3, a rough estimation of  $\mathbb{E} \left[ \left\| e_1^{N,h}(t, \cdot) \right\|_{L^1} \right]$  where we lose a  $\ln(h)$  factor.

### 4.4.3 Estimation of the bias

We recall Equation (4.4.7):

$$\begin{aligned} F^{N,h}(t, x) - F(t, x) &= G_t * \left( F_0^{N,h} - F_0 \right) (x) - \int_0^t \partial_x G_{t-s} * \left( \Lambda(F^{N,h}(s, \cdot)) - \Lambda(F(s, \cdot)) \right) (x) ds \\ &\quad + R^{N,h}(t, x) + E^{N,h}(t, x), \end{aligned}$$

and we shall use the expression of  $E^{N,h}(t, x)$  proved in Lemma 4.4.7. The next lemma provides an upper-bound of  $\left\| \mathbb{E} [E^{N,h}(t, \cdot)] \right\|_{L^1}$ .

**Lemma 4.4.15.** Assume that  $\lambda$  is Lipschitz continuous and the initial positions are

- either i.i.d. according to  $m$  and  $\int_{\mathbb{R}} \sqrt{F_0(x)(1 - F_0(x))} dx < \infty$ ,
- or optimal deterministic and  $\sup_{x \geq 1} x \int_x^{+\infty} (F_0(-y) + 1 - F_0(y)) dy < \infty$ .

Then

$$\exists Z_b < \infty, \forall N \in \mathbb{N}^*, \forall h \in (0, T], \quad \sup_{t \leq T} \left\| \mathbb{E} [E^{N,h}(t, \cdot)] \right\|_{L^1} \leq Z_b \left( \sqrt{\frac{h}{N}} + h \right).$$

*Proof.* To estimate  $\|\mathbb{E}[E^{N,h}(t, \cdot)]\|_{L^1}$ , we estimate each  $\|\mathbb{E}[e_p^{N,h}(t, \cdot)]\|_{L^1}$ ,  $p \in \llbracket 0, 5 \rrbracket$ . From Lemmas 4.4.8, 4.4.9 and 4.4.10 we have  $\|\mathbb{E}[E^{N,h}(t, \cdot)]\|_{L^1} \leq \|\mathbb{E}[e_1^{N,h}(t, \cdot)]\|_{L^1} + 2L_\Lambda h + 2C_{4,5}h$ . By Lemma 4.4.11 and Corollary 4.2.7, we have:

$$\begin{aligned} \|\mathbb{E}[e_1^{N,h}(t, \cdot)]\|_{L^1} &\leq \mathbb{E}[\|e_1^{N,h}(t, \cdot)\|_{L^1}] \\ &\leq C_1 h + C_1 \sqrt{h} \mathbb{E}[\|F^{N,h}(t, \cdot) - F(t, \cdot)\|_{L^1}] + C_1 \mathbf{1}_{\{t \geq h\}} \int_0^{t-h} \frac{h}{2(t-s)^{3/2}} \mathbb{E}[\|F^{N,h}(s, \cdot) - F(s, \cdot)\|_{L^1}] ds \\ &\leq C_1 h + C_1 \sqrt{h} C \left( \frac{1}{\sqrt{N}} + h \right) + C_1 \mathbf{1}_{\{t \geq h\}} \int_0^{t-h} \frac{h}{2(t-s)^{3/2}} C \left( \frac{1}{\sqrt{N}} + h \right) ds \\ &\leq C_1 h + C_1 C h^{3/2} + C_1 C \sqrt{\frac{h}{N}} + C_1 C \left( \frac{1}{\sqrt{N}} + h \right) \sqrt{h} \\ &\leq \left( C_1 + 2C_1 C \sqrt{T} \right) h + 2C_1 C \sqrt{\frac{h}{N}}. \end{aligned}$$

The conclusion holds with  $Z_b = \max(2L_\Lambda + 2C_{4,5} + C_1 + 2C_1 C \sqrt{T}, 2C_1 C)$ .  $\square$

The proof of Theorem 4.2.8 relies on the following Proposition that we will prove in Section 4.4.3.

**Proposition 4.4.16.** *Assume that  $\int_{\mathbb{R}} |x| m(dx) < \infty$  and  $\lambda$  is Lipschitz continuous. Then:*

$$\exists M_b < \infty, \forall N \in \mathbb{N}^*, \forall h \in (0, T], \quad \sup_{t \leq T} \mathbb{E}[\|F^{N,h}(t, \cdot) - F(t, \cdot)\|_{L^2}^2] \leq M_b \left( \frac{1}{N} + h \right).$$

#### Proof of Theorem 4.2.8

Taking the expectation of Equation (4.4.7) and using Lemma 4.4.5, we obtain that  $dx$  a.e.:

$$\begin{aligned} \mathbb{E}[F^{N,h}(t, x)] - F(t, x) &= G_t * \mathbb{E}[F_0^N(x) - F_0(x)] - \int_0^t \partial_x G_{t-s} * \mathbb{E}[(\Lambda(F^{N,h}(s, \cdot)) - \Lambda(F(s, \cdot)))(x)] ds \\ &\quad + \mathbb{E}[E^{N,h}(t, x)]. \end{aligned}$$

Besides, using Taylor-Young's inequality, we have that:

$$|\Lambda(F^{N,h}(s, \cdot)) - \Lambda(F(s, \cdot)) - \lambda(F(s, \cdot)) [F^{N,h}(s, \cdot) - F(s, \cdot)]| \leq \frac{L_\lambda}{2} |(F^{N,h}(s, \cdot) - F(s, \cdot))^2|$$

which implies:

$$\begin{aligned} &\|\mathbb{E}[\Lambda(F^{N,h}(s, \cdot)) - \Lambda(F(s, \cdot))]\|_{L^1} \\ &\leq \|\lambda(F(s, \cdot))\|_{L^\infty} \|\mathbb{E}[F^{N,h}(s, \cdot) - F(s, \cdot)]\|_{L^1} + \frac{L_\lambda}{2} \mathbb{E}[\|F^{N,h}(s, \cdot) - F(s, \cdot)\|_{L^2}^2]. \end{aligned}$$

Therefore, using the fact that  $G_t$  is a probability density and the estimate (B.2.4) from Lemma B.2.2, we obtain:

$$\begin{aligned} \|\mathbb{E}[F^{N,h}(t, \cdot)] - F(t, \cdot)\|_{L^1} &\leq \|\mathbb{E}[F_0^N] - F_0\|_{L^1} + \sqrt{\frac{2}{\pi\sigma^2}} \int_0^t \frac{1}{\sqrt{t-s}} \left\{ L_\lambda \|\mathbb{E}[F^{N,h}(s, \cdot)] - F(s, \cdot)\|_{L^1} \right. \\ &\quad \left. + \frac{L_\lambda}{2} \mathbb{E}[\|F^{N,h}(s, \cdot) - F(s, \cdot)\|_{L^2}^2] \right\} ds + \|\mathbb{E}[E^{N,h}(t, x)]\|_{L^1}. \end{aligned}$$

Using Lemma 4.4.15 and Proposition 4.4.16 then Young's inequality, we deduce that:

$$\begin{aligned}
& \|\mathbb{E} [F^{N,h}(t, \cdot)] - F(t, \cdot)\|_{L^1} \\
& \leq \|\mathbb{E} [F_0^N] - F_0\|_{L^1} + \sqrt{\frac{2}{\pi\sigma^2}} \int_0^t \frac{1}{\sqrt{t-s}} \left\{ L_\Lambda \|\mathbb{E} [F^{N,h}(s, \cdot)] - F(s, \cdot)\|_{L^1} + \frac{L_\Lambda}{2} M_b \left( \frac{1}{N} + h \right) \right\} ds \\
& \quad + Z_b \left( \sqrt{\frac{h}{N}} + h \right) \\
& \leq \|\mathbb{E} [F_0^N] - F_0\|_{L^1} + \left( \frac{1+\sqrt{2}}{2} Z_b + \frac{L_\Lambda M_b}{\sigma} \sqrt{\frac{2t}{\pi}} \right) \left( \frac{1}{N} + h \right) \\
& \quad + \frac{L_\Lambda}{\sigma} \sqrt{\frac{2}{\pi}} \int_0^t \frac{1}{\sqrt{t-s}} \|\mathbb{E} [F^{N,h}(s, \cdot)] - F(s, \cdot)\|_{L^1} ds.
\end{aligned}$$

We iterate this inequality and use that  $\int_0^t \frac{\sqrt{s}}{\sqrt{t-s}} ds = \frac{t\pi}{2}$  to obtain:

$$\begin{aligned}
\|\mathbb{E} [F^{N,h}(t, \cdot)] - F(t, \cdot)\|_{L^1} & \leq \left( \frac{1+\sqrt{2}}{2} Z_b + \frac{L_\Lambda M_b + (1+\sqrt{2})L_\Lambda Z_b}{\sigma} \sqrt{\frac{2t}{\pi}} + \frac{L_\Lambda L_\Lambda M_b t}{\sigma^2} \right) \left( \frac{1}{N} + h \right) \\
& \quad + \left( 1 + \frac{L_\Lambda}{\sigma} \sqrt{\frac{8t}{\pi}} \right) \|\mathbb{E} [F_0^N] - F_0\|_{L^1} + \frac{2L_\Lambda^2}{\sigma^2} \int_0^t \|\mathbb{E} [F^{N,h}(r, \cdot)] - F(r, \cdot)\|_{L^1} dr.
\end{aligned}$$

By Lemma 4.4.3, the application  $t \mapsto \|\mathbb{E} [F^{N,h}(t, \cdot)] - F(t, \cdot)\|_{L^1}$  is locally integrable  $\forall h \in [0, T], N \in \mathbb{N}^*$ . Therefore, we can apply Gronwall's lemma and choosing

$$C_b = \max \left( 1 + \frac{L_\Lambda}{\sigma} \sqrt{\frac{8T}{\pi}}, \frac{1+\sqrt{2}}{2} Z_b + \frac{L_\Lambda M_b + (1+\sqrt{2})L_\Lambda Z_b}{\sigma} \sqrt{\frac{2T}{\pi}} + \frac{4L_\Lambda L_\Lambda M_b T}{\sigma^2} \right) \exp \left( \frac{2L_\Lambda^2}{\sigma^2} T \right)$$

concludes the proof of the theorem.

#### Proof of Proposition 4.4.16

For all  $t, h \in [0, T], N \in \mathbb{N}^*$ , we use Jensen's inequality upon Equation (4.4.7) and obtain:

$$\begin{aligned}
& \mathbb{E} \left[ \|\mathbb{E} [F^{N,h}(t, \cdot)] - F(t, \cdot)\|_{L^2}^2 \right] \\
& \leq 4 \int_{\mathbb{R}} \mathbb{E} \left[ G_t * \left( F_0^{N,h} - F_0 \right)^2(x) \right] dx + 4 \int_{\mathbb{R}} \mathbb{E} [R^{N,h}(t, x)^2] dx + 4 \int_{\mathbb{R}} \mathbb{E} [E^{N,h}(t, x)^2] dx \\
& \quad + 4 \int_{\mathbb{R}} \mathbb{E} \left[ \left( \int_0^t \partial_x G_{t-s} * (\Lambda(F^{N,h}(s, \cdot)) - \Lambda(F(s, \cdot))) (x) ds \right)^2 \right] dx. \quad (4.4.17)
\end{aligned}$$

On the one hand, we have using the definition (4.4.5) of  $R^{N,h}(t, x)$ , Itô's isometry and the estimate (B.2.7) from Lemma B.2.2 that:

$$\int_{\mathbb{R}} \mathbb{E} [R^{N,h}(t, x)^2] dx = \frac{\sigma^2}{N^2} \sum_{i=1}^N \int_0^t \mathbb{E} \left[ \int_{\mathbb{R}} G_{t-s}^2 (X_s^{i,N,h} - x) dx \right] ds = \frac{\sigma}{N} \sqrt{\frac{t}{\pi}}.$$

On the other hand, using Minkowski's, Young's and Cauchy-Schwarz's inequalities in addition to the estimate (B.2.4) from Lemma B.2.2, we get:

$$\begin{aligned}
& \int_{\mathbb{R}} \mathbb{E} \left[ \left( \int_0^t \partial_x G_{t-s} * (\Lambda(F^{N,h}(s, \cdot)) - \Lambda(F(s, \cdot))) (x) ds \right)^2 \right] dx \\
&= \mathbb{E} \left[ \left\| \int_0^t \partial_x G_{t-s} * (\Lambda(F^{N,h}(s, \cdot)) - \Lambda(F(s, \cdot))) ds \right\|_{L^2}^2 \right] \\
&\leq \mathbb{E} \left[ \left( \int_0^t \|\partial_x G_{t-s}\|_{L^1} \|\Lambda(F^{N,h}(s, \cdot)) - \Lambda(F(s, \cdot))\|_{L^2} ds \right)^2 \right] \\
&\leq \mathbb{E} \left[ \left( \int_0^t \sqrt{\frac{2L_{\Lambda}^2}{\pi\sigma^2(t-s)}} \|F^{N,h}(s, \cdot) - F(s, \cdot)\|_{L^2} ds \right)^2 \right] \\
&\leq \frac{2L_{\Lambda}^2}{\pi\sigma^2} \int_0^t \frac{du}{\sqrt{t-u}} \int_0^t \frac{1}{\sqrt{t-s}} \mathbb{E} \left[ \|F^{N,h}(s, \cdot) - F(s, \cdot)\|_{L^2}^2 \right] ds \\
&= \frac{4L_{\Lambda}^2\sqrt{t}}{\pi\sigma^2} \int_0^t \frac{1}{\sqrt{t-s}} \mathbb{E} \left[ \|F^{N,h}(s, \cdot) - F(s, \cdot)\|_{L^2}^2 \right] ds.
\end{aligned}$$

Therefore, Inequality (4.4.17) becomes:

$$\begin{aligned}
\mathbb{E} \left[ \|F^{N,h}(t, \cdot) - F(t, \cdot)\|_{L^2}^2 \right] &\leq 4 \int_{\mathbb{R}} \mathbb{E} \left[ \left( F_0^{N,h} - F_0 \right)^2 (x) \right] dx + \frac{4\sigma}{N} \sqrt{\frac{t}{\pi}} + 4 \int_{\mathbb{R}} \mathbb{E} [E^{N,h}(t, x)^2] dx \\
&\quad + \frac{16L_{\Lambda}^2\sqrt{t}}{\pi\sigma^2} \int_0^t \frac{1}{\sqrt{t-s}} \mathbb{E} \left[ \|F^{N,h}(s, \cdot) - F(s, \cdot)\|_{L^2}^2 \right] ds.
\end{aligned}$$

As for the intialization term, when choosing either initial positions that are i.i.d. according to  $m$  or optimal deterministic, according to Lemma 4.5.1 and Remark 4.5.2, we have  $\int_{\mathbb{R}} \mathbb{E} \left[ \left( F_0^{N,h} - F_0 \right)^2 (x) \right] dx \leq \frac{1}{N} \int_{\mathbb{R}} |x| m(dx)$ . With Lemma 4.4.17 below which provides an estimation of the term  $\int_{\mathbb{R}} \mathbb{E} [E^{N,h}(t, x)^2] dx$ , we deduce that:

$$\begin{aligned}
& \mathbb{E} \left[ \|F^{N,h}(t, \cdot) - F(t, \cdot)\|_{L^2}^2 \right] \\
&\leq \frac{4}{N} \int_{\mathbb{R}} |x| m(dx) + \frac{4\sigma}{N} \sqrt{\frac{t}{\pi}} + 4Q_b h + \frac{16L_{\Lambda}^2\sqrt{t}}{\pi\sigma^2} \int_0^t \frac{1}{\sqrt{t-s}} \mathbb{E} \left[ \|F^{N,h}(s, \cdot) - F(s, \cdot)\|_{L^2}^2 \right] ds.
\end{aligned}$$

Iterating the previous inequality, we obtain:

$$\begin{aligned}
\mathbb{E} \left[ \|F^{N,h}(t, \cdot) - F(t, \cdot)\|_{L^2}^2 \right] &\leq 4 \left( 1 + \frac{32L_{\Lambda}^2 t}{\pi\sigma^2} \right) \left( \frac{1}{N} \int_{\mathbb{R}} |x| m(dx) + \frac{\sigma}{N} \sqrt{\frac{t}{\pi}} + Q_b h \right) \\
&\quad + \frac{256L_{\Lambda}^4 t}{\pi\sigma^4} \int_0^t \mathbb{E} \left[ \|F^{N,h}(r, \cdot) - F(r, \cdot)\|_{L^2}^2 \right] dr
\end{aligned}$$

By Lemma 4.4.3 and since  $|F^{N,h}(t, \cdot) - F(t, \cdot)| \leq 1$ , the function  $t \mapsto \mathbb{E} \left[ \|F^{N,h}(t, \cdot) - F(t, \cdot)\|_{L^2}^2 \right]$  is locally integrable for all  $h \in [0, T]$ ,  $N \in \mathbb{N}^*$ . We use Gronwall's lemma once again and conclude for the choice

$$M_b = 4 \max \left\{ Q_b \left( 1 + \frac{32L_{\Lambda}^2 T}{\pi\sigma^2} \right), \left( 1 + \frac{32L_{\Lambda}^2 t}{\pi\sigma^2} \right) \left( \int_{\mathbb{R}} |x| m(dx) + \sigma \sqrt{\frac{T}{\pi}} \right) \right\} \exp \left( \frac{256L_{\Lambda}^4 T^2}{\pi\sigma^4} \right).$$

**Lemma 4.4.17.** *Assume that  $\lambda$  is Lipschitz continuous. Then*

$$\exists Q_b, \forall h \in (0, T], \forall t \in [0, T], \quad \int_{\mathbb{R}} \mathbb{E} [E^{N,h}(t, x)^2] dx \leq Q_b h.$$

*Proof.* We have that  $\int_{\mathbb{R}} \mathbb{E} [E^{N,h}(t, x)^2] dx \leq 6 \sum_{p=0}^5 \int_{\mathbb{R}} \mathbb{E} [e_p^{N,h}(t, x)^2] dx$ . For this reason, we shall estimate, in what follows, each  $\int_{\mathbb{R}} \mathbb{E} [e_p^{N,h}(t, x)^2] dx, p \in \llbracket 0, 5 \rrbracket$ .

On the one hand, we have using Itô's isometry and the estimate (B.2.3) from Lemma B.2.2 that  $\forall h \in (0, T]$ :

$$\begin{aligned} \int_{\mathbb{R}} \mathbb{E} [e_2^{N,h}(t, x)^2] dx &= \frac{\sigma^2}{N^2} \sum_{i=1}^N \int_0^t (t \wedge \bar{\tau}_s^h - s)^2 (\lambda^N(i))^2 \mathbb{E} \left[ \int_{\mathbb{R}} (\partial_x G_{t-s} (Y_s^{i,N,h} - x))^2 dx \right] ds \\ &\leq \frac{L_\lambda^2}{4\sigma\sqrt{\pi}N} \int_0^t \frac{(t \wedge \bar{\tau}_s^h - s)^2}{(t-s)^{3/2}} ds \leq \frac{2L_\lambda^2}{3\sigma\sqrt{\pi}} \frac{h^{3/2}}{N}, \end{aligned}$$

where the last inequality has already been derived at the end of the proof of Lemma 4.4.10. The same estimation can be derived in the same way for  $\int_{\mathbb{R}} \mathbb{E} [e_3^{N,h}(t, x)^2] dx$ . The Cauchy-Schwarz inequality then a similar reasoning implies that, for  $r \in \{4, 5\}$ ,  $\int_{\mathbb{R}} \mathbb{E} [e_r^{N,h}(t, x)^2] dx \leq \frac{2L_\lambda^4 T}{3\sigma^3\sqrt{\pi}} h^{3/2}$ . As for the term  $e_0^{N,h}$ , we have using the estimate (B.2.6) that  $\int_{\mathbb{R}} \mathbb{E} [e_0^{N,h}(t, x)^2] dx \leq \frac{4L_\lambda^2}{\sigma\sqrt{\pi}} h^{3/2}$ .

On the other hand, using Cauchy-Schwarz's inequality twice, (B.2.3) and the estimation (4.4.14), we obtain:

$$\begin{aligned} \int_{\mathbb{R}} \mathbb{E} [e_1^{N,h}(t, x)^2] dx &= \mathbb{E} \left[ \int_{\mathbb{R}} \left( \frac{1}{N} \sum_{i=2}^N \int_0^t (t \wedge \bar{\tau}_s^h - s) (\lambda^N(i) - \lambda^N(i-1)) \partial_x G_{t-s} (Y_s^{i,N,h} - x) \gamma_s^i d|K|_s \right)^2 dx \right] \\ &\leq \mathbb{E} \left[ \int_{\mathbb{R}} \frac{1}{N} \sum_{i=2}^N \left( \int_0^t (t \wedge \bar{\tau}_s^h - s)^2 (\lambda^N(i) - \lambda^N(i-1))^2 (\partial_x G_{t-s})^2 (Y_s^{i,N,h} - x) \gamma_s^i d|K|_s \right) \left( \int_0^t \gamma_r^i d|K|_r \right) dx \right] \\ &\leq \frac{L_\lambda^2}{4\sigma^3\sqrt{\pi}} \mathbb{E} \left[ \frac{1}{N^3} \sum_{i=2}^N \left( \int_0^t \frac{(t \wedge \bar{\tau}_s^h - s)^2}{(t-s)^{3/2}} \gamma_s^i d|K|_s \right) \left( \int_0^t \gamma_r^i d|K|_r \right) \right] \\ &\leq \frac{L_\lambda^2}{4\sigma^3\sqrt{\pi}} \frac{1}{N^3} \sum_{i=2}^N \mathbb{E}^{1/2} \left[ \left( \int_0^t \frac{(t \wedge \bar{\tau}_s^h - s)^2}{(t-s)^{3/2}} \gamma_s^i d|K|_s \right)^2 \right] \mathbb{E}^{1/2} \left[ \left( \int_0^t \gamma_r^i d|K|_r \right)^2 \right] \\ &\leq \frac{3L_\lambda^2}{4\sigma^3} \sqrt{\frac{t(\sigma^2 + L_\lambda^2 T)}{\pi}} \frac{1}{N^2} \sum_{i=2}^N \mathbb{E}^{1/2} \left[ \left( \int_0^t \frac{(t \wedge \bar{\tau}_s^h - s)^2}{(t-s)^{3/2}} \gamma_s^i d|K|_s \right)^2 \right]. \end{aligned}$$

We denote by  $m = \lceil \log_2(t/h) \rceil$  and rewrite the integral the following way:

$$\int_0^t \frac{(t \wedge \bar{\tau}_s^h - s)^2}{(t-s)^{3/2}} \gamma_s^i d|K|_s = \sum_{k=0}^{m-1} \int_{t-t/2^k}^{t-t/2^{k+1}} \frac{(t \wedge \bar{\tau}_s^h - s)^2}{(t-s)^{3/2}} \gamma_s^i d|K|_s + \int_{t-t/2^m}^t \frac{(t \wedge \bar{\tau}_s^h - s)^2}{(t-s)^{3/2}} \gamma_s^i d|K|_s.$$



Therefore,

$$\begin{aligned}
& \mathbb{E}^{1/2} \left[ \left( \int_0^t \frac{(t \wedge \bar{\tau}_s^h - s)^2}{(t-s)^{3/2}} \gamma_s^i d|K|_s \right)^2 \right] \\
& \leq \sum_{k=0}^{m-1} \frac{h^2}{\left(\frac{t}{2^{k+1}}\right)^{3/2}} \mathbb{E}^{1/2} \left[ \left( \int_{t-2^{k+1}}^{t-2^k} \gamma_s^i d|K|_s \right)^2 \right] + \sqrt{h} \mathbb{E}^{1/2} \left[ \left( \int_{t-2^m}^t \gamma_s^i d|K|_s \right)^2 \right] \\
& \leq 3N \sqrt{(\sigma^2 + L_\Lambda^2 T)} \left( h^2 \sum_{k=0}^{m-1} \left( \frac{t}{2^{k+1}} \right)^{-1} + h \right) \\
& = 3N \sqrt{(\sigma^2 + L_\Lambda^2 T)} \left( \frac{2h^2}{t} (2^m - 1) + h \right) \\
& \leq 15Nh \sqrt{(\sigma^2 + L_\Lambda^2 T)}.
\end{aligned}$$

We then have  $\int_{\mathbb{R}} \mathbb{E} \left[ e_1^{N,h}(t, x)^2 \right] dx \leq \frac{45L_\Lambda^2 (\sigma^2 + L_\Lambda^2 T)}{4\sigma^3} \sqrt{\frac{T}{\pi}} h.$

The conclusion holds for the choice  $Q_b = \frac{1}{\sigma} \sqrt{\frac{T}{\pi}} \left\{ \frac{16L_\Lambda^2}{3} \left( 1 + \frac{L_\Lambda^2 T}{4\sigma^2} \right) + \frac{45L_\Lambda^2}{4} \left( 1 + \frac{L_\Lambda^2 T}{\sigma^2} \right) \right\}.$   $\square$

## 4.5 Particle initialization

In this section, we are interested in the strong and weak initialization errors  $\mathbb{E} [\mathcal{W}_1(\mu_0^N, m)]$  and  $\mathcal{W}_1(\mathbb{E} [\mu_0^N], m)$ . For initial positions i.i.d. according to  $m$ ,  $\mathbb{E} [\hat{\mu}_0^N] = m$  and the weak error is zero. For optimal deterministic initial positions, both are equal to  $\mathcal{W}_1(\tilde{\mu}_0^N, m)$ . In Section 4.5.1, we check that the strong error is bounded if and only if  $m$  has a finite first order moment. Section 4.5.2 is devoted to the proof of Proposition 4.2.4 which gives conditions for the strong initialization error to be of order  $N^{-1/2}$ . Last, in Section 4.5.3, we state necessary conditions for the optimal deterministic initialization error to be of order  $N^{-1}$  and study the asymptotic behaviour of moments under this initialization as  $N \rightarrow \infty$ .

### 4.5.1 Finite Wassertein distance

We recall that  $F_0(x) = m((-\infty, x])$  is the cumulative distribution function of the probability measure  $m$  on the real line and  $F_0^{-1}(u) = \inf\{x \in \mathbb{R} : F_0(x) \geq u\}$ ,  $u \in (0, 1)$  its quantile function. The following lemma implies that the finiteness of  $\int_{\mathbb{R}} |x| m(dx)$  is a necessary and sufficient condition for the finiteness of  $\sup_{N \in \mathbb{N}^*} \mathbb{E} [\mathcal{W}_1(\mu_0^N, m)]$  and a sufficient condition for the finiteness of  $\sup_{N \in \mathbb{N}^*} N \mathbb{E} [\|F_0^N - F_0\|_{L^2}^2]$ .

**Lemma 4.5.1.** *For the optimal deterministic initial positions, we have*

$$\forall N \in \mathbb{N}^*, \mathcal{W}_1(\tilde{\mu}_0^N, m) \leq \int_{\mathbb{R}} |x| m(dx) \text{ and } \|\tilde{F}_0^N - F_0\|_{L^2}^2 \leq \frac{1}{2N} \int_{\mathbb{R}} |x| m(dx).$$

*For initial positions i.i.d. according to  $m$ , we have*

$$\forall N \in \mathbb{N}^*, \mathbb{E} [\mathcal{W}_1(\hat{\mu}_0^N, m)] \leq 2 \int_{\mathbb{R}} F_0(x)(1 - F_0(x)) dx \text{ and } \mathbb{E} [\|\hat{F}_0^N - F_0\|_{L^2}^2] = \frac{1}{N} \int_{\mathbb{R}} F_0(x)(1 - F_0(x)) dx.$$

*Conversely, the existence of  $N \in \mathbb{N}^*$  such that  $\mathcal{W}_1(\tilde{\mu}_0^N, m)$  or  $\mathbb{E} [\mathcal{W}_1(\hat{\mu}_0^N, m)]$  is finite implies the finiteness of  $\int_{\mathbb{R}} |x| m(dx)$ .*

**Remark 4.5.2.** *Notice that  $\int_{\mathbb{R}} F_0(x)(1 - F_0(x)) dx \leq \int_{-\infty}^0 F_0(x) dx + \int_0^{+\infty} (1 - F_0(x)) dx = \int_{\mathbb{R}} |x| m(dx)$  (see (4.5.6) below for a short proof of this well-known equality). On the other hand, since*

$$\int_{\mathbb{R}} F_0(x)(1 - F_0(x)) dx \geq \frac{1}{2} \int_{-\infty}^{F_0^{-1}(1/2)} F_0(x) dx + \frac{1}{2} \int_{F_0^{-1}(1/2)}^{+\infty} (1 - F_0(x)) dx \geq \frac{1}{2} \left( \int_{\mathbb{R}} |x| m(dx) - |F_0^{-1}(1/2)| \right),$$

$\int_{\mathbb{R}} F_0(x)(1 - F_0(x)) dx$  and  $\int_{\mathbb{R}} |x|m(dx)$  are simultaneously finite (or infinite).

*Proof.* We recall that  $F_0(x) = m((-\infty, x])$  is the cumulative distribution function of the probability measure  $m$  on the real line and  $F_0^{-1}(u) = \inf\{x \in \mathbb{R} : F_0(x) \geq u\}$ ,  $u \in (0, 1)$  its quantile function.

► Since for  $i \in \{1, \dots, N\}$ ,  $x_i^N = F_0^{-1}\left(\frac{2i-1}{2N}\right)$  minimizes  $\mathbb{R} \ni y \mapsto \int_{\frac{i-1}{N}}^{\frac{i}{N}} |y - F_0^{-1}(u)| du$ , we have that:

$$\mathcal{W}_1(\tilde{\mu}_0^N, m) = \sum_{i=1}^N \int_{\frac{i-1}{N}}^{\frac{i}{N}} |x_i^N - F_0^{-1}(u)| du \leq \sum_{i=1}^N \int_{\frac{i-1}{N}}^{\frac{i}{N}} |F_0^{-1}(u)| du = \int_{\mathbb{R}} |x|m(dx).$$

Since  $|\tilde{F}_0^N - F_0|$  is not greater than  $1/2N$ , we deduce that

$$\left\| \tilde{F}_0^N - F_0 \right\|_{L^2}^2 \leq \frac{\left\| \tilde{F}_0^N - F_0 \right\|_{L^1}^2}{2N} = \frac{\mathcal{W}_1(\tilde{\mu}_0^N, m)^2}{2N} \leq \frac{1}{2N} \int_{\mathbb{R}} |x|m(dx).$$

On the other hand,

$$\begin{aligned} \int_{\mathbb{R}} |x|m(dx) &= \int_0^1 |F_0^{-1}(u)| du \leq \int_0^1 \left| F_0^{-1}(u) - \left( \tilde{F}_0^N \right)^{-1}(u) \right| du + \int_0^1 \left| \left( \tilde{F}_0^N \right)^{-1}(u) \right| du \\ &= \mathcal{W}_1(\tilde{\mu}_0^N, m) + \frac{1}{N} \sum_{i=1}^N \left| F_0^{-1}\left(\frac{2i-1}{2N}\right) \right|. \end{aligned}$$

Since the last sum is finite, the finiteness of  $\mathcal{W}_1(\tilde{\mu}_0^N, m)$  implies that  $\int_{\mathbb{R}} |x|m(dx)$  is finite.

► When choosing initial positions i.i.d. according to  $m$ , we first have the following results:

- Since  $\hat{F}_0^N(x) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{X_0^i \leq x\}}$  where the random variables  $(\mathbf{1}_{\{X_0^i \leq x\}})_{i \geq 1}$  are i.i.d. according to the Bernoulli law with parameter  $F_0(x)$  and variance  $F_0(x)(1 - F_0(x))$ ,

$$\mathbb{E} \left[ \left| \hat{F}_0^N(x) - F_0(x) \right| \right] \leq \mathbb{E}^{1/2} \left[ \left( \hat{F}_0^N(x) - F_0(x) \right)^2 \right] = \sqrt{\frac{F_0(x)(1 - F_0(x))}{N}}, \quad (4.5.1)$$

The equality ensures that  $\mathbb{E} \left[ \left\| \hat{F}_0^N - F_0 \right\|_{L^2}^2 \right] = \frac{\int_{\mathbb{R}} F_0(x)(1 - F_0(x)) dx}{N}$ .

- When  $F_0(x) \leq \frac{1}{N}$ , since  $N\hat{F}_0^N(x)$  is distributed according to the binomial law with parameter  $(N, F_0(x))$  and expectation  $NF_0(x)$ , one has

$$\begin{aligned} \mathbb{E} \left[ \left| \hat{F}_0^N(x) - F_0(x) \right| \right] &= F_0(x) \mathbb{P} \left( \hat{F}_0^N(x) = 0 \right) + \sum_{k=1}^N \left( \frac{k}{N} - F_0(x) \right) \mathbb{P} \left( \hat{F}_0^N(x) = \frac{k}{N} \right) \\ &= F_0(x) \left( \mathbb{P} \left( \hat{F}_0^N(x) = 0 \right) - \left( 1 - \mathbb{P} \left( \hat{F}_0^N(x) = 0 \right) \right) \right) + \mathbb{E} \left[ \hat{F}_0^N(x) \right] \\ &= 2F_0(x) \mathbb{P} \left( \hat{F}_0^N(x) = 0 \right) \\ &= 2F_0(x) (1 - F_0(x))^N. \end{aligned} \quad (4.5.2)$$

- When  $F_0(x) \geq 1 - \frac{1}{N}$ , we obtain in a symmetric way that:

$$\mathbb{E} \left[ \left| \hat{F}_0^N(x) - F_0(x) \right| \right] = 2(1 - F_0(x)) (F_0(x))^N. \quad (4.5.3)$$

Using these results for the first inequality then the fact that when  $1/N < F_0(x) < 1 - 1/N$ , then  $\frac{1}{\sqrt{N}} \leq \sqrt{2F_0(x)(1 - F_0(x))}$  for the second one, we obtain that

$$\begin{aligned} \mathbb{E} [\mathcal{W}_1(\hat{\mu}_0^N, m)] &\leq \int_{\mathbb{R}} \mathbf{1}_{\{1/N < F_0(x) < 1-1/N\}} \sqrt{\frac{F_0(x)(1 - F_0(x))}{N}} dx \\ &\quad + \int_{\mathbb{R}} (\mathbf{1}_{\{F_0(x) \leq 1/N\}} + \mathbf{1}_{\{F_0(x) \geq 1-1/N\}}) 2F_0(x)(1 - F_0(x)) dx \\ &\leq \int_{\mathbb{R}} \mathbf{1}_{\{1/N < F_0(x) < 1-1/N\}} \sqrt{2} F_0(x)(1 - F_0(x)) dx \\ &\quad + \int_{\mathbb{R}} (\mathbf{1}_{\{F_0(x) \leq 1/N\}} + \mathbf{1}_{\{F_0(x) \geq 1-1/N\}}) 2F_0(x)(1 - F_0(x)) dx \\ &\leq 2 \int_{\mathbb{R}} F_0(x)(1 - F_0(x)) dx. \end{aligned}$$

On the other hand, using once again Equations (4.5.2) and (4.5.3), we have that:

$$\mathbb{E} [\mathcal{W}_1(\hat{\mu}_0^N, m)] \geq 2 \left(1 - \frac{1}{N}\right)^N \left( \int_{-\infty}^0 \mathbf{1}_{\{F_0(x) \leq \frac{1}{N}\}} F_0(x) dx + \int_0^{+\infty} \mathbf{1}_{\{F_0(x) \geq 1-\frac{1}{N}\}} (1 - F_0(x)) dx \right).$$

With the inequality

$$\begin{aligned} &\int_{-\infty}^0 \mathbf{1}_{\{F_0(x) > \frac{1}{N}\}} F_0(x) dx + \int_0^{+\infty} \mathbf{1}_{\{F_0(x) < 1-\frac{1}{N}\}} (1 - F_0(x)) dx \\ &\leq \int_{-\infty}^0 \mathbf{1}_{\{F_0(x) \geq \frac{1}{N}\}} dx + \int_0^{+\infty} \mathbf{1}_{\{F_0(x) < 1-\frac{1}{N}\}} dx = (F_0^{-1}(1/N))^- + (F_0^{-1}(1 - 1/N))^+ < \infty, \end{aligned}$$

we conclude that  $\infty > \mathbb{E} [\mathcal{W}_1(\hat{\mu}_0^N, m)]$  implies that  $\infty > \int_{-\infty}^0 F_0(x) dx + \int_0^{+\infty} (1 - F_0(x)) dx = \int_{\mathbb{R}} |x|m(dx)$ .  $\square$

When  $m(dx) = \delta_y(dx)$  for some  $y \in \mathbb{R}$ , then  $F_0^{-1}(\frac{2i-1}{2N}) = y$  for all  $1 \leq i \leq N$  and  $\mathcal{W}_1(\hat{\mu}_0^N, m) = 0$  for each  $N \geq 1$ . Otherwise,  $\int_{\mathbb{R}} (F_0(x) \wedge (1 - F_0(x))) dx > 0$  and, according to the next Lemma, the strong error of the optimal deterministic initialization error cannot behave better than  $\mathcal{O}(N^{-1})$ .

**Lemma 4.5.3.**

$$\forall N \geq 1, \quad N \int_{\mathbb{R}} |\tilde{F}_0^N(x) - F_0(x)| dx + (N+1) \int_{\mathbb{R}} |\tilde{F}_0^{N+1}(x) - F_0(x)| dx \geq \frac{1}{2} \int_{\mathbb{R}} (F_0(x) \wedge (1 - F_0(x))) dx. \quad (4.5.4)$$

As  $\int_{\mathbb{R}} |x|m(dx) = \int_{-\infty}^0 F_0(x) dx + \int_0^{+\infty} (1 - F_0(x)) dx$ ,  $\int_{\mathbb{R}} |x|m(dx) < \infty \Leftrightarrow \int_{\mathbb{R}} (F_0(x) \wedge (1 - F_0(x))) dx < \infty$ . So the above statement can be seen as a refinement of the necessary condition in Lemma 4.5.1.

*Proof.* Formula (4.2.5) rewrites:

$$\int_{\mathbb{R}} |\tilde{F}_0^N(x) - F_0(x)| dx = \frac{1}{N} \int_{\mathbb{R}} \min_{j \in \mathbb{N}} |NF_0(x) - j| dx. \quad (4.5.5)$$

For  $v \in (0, 1)$ ,

- Either  $\lfloor Nv \rfloor \leq Nv < (N+1)v \leq \lfloor Nv \rfloor + 1$  which implies that  $(Nv - \lfloor Nv \rfloor) \vee (\lfloor Nv \rfloor + 1 - (N+1)v) \geq \frac{1-v}{2}$  while  $\lfloor Nv \rfloor + 1 - Nv = \lfloor Nv \rfloor + 1 - (N+1)v + v \geq v$  and  $(N+1)v - \lfloor Nv \rfloor = Nv - \lfloor Nv \rfloor + v \geq v$  so that

$$\min_{j \in \mathbb{N}} |Nv - j| \vee \min_{j \in \mathbb{N}} |(N+1)v - j| \geq v \wedge \frac{1-v}{2}.$$

- Or  $\lfloor Nv \rfloor \leq Nv < \lfloor Nv \rfloor + 1 \leq (N+1)v$ , which implies that  $(\lfloor Nv \rfloor + 1 - Nv) \vee ((N+1)v - (\lfloor Nv \rfloor + 1)) \geq \frac{v}{2}$  while  $Nv - \lfloor Nv \rfloor = (N+1)v - (\lfloor Nv \rfloor + 1) + 1 - v \geq 1 - v$  and  $\lfloor Nv \rfloor + 2 - (N+1)v = \lfloor Nv \rfloor + 1 - Nv + 1 - v > 1 - v$  so that

$$\min_{j \in \mathbb{N}} |Nv - j| \vee \min_{j \in \mathbb{N}} |(N+1)v - j| \geq \frac{v}{2} \wedge (1 - v).$$

Synthetising the two cases and remarking that the inequality still holds for  $v \in \{0, 1\}$ , we deduce that:

$$\forall v \in [0, 1], \forall N \geq 1, \quad \min_{j \in \mathbb{N}} |Nv - j| \vee \min_{j \in \mathbb{N}} |(N+1)v - j| \geq \frac{v \wedge (1 - v)}{2}.$$

Inserting this inequality with  $v = F_0(x)$  into (4.5.5), we conclude that for each  $N \geq 1$ ,

$$N \int_{\mathbb{R}} \left| \tilde{F}_0^N(x) - F_0(x) \right| dx + (N+1) \int_{\mathbb{R}} \left| \tilde{F}_0^{N+1}(x) - F_0(x) \right| dx \geq \frac{1}{2} \int_{\mathbb{R}} (F_0(x) \wedge (1 - F_0(x))) dx.$$

□

## 4.5.2 Strong errors of order $N^{-1/2}$

In this section, we shall prove each implication in Proposition 4.2.4. Since  $\mathcal{W}_1(\hat{\mu}_0^N, m) = \left\| \hat{F}_0^N(\cdot) - F_0(\cdot) \right\|_{L^1}$  the equivalence concerning the strong random initialization error  $\mathbb{E}[\mathcal{W}_1(\hat{\mu}_0^N, m)]$  is a direct consequence of the following lemma.

**Lemma 4.5.4.** *One has:*

$$\begin{aligned} \sqrt{N} \mathbb{E} \left[ \left\| \hat{F}_0^N(\cdot) - F_0(\cdot) \right\|_{L^1} \right] &\leq \int_{\mathbb{R}} \sqrt{F_0(x)(1 - F_0(x))} dx, \\ \lim_{N \rightarrow \infty} \sqrt{N} \mathbb{E} \left[ \left\| \hat{F}_0^N(\cdot) - F_0(\cdot) \right\|_{L^1} \right] &= \sqrt{2/\pi} \int_{\mathbb{R}} \sqrt{F_0(x)(1 - F_0(x))} dx. \end{aligned}$$

**Remark 4.5.5.** *Let us illustrate by an example how to derive slower rates of convergence under weaker integrability conditions. For the Pareto law  $m(dx) = \mathbf{1}_{\{x \geq 1\}} \frac{\alpha dx}{x^{1+\alpha}}$  with  $\alpha > 0$ ,  $(1 - F_0(x)) = \mathbf{1}_{\{x \geq 1\}} x^{-\alpha}$ .*

*With (4.5.3) and (4.5.1), we deduce that for  $N \geq 2$  so that  $2(1 - \frac{1}{N})^{N-1} \leq 1$ ,*

$$\begin{aligned} \int_{\mathbb{R}} \mathbb{E} \left[ \left| \hat{F}_0^N(x) - F_0(x) \right| \right] dx &\leq \int_{\mathbb{R}} \left( \mathbf{1}_{\{F_0(x) < 1 - \frac{1}{N}\}} \sqrt{\frac{1 - F_0(x)}{N}} + \mathbf{1}_{\{F_0(x) \geq 1 - \frac{1}{N}\}} (1 - F_0(x)) \right) dx \\ &= \int_1^{N^{1/\alpha}} \frac{x^{-\alpha/2}}{\sqrt{N}} dx + \int_{N^{1/\alpha}}^{+\infty} x^{-\alpha} dx \\ &= \frac{2(N^{-1+1/\alpha} - N^{-1/2})}{2 - \alpha} + \frac{N^{-1+1/\alpha}}{\alpha - 1} \\ &= \frac{\alpha N^{-1+1/\alpha}}{(2 - \alpha)(\alpha - 1)} + \frac{2N^{-1/2}}{\alpha - 2}. \end{aligned}$$

*We conclude that, for  $\alpha \in (1, 2)$ ,  $\int_{\mathbb{R}} \mathbb{E} \left[ \left| \hat{F}_0^N(x) - F_0(x) \right| \right] dx = \mathcal{O}(N^{-1+1/\alpha})$  using that its lower bound  $\int_{\mathbb{R}} \left| \tilde{F}_0^N(x) - F_0(x) \right| dx$  also is  $\mathcal{O}(N^{-1+1/\alpha})$  according to Remark 2.2 [49].*

Let us now prove Lemma 4.5.4.

*Proof.* The first assertion is an immediate consequence of (4.5.1). Moreover, the central limit theorem implies that, for each  $x \in \mathbb{R}$ ,  $\mathbb{E} \left[ \sqrt{N} \left| \hat{F}_0^N(x) - F_0(x) \right| \right]$  converges to  $\sqrt{2F_0(x)(1 - F_0(x))/\pi}$  as  $N \rightarrow \infty$ . When  $\int_{\mathbb{R}} \sqrt{F_0(x)(1 - F_0(x))} dx = +\infty$ , one concludes by applying Fatou's lemma to the spatial integral with respect to  $dx$ . Otherwise, one concludes by Lebesgue's theorem, using domination deduced from the inequality (4.5.1). □

Let us now check the implications involving the finiteness of  $\int_{\mathbb{R}} \sqrt{F_0(x)(1-F_0(x))} dx$ .

**Lemma 4.5.6.**

$$\int_{\mathbb{R}} |x|^{2+} m(dx) < \infty \Rightarrow \int_{\mathbb{R}} \sqrt{F_0(x)(1-F_0(x))} dx < \infty \Rightarrow \int_{\mathbb{R}} x^2 m(dx) < \infty.$$

**Remark 4.5.7.** *These implications have been demonstrated by Bobkov-Ledoux (see for instance Section 3.1 in [7]), but we are leaving the proof that we have achieved before realizing that it has already been done.*

The following examples show that the assertions are not equivalent.

**Example 4.5.8.** Let  $m(dx) = \mathbf{1}_{\{x \geq 2\}} \frac{c}{x^3 \ln^2(x)} dx$  where  $\frac{1}{c} = \int_2^{+\infty} \frac{dx}{x^3 \ln^2(x)}$ .

Since  $\int_2^{+\infty} \frac{dx}{x \ln^2(x)} = \int_{\ln(2)}^{+\infty} \frac{dy}{y^2} = \frac{1}{\ln(2)} < \infty$ , one has  $\int_{\mathbb{R}} x^2 m(dx) < \infty$ . On the other hand, for  $x \geq 2$ , one has, by integration by parts,  $\frac{1-F_0(x)}{c} = \int_x^{+\infty} \frac{dy}{y^3 \ln^2(y)} = \frac{1}{2x^2 \ln^2(x)} - \int_x^{+\infty} \frac{dy}{y^3 \ln^3(y)}$ . Since  $0 \leq \int_x^{+\infty} \frac{dy}{y^3 \ln^3(y)} \leq \frac{1}{\ln(x)} \int_x^{+\infty} \frac{dy}{y^3 \ln^2(y)}$ , we deduce that, as  $x \rightarrow \infty$ ,  $\frac{1-F_0(x)}{c} \sim \frac{1}{2x^2 \ln^2(x)}$  so that  $\sqrt{F_0(x)(1-F_0(x))} \sim \frac{\sqrt{c/2}}{x \ln(x)}$ . We conclude that  $\int_{\mathbb{R}} \sqrt{F_0(x)(1-F_0(x))} dx = +\infty$ .

For  $m(dx) = \mathbf{1}_{\{x \geq 2\}} \frac{c}{x^3 \ln^3(x)} dx$  where  $\frac{1}{c} = \int_2^{+\infty} \frac{dx}{x^3 \ln^3(x)}$ , one has, by similar computations,  $\frac{1-F_0(x)}{c} \sim \frac{1}{2x^2 \ln^3(x)}$  as  $x \rightarrow +\infty$  so that  $\int_{\mathbb{R}} \sqrt{F_0(x)(1-F_0(x))} dx < \infty$  whereas  $\int_{\mathbb{R}} |x|^{2+\varepsilon} m(dx) = +\infty$  for each  $\varepsilon > 0$ .

For the sake of completeness, we are going to reproduce the short proof of the first implication in Lemma 4.5.6 given in Remark 2.2 [49]. The proof of the second implication relies on the following result, the proof of which is postponed.

**Lemma 4.5.9.** *One has*

$$\forall y \in \mathbb{R}, \frac{1}{2} \int_{\mathbb{R}} (x-y)^2 m(dx) \leq \left( \int_{-\infty}^y \sqrt{F_0(x)} dx \right)^2 + \left( \int_y^{+\infty} \sqrt{1-F_0(x)} dx \right)^2.$$

Let us prove Lemma 4.5.6.

*Proof.* Since for  $y = F_0^{-1}(1/2)$ ,  $\sqrt{2} \int_{\mathbb{R}} \sqrt{F_0(x)(1-F_0(x))} dx \geq \int_{-\infty}^y \sqrt{F_0(x)} dx + \int_y^{+\infty} \sqrt{1-F_0(x)} dx$ ,

the second implication is easily deduced from Lemma 4.5.9.

Let  $\varepsilon \geq -1$ . By Fubini's theorem,

$$\begin{aligned} \int_{-\infty}^0 |x|^{1+\varepsilon} F_0(x) dx + \int_0^{+\infty} |x|^{1+\varepsilon} (1-F_0(x)) dx \\ = \int_{-\infty}^0 |x|^{1+\varepsilon} \int_{-\infty}^0 \mathbf{1}_{\{y \leq x\}} m(dy) dx + \int_0^{+\infty} |x|^{1+\varepsilon} \int_0^{+\infty} \mathbf{1}_{\{y > x\}} m(dy) dx \\ = \frac{1}{2+\varepsilon} \int_{\mathbb{R}} |y|^{2+\varepsilon} m(dy). \end{aligned} \quad (4.5.6)$$

By the Cauchy-Schwarz inequality,

$$\left( \int_{\mathbb{R}} \sqrt{F_0(x)(1-F_0(x))} dx \right)^2 \leq \int_{\mathbb{R}} \frac{dx}{1+|x|^{1+\varepsilon}} \int_{\mathbb{R}} (1+|x|^{1+\varepsilon}) F_0(x) (1-F_0(x)) dx,$$

where the first integral in the right-hand side is finite when  $\varepsilon > 0$  and, according to (4.5.6), the second one is finite when  $\int_{\mathbb{R}} |x|^{2+\varepsilon} m(dx) < \infty$ .  $\square$

The proof of Lemma 4.5.9 relies on the following integral formulas for the square roots of the cumulative distribution function and the survival function.

**Lemma 4.5.10.** *Let  $m \in \mathcal{P}(\mathbb{R})$  with cumulative distribution function  $F_0(x)$ ,  $x \in \mathbb{R}$ . Then  $\forall x \in \mathbb{R}$ ,*

$$\sqrt{F_0(x)} = \int_{u=0}^1 \int_{\mathbb{R}} \frac{\mathbf{1}_{\{z \leq x\}}}{2\sqrt{F_0(z-) + u(F_0(z) - F_0(z-))}} m(dz) du, \quad (4.5.7)$$

$$\sqrt{1 - F_0(x)} = \int_{u=0}^1 \int_{\mathbb{R}} \frac{\mathbf{1}_{\{z > x\}}}{2\sqrt{1 - F_0(z-) - u(F_0(z) - F_0(z-))}} m(dz) du. \quad (4.5.8)$$

Let us show (4.5.7) ((4.5.8) is obtained by a symmetric reasoning) before checking Lemma 4.5.9.

*Proof.* If  $F_0(x) = 0$ , then  $m((-\infty, x]) = 0$  and although  $\int_{u=0}^1 \frac{du}{2\sqrt{F_0(z-) + u(F_0(z) - F_0(z-))}} = +\infty$  for  $z \in (-\infty, x]$ , the integral in the right-hand side of (4.5.7) is equal to 0 by the usual convention in measure theory. If the limit  $x_0 := F_0^{-1}(0+)$  of the left-continuous function  $F_0^{-1}$  at point 0 is larger than  $-\infty$  and such that  $m(\{x_0\}) > 0$  then since  $F_0(x_0-) = m((-\infty, x_0)) = 0$ , one has:

$$\sqrt{F_0(x_0)} = \sqrt{F_0(x_0)} - \sqrt{F_0(x_0-)} = m(\{x_0\}) \int_0^1 \frac{du}{2\sqrt{F_0(x_0-) + u(F_0(x_0) - F_0(x_0-))}},$$

with the right-hand side equal to the one of (4.5.7) with  $x = x_0$ . So, as soon as  $x_0 > -\infty$ , (4.5.7) holds for  $x = x_0$ .

It is enough to deal with the case  $x > x_0$  to conclude the proof. Let now  $\varphi : [0, 1] \rightarrow \mathbb{R}$  be  $C^1$ . The chain rule for càdlàg functions with finite variation (see for instance Proposition 4.6 Chapter 0 [78]) writes

$$\begin{aligned} d\varphi(F_0(z)) &= \varphi'(F_0(z-))dF_0(z) + \varphi(F_0(z)) - \varphi(F_0(z-)) - \varphi'(F_0(z-))(F_0(z) - F_0(z-)) \\ &= \mathbf{1}_{\{F_0(z-) = F_0(z)\}} \varphi'(F_0(z-))dF_0(z) + \mathbf{1}_{\{F_0(z-) < F_0(z)\}} \frac{\varphi(F_0(z)) - \varphi(F_0(z-))}{F_0(z) - F_0(z-)} dF_0(z) \end{aligned}$$

where  $dF_0(z) = m(dz)$ . Since

$$\int_{u=0}^1 \varphi'(F_0(z-) + u(F_0(z) - F_0(z-))) du = \mathbf{1}_{\{F_0(z-) = F_0(z)\}} \varphi'(F_0(z-)) + \mathbf{1}_{\{F_0(z-) < F_0(z)\}} \frac{\varphi(F_0(z)) - \varphi(F_0(z-))}{F_0(z) - F_0(z-)},$$

we deduce that:

$$d\varphi(F_0(z)) = \int_{u=0}^1 \varphi'(F_0(z-) + u(F_0(z) - F_0(z-))) du m(dz).$$

Let  $x > x_0$  and  $y \in (x_0, x)$ . By definition of  $x_0$ , one has  $0 < \sqrt{F_0(y)} \leq \sqrt{F_0(x)}$  and, by choosing some  $C^1$  function  $\varphi$  which coincides with the square root on  $[F_0(y), +\infty)$ , we deduce that:

$$\sqrt{F_0(x)} = \sqrt{F_0(y)} + \int_{u=0}^1 \int_{\mathbb{R}} \frac{\mathbf{1}_{\{y < z \leq x\}}}{2\sqrt{F_0(z-) + u(F_0(z) - F_0(z-))}} m(dz) du$$

We conclude by letting  $y$  decrease to  $x_0$  in this inequality using monotone convergence to deal with the integral and using the right-continuity of  $F_0$  together with (4.5.7) for  $x = x_0$  when  $x_0 > -\infty$ .  $\square$

Let us now prove Lemma 4.5.9.

*Proof.* With Fubini's theorem, one deduces from the first equality in Lemma 4.5.10 that

$$\int_{-\infty}^y \sqrt{F_0(x)} dx = \int_{u=0}^1 \int_{\mathbb{R}} \frac{\mathbf{1}_{\{z \leq y\}}(y - z)}{2\sqrt{F_0(z-) + u(F_0(z) - F_0(z-))}} m(dz) du. \quad (4.5.9)$$

By the monotonicity of  $F_0$ ,  $\forall z < y$ ,  $\int_{-\infty}^y \sqrt{F_0(x)} dx \geq \int_z^y \sqrt{F_0(x)} dx \geq \sqrt{F_0(z)}(y - z)$  which implies that

$\forall z < y, \quad \forall u \in [0, 1], \quad \frac{1}{\sqrt{F_0(z-) + u(F_0(z) - F_0(z-))}} \int_{-\infty}^y \sqrt{F_0(x)} dx \geq (y - z)$ . With (4.5.9), one deduces that  $\left( \int_{-\infty}^y \sqrt{F_0(x)} dx \right)^2 \geq \frac{1}{2} \int_{\mathbb{R}} \mathbf{1}_{\{z \leq y\}} (y - z)^2 m(dz)$ . One concludes by summing this inequality with  $\left( \int_y^{+\infty} \sqrt{1 - F_0(x)} dx \right)^2 \geq \frac{1}{2} \int_{\mathbb{R}} \mathbf{1}_{\{z \geq y\}} (y - z)^2 m(dz)$ , obtained in a symmetric way by using the second equality in Lemma 4.5.10.  $\square$

Let us now deal with the implication in Proposition 4.2.4 concerning the finiteness of

$$\sup_{N \geq 1} \sqrt{N} \int_{\mathbb{R}} \left| \tilde{F}_0^N(x) - F_0(x) \right| dx.$$

**Lemma 4.5.11.** *If  $\sup_{x \geq 1} x \int_x^{+\infty} (F_0(-y) + 1 - F_0(y)) dy < \infty$ , then  $\sup_{N \geq 1} \sqrt{N} \int_{\mathbb{R}} \left| \tilde{F}_0^N(x) - F_0(x) \right| dx < \infty$ . Moreover,*

$$\int_{\mathbb{R}} |x|^2 m(dx) < \infty \Rightarrow \sup_{x \geq 1} x \int_x^{+\infty} (F_0(-y) + 1 - F_0(y)) dy < \infty \Rightarrow \int_{\mathbb{R}} |x|^{2-} m(dx) < \infty.$$

**Remark 4.5.12.** *By [49] p.4975, if for some  $y \in \mathbb{R}$ , the restriction of  $m$  to  $[y, +\infty)$  (resp.  $(-\infty, y]$ ) has a positive non-increasing (resp. non-decreasing) density with respect to the Lebesgue measure then for  $N$  large enough,*

$$\liminf_{N \rightarrow \infty} 8\sqrt{N} \int_{\mathbb{R}} \left| \tilde{F}_0^N(x) - F_0(x) \right| dx \geq \limsup_{N \rightarrow \infty} \frac{F_0^{-1}(1 - 1/4N)}{\sqrt{N}} \left( \text{resp. } \geq \limsup_{N \rightarrow \infty} |F_0^{-1}(1/4N)| / \sqrt{N} \right).$$

Hence  $\sup_{N \geq 1} \sqrt{N} \int_{\mathbb{R}} \left| \tilde{F}_0^N(x) - F_0(x) \right| dx < \infty$  implies the existence of  $C \in (0, +\infty)$  such that for  $N$  large enough,  $F_0^{-1}(1 - 1/4N) \leq C\sqrt{N}$  and  $\frac{1}{4N} = 1 - F_0(F_0^{-1}(1 - 1/4N)) \geq 1 - F_0(C\sqrt{N})$ . Since for  $x \in [C\sqrt{N}, C\sqrt{N+1})$ ,  $\frac{1}{4N} \leq \frac{C^2(N+1)}{4x^2N} \leq \frac{C^2}{2x^2}$ , we deduce that  $1 - F_0(x) \leq \frac{C^2}{2x^2}$  (resp.  $F_0(-x) \leq \frac{C^2}{2x^2}$ ) for  $x$  large enough so that  $\sup_{x \geq 1} x \int_x^{+\infty} (1 - F_0(y)) dy < \infty$  (resp.  $\sup_{x \geq 1} x \int_x^{+\infty} F_0(-y) dy < \infty$ ).

Before proving the lemma, let us exhibit a measure  $m$  such that  $\int_{\mathbb{R}} |x|^{2-} m(dx) < \infty$  and  $\sup_{N \geq 1} \sqrt{N} \int_{\mathbb{R}} \left| \tilde{F}_0^N(x) - F_0(x) \right| dx = \infty$  so that, by the first assertion in Lemma 4.5.11,  $\sup_{x \geq 1} x \int_x^{+\infty} (F_0(-y) + 1 - F_0(y)) dy < \infty$ . We first exhibit a measure  $m$  such that  $\sup_{x \geq 1} x \int_x^{+\infty} (F_0(-y) + 1 - F_0(y)) dy < \infty$  and  $\int_{\mathbb{R}} |x|^2 m(dx) = \infty$ .

**Example 4.5.13.** *If  $m(dx) = \mathbf{1}_{\{x \geq 1\}} \frac{2}{x^3} dx$ , then  $\int_{\mathbb{R}} |x|^2 m(dx) = +\infty$  whereas  $\forall x \geq 1, F_0(-x) = 0$  and  $1 - F_0(x) = \frac{1}{x^2}$  so that  $x \int_x^{+\infty} (1 - F_0(y)) dy = 1$ . Let now  $m(dx) = \mathbf{1}_{\{x \geq 1\}} \frac{c \ln(x)}{x^3} dx$  where  $1/c = \int_1^{+\infty} \frac{\ln(x)}{x^3} dx$ . One has  $\int_{\mathbb{R}} |x|^{2-\varepsilon} m(dx) < \infty$  for all  $\varepsilon \in [0, 2)$ . let us check that  $\lim_{N \rightarrow \infty} \sqrt{N} \int_{F_0^{-1}(1-1/2N)}^{+\infty} (1 - F_0(x)) dx = +\infty$ , which, in view of (4.2.5), implies that  $\lim_{N \rightarrow \infty} \sqrt{N} \int_{\mathbb{R}} \left| \tilde{F}_0^N(x) - F_0(x) \right| dx = +\infty$  and, by the first assertion in Lemma 4.5.11, that  $\sup_{x \geq 1} x \int_x^{+\infty} (F_0(-y) + 1 - F_0(y)) dy = +\infty$ . Using the integration by parts formula like in Examples*

4.5.8, we check that, as  $x \rightarrow +\infty$ ,  $1 - F_0(x) \sim \frac{c \ln(x)}{2x^2}$  and  $\int_x^{+\infty} (1 - F_0(y)) dy \sim \frac{c \ln(x)}{2x}$ . Since  $\lim_{u \rightarrow 1-} F_0^{-1}(u) = +\infty$ , one has, for  $N$  large enough,  $\frac{c \ln(F_0^{-1}(1 - 1/2N))}{(F_0^{-1}(1 - 1/2N))^2} \geq 1 - F_0(F_0^{-1}(1 - 1/2N)) = \frac{1}{2N}$  and  $\int_{F_0^{-1}(1-1/2N)}^{+\infty} (1 - F_0(x)) dx \geq \frac{c \ln(F_0^{-1}(1 - 1/2N))}{4F_0^{-1}(1 - 1/2N)} \geq \frac{\sqrt{c \ln(F_0^{-1}(1 - 1/2N))}}{4\sqrt{2N}}$ .

*Proof.* Let us first assume the existence of  $C \in (0, +\infty)$  s.t.  $\forall x \geq 1$ ,  $\int_{-\infty}^{-x} F_0(y) dy + \int_x^{+\infty} (1 - F_0(y)) dy < \frac{C}{x}$ . We have:

$$\begin{aligned} \int_{F_0^{-1}(\frac{2N-1}{2N})}^{+\infty} (1 - F_0(x)) dx &\leq \int_{F_0^{-1}(\frac{2N-1}{2N}) \wedge \sqrt{N}}^{\sqrt{N}} (1 - F_0(x)) dx + \int_{\sqrt{N}}^{+\infty} (1 - F_0(x)) dx \\ &\leq \frac{(\sqrt{N} - F_0^{-1}(1/2))^+}{2N} + \frac{C}{\sqrt{N}}. \end{aligned} \quad (4.5.10)$$

where, for the second inequality, we used that  $1 - F_0(x) \leq 1/2N$  for  $x \geq F_0^{-1}(1 - 1/2N)$  and  $F_0^{-1}(1 - 1/2N) \geq F_0^{-1}(1/2)$  to deal with the first integral and the hypothesis to deal with the second one. In a symmetric way, we check that  $\int_{-\infty}^{F_0^{-1}(1/2N)} F_0(x) dx \leq \frac{(F_0^{-1}(1/2) - \sqrt{N})^+}{2N} + \frac{C}{\sqrt{N}}$ . For  $x \geq 2$ , we have, using the monotonicity of  $F_0$  for the first inequality, the fact that  $F_0 \leq 1$  for the second and the assumption for the third,

$$(1 - F_0(x)) \leq \frac{2}{x} \int_{x/2}^x (1 - F_0(y)) dy \leq \frac{2}{x} \int_{x/2}^{+\infty} (1 - F_0(y)) dy \leq \frac{4C}{x^2},$$

and, by left-continuity of the right-hand side,  $\frac{4C}{x^2} \geq 1 - F_0(x-)$ . Either  $F_0(1 - 1/2N) \leq 2$  or  $\frac{4C}{(F_0^{-1}(1 - 1/2N))^2} \geq 1 - F_0(F_0^{-1}(1 - 1/2N)-) \geq \frac{1}{2N}$  so that  $F_0^{-1}(1 - 1/2N) \leq 2(1 \vee \sqrt{2CN})$ . By a symmetric reasoning,  $F_0(x) \leq \frac{4C}{x^2}$  for  $x \leq -2$  and  $F_0^{-1}(1/2N) \geq -2(1 \vee \sqrt{2CN})$ . Since the integrand in (4.2.5) is smaller than  $1/2N$ , combining these bounds on  $F_0^{-1}(1/2N)$  and  $F_0^{-1}(1 - 1/2N)$  with (4.5.10) and the symmetric estimation, we deduce that

$$\int_{\mathbb{R}} |\tilde{F}_0^N(x) - F_0(x)| dx \leq \frac{2(1 \vee \sqrt{2CN})}{N} + \frac{|\sqrt{N} - F_0^{-1}(1/2)|}{2N} + \frac{2C}{\sqrt{N}},$$

which concludes the proof of the first assertion.

Since, according to (4.5.6) for  $\varepsilon = 0$ ,  $\int_{\mathbb{R}} y^2 m(dy) = 2 \left( \int_0^{+\infty} y(1 - F_0(y)) dy - \int_{-\infty}^0 y F_0(y) dy \right)$ , one has

$$\forall x > 0, \quad \int_{-\infty}^{-x} F_0(y) dy + \int_x^{+\infty} (1 - F_0(y)) dy \leq \frac{1}{2x} \int_{\mathbb{R}} y^2 m(dy).$$

Let us last check that the existence of  $C \in (0, +\infty)$ ,  $\forall x \geq 1$ ,  $\int_{-\infty}^{-x} F_0(y) dy + \int_x^{+\infty} (1 - F_0(y)) dy \leq \frac{C}{x}$  implies that  $\int_{\mathbb{R}} |x|^{2-\varepsilon} m(dx) < \infty$  for all  $\varepsilon \in (0, 1]$  (and therefore all  $\varepsilon \in (0, 2]$ ). For  $\varepsilon = 1$ , we have

$$\int_{\mathbb{R}} |x| m(dx) = \int_0^{+\infty} (F_0(-y) + (1 - F_0(y))) dy \leq 1 + \int_1^{+\infty} (F_0(-y) + (1 - F_0(y))) dy \leq 1 + C.$$



Let now  $\varepsilon \in (0, 1)$ . Using Fubini's theorem then the integration by parts formula and the fact that  $\lim_{y \rightarrow +\infty} y^{1-\varepsilon} \int_y^{+\infty} (1 - F_0(z)) dz = 0$ , we conclude that

$$\begin{aligned} \frac{1}{2-\varepsilon} \int_0^{+\infty} x^{2-\varepsilon} m(dx) &= \int_0^{+\infty} y^{1-\varepsilon} (1 - F_0(y)) dy = (1-\varepsilon) \int_0^{+\infty} y^{-\varepsilon} \int_y^{+\infty} (1 - F_0(z)) dz dy \\ &\leq (1-\varepsilon) \int_0^1 y^{-\varepsilon} dy \int_0^{+\infty} (1 - F_0(z)) dz + (1-\varepsilon) \int_1^{+\infty} y^{-\varepsilon} \int_y^{+\infty} (1 - F_0(z)) dz dy. \end{aligned}$$

Combining this inequality with the symmetric one then using the above estimation of  $\int_0^{+\infty} (F_0(-y) + (1 - F_0(y))) dy$ , we conclude that

$$\frac{1}{2-\varepsilon} \int_{\mathbb{R}} x^{2-\varepsilon} m(dx) \leq \int_0^{+\infty} (F_0(-z) + (1 - F_0(z))) dz + (1-\varepsilon) C \int_1^{+\infty} y^{-1-\varepsilon} dy \leq 1 + \frac{C}{\varepsilon}.$$

□

### 4.5.3 Further properties of the optimal deterministic initialization

According to Proposition 4.2.5, when  $m$  is compactly supported, then  $\sup_{N \geq 1} N \int_{\mathbb{R}} |\tilde{F}_0^N(x) - F_0(x)| dx < \infty$ . The next proposition states that for the latter property to hold when  $m$  has a density with respect to the Lebesgue measure, then the Lebesgue measure of the set where this density is finite must be finite.

**Proposition 4.5.14.** *If  $m(dx) = f(x) dx$ , then*

$$\liminf_{N \rightarrow \infty} N \int_{\mathbb{R}} |\tilde{F}_0^N(x) - F_0(x)| dx \geq \frac{1}{4} \int_{\mathbb{R}} \mathbf{1}_{\{f(x) > 0\}} dx.$$

**Remark 4.5.15.** *Lemma 2.1 [49] refines the statement when  $f$  is positive on a bounded interval  $[c, d]$  and equal to 0 outside by stating that  $\lim_{N \rightarrow \infty} N \int_{\mathbb{R}} |\tilde{F}_0^N(x) - F_0(x)| dx = \frac{d-c}{4}$ .*

*Proof.* Let us suppose that  $m(dx) = f(x) dx$ . The continuity of  $F_0$  implies that  $F_0(F_0^{-1}(v)) = v$  for each  $v \in (0, 1)$ . Using this equation in the first equality then the inverse transform sampling for the second equality and last that  $F_0^{-1}(v) \leq y \Leftrightarrow v \leq F_0(y)$ , we have, for  $u \in (\frac{1}{2N}, 1)$ ,

$$\begin{aligned} \int_{u-\frac{1}{2N}}^u \frac{dv}{f(F_0^{-1}(v))} &= \int_0^1 \mathbf{1}_{\{u-\frac{1}{2N} \leq F_0(F_0^{-1}(v)) < u\}} \frac{dv}{f(F_0^{-1}(v))} \\ &= \int_{\mathbb{R}} \mathbf{1}_{\{u-\frac{1}{2N} \leq F_0(x) < u\}} \frac{m(dx)}{f(x)} \\ &= \int_{\mathbb{R}} \mathbf{1}_{\{F_0^{-1}(u-\frac{1}{2N}) \leq x < F_0^{-1}(u)\}} \mathbf{1}_{\{f(x) > 0\}} dx \\ &= F_0^{-1}(u) - F_0^{-1}\left(u - \frac{1}{2N}\right) - \int_{\mathbb{R}} \mathbf{1}_{\{u-\frac{1}{2N} \leq F_0(x) < u\}} \mathbf{1}_{\{f(x)=0\}} dx. \end{aligned}$$

With (4.2.6), we deduce that

$$\begin{aligned} \int_{\mathbb{R}} |\tilde{F}_0^N(x) - F_0(x)| dx &\geq \sum_{i=1}^N \int_{\frac{2i-1}{2N}}^{\frac{i}{N}} \int_{u-\frac{1}{2N}}^u \frac{dv}{f(F_0^{-1}(v))} du = \sum_{i=1}^N \int_{\frac{i-1}{N}}^{\frac{i}{N}} \left(\frac{i}{N} - v\right) \wedge \left(v - \frac{i-1}{N}\right) \frac{dv}{f(F_0^{-1}(v))} \\ &= \frac{1}{N} \int_0^1 \min_{j \in \mathbb{N}} |Nv - j| \frac{dv}{f(F_0^{-1}(v))}, \end{aligned} \quad (4.5.11)$$

where we used Fubini's theorem for the first equality.

Let  $K \in (0, +\infty)$  be some cutoff parameter and  $\varphi_K(v) = \frac{1}{f(F_0^{-1}(v))} \wedge K$ . For  $\varepsilon > 0$ , by density of the

continuous functions in the space of integrable functions on  $[0, 1]$  endowed with the Lebesgue measure, there exists a continuous function  $\varphi_{K,\varepsilon}$  such that  $\int_0^1 |\varphi_{K,\varepsilon}(v) - \varphi_K(v)| dv \leq \varepsilon$ . One has

$$\begin{aligned} \left| \int_0^1 \varphi_K(v) 4 \min_{j \in \mathbb{N}} |Nv - j| dv - \int_0^1 \varphi_K(v) dv \right| &\leq \int_0^1 |\varphi_K(v) - \varphi_{K,\varepsilon}(v)| 4 \min_{j \in \mathbb{N}} |Nv - j| dv \\ &+ \left| \int_0^1 4 \min_{j \in \mathbb{N}} |Nv - j| \varphi_{K,\varepsilon}(v) dv - \int_0^1 \varphi_{K,\varepsilon}(v) dv \right| + \int_0^1 |\varphi_{K,\varepsilon}(v) - \varphi_K(v)| dv. \end{aligned}$$

Since for each  $N \geq 1$ ,  $\sup_{v \in [0,1]} 4 \min_{j \in \mathbb{N}} |Nv - j| \leq 2$ , the sum of the first and third terms in the right-hand side is smaller than  $3\varepsilon$ . On the other hand, the second term goes to 0 as  $N \rightarrow \infty$ , since the probability measures with densities  $4 \mathbf{1}_{\{0 \leq v \leq 1\}} \min_{j \in \mathbb{N}} |Nv - j|$  with respect to the Lebesgue measure converge weakly to the uniform distribution on  $[0, 1]$ . Hence for each  $K \in (0, +\infty)$ ,  $\lim_{N \rightarrow \infty} \int_0^1 \varphi_K(v) 4 \min_{j \in \mathbb{N}} |Nv - j| dv = \int_0^1 \varphi_K(v) dv$ . We deduce that  $\liminf_{N \rightarrow \infty} \int_0^1 4 \min_{j \in \mathbb{N}} |Nv - j| \frac{dv}{f(F_0^{-1}(v))} \geq \int_0^1 \varphi_K(v) dv$  where, by monotone convergence, the right-hand side converges to  $\int_0^1 \frac{dv}{f(F_0^{-1}(v))}$  as  $K \rightarrow \infty$ . With (4.5.11) and the inverse transform sampling, we conclude that

$$\liminf_{N \rightarrow \infty} N \int_{\mathbb{R}} |\tilde{F}_0^N(x) - F_0(x)| dx \geq \int_0^1 \frac{dv}{4f(F_0^{-1}(v))} = \frac{1}{4} \int_{\mathbb{R}} \frac{1}{f(x)} m(dx) = \frac{1}{4} \int_{\mathbb{R}} \mathbf{1}_{\{f(x) > 0\}} dx.$$

□

Let us finally check that the moments of the empirical measure of the optimal deterministic initial positions converge to those of  $m$  as  $N \rightarrow \infty$ .

**Lemma 4.5.16.** *For  $\rho > 0$ , one has  $\lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{i=1}^N \left| F_0^{-1} \left( \frac{2i-1}{2N} \right) \right|^\rho = \int_{\mathbb{R}} |x|^\rho m(dx)$  and*

$$\forall N \in \mathbb{N}^*, \frac{1}{N} \sum_{i=1}^N \left| F_0^{-1} \left( \frac{2i-1}{2N} \right) \right|^\rho \leq 2 \int_{\mathbb{R}} |x|^\rho m(dx).$$

*Proof.* One has  $\frac{1}{N} \sum_{i=1}^N \left| F_0^{-1} \left( \frac{2i-1}{2N} \right) \right|^\rho = \int_0^1 \left| F_0^{-1} \left( \frac{2\lfloor Nu \rfloor + 1}{2N} \right) \right|^\rho du$ . Since for  $u \in (0, 1)$ ,  $\lim_{N \rightarrow \infty} \frac{2\lfloor Nu \rfloor + 1}{2N} = u$  and the set of discontinuities of the non-decreasing function  $F_0^{-1}$  is at most countable, we deduce from Fatou's lemma that  $\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left| F_0^{-1} \left( \frac{2i-1}{2N} \right) \right|^\rho \geq \int_0^1 |F_0^{-1}(u)|^\rho du = \int_{\mathbb{R}} |x|^\rho m(dx)$ , where the last equality follows from the inverse transform sampling. This concludes the proof of the first assertion when  $\int_0^1 |F_0^{-1}(u)|^\rho du < \infty$ . Since

$$\begin{aligned} \forall i \in \{1, \dots, N-1\}, \quad F_0^{-1} \left( \frac{2i-1}{2N} \right) \geq 0 &\Rightarrow \frac{|F_0^{-1} \left( \frac{2i-1}{2N} \right)|^\rho}{N} \leq \int_{\frac{2i-1}{2N}}^{\frac{2i+1}{2N}} |F_0^{-1}(u)|^\rho du, \\ \forall i \in \{2, \dots, N\}, \quad F_0^{-1} \left( \frac{2i-1}{2N} \right) \leq 0 &\Rightarrow \frac{|F_0^{-1} \left( \frac{2i-1}{2N} \right)|^\rho}{N} \leq \int_{\frac{2i-3}{2N}}^{\frac{2i-1}{2N}} |F_0^{-1}(u)|^\rho du, \end{aligned}$$

one has

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \left| F_0^{-1} \left( \frac{2i-1}{2N} \right) \right|^\rho &\leq \mathbf{1}_{\{F_0^{-1}(\frac{1}{2N}) \leq 0\}} \frac{|F_0^{-1}(\frac{1}{2N})|^\rho}{N} + \int_{\frac{1}{2N}}^{\frac{2N-1}{2N}} |F_0^{-1}(u)|^\rho du + \mathbf{1}_{\{F_0^{-1}(\frac{2N-1}{2N}) \geq 0\}} \frac{|F_0^{-1}(\frac{2N-1}{2N})|^\rho}{N} \\ &\leq \int_0^{\frac{1}{2N}} |F_0^{-1}(u)|^\rho du + \int_0^1 |F_0^{-1}(u)|^\rho du + \int_{\frac{2N-1}{2N}}^1 |F_0^{-1}(u)|^\rho du, \end{aligned}$$

from which, we deduce the second assertion. When  $\int_0^1 |F_0^{-1}(u)|^\rho du < \infty$ , by Lebesgue's theorem,

$$\lim_{N \rightarrow \infty} \left( \int_0^{\frac{1}{2N}} |F_0^{-1}(u)|^\rho du + \int_{\frac{2N-1}{2N}}^1 |F_0^{-1}(u)|^\rho du \right) = 0.$$

We deduce that  $\limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left| F_0^{-1} \left( \frac{2i-1}{2N} \right) \right|^\rho \leq \int_0^1 |F_0^{-1}(u)|^\rho du$ , which concludes the proof.  $\square$

## 4.6 Numerical experiments for the Burgers equation

In order to confirm our theoretical estimates for the strong and the weak  $L^1$ -error between  $F^{N,h}$  and its limit  $F$ , we consider, for the choice  $\Lambda(u) = -(1-u)^2/2$  and the initial condition  $F(0, x) = \mathbf{1}_{\{x \geq 0\}}$ , the following equation:

$$\begin{cases} \partial_t F(t, x) + \partial_x F(t, x) (1 - F(t, x)) = \frac{\sigma^2}{2} \partial_{xx} F(t, x), \\ F(0, x) = \mathbf{1}_{\{x \geq 0\}}. \end{cases}$$

We can notice that the function  $(1 - F(t, \cdot))$  is solution of the Burgers equation that was also used in [9]. The Cole-Hopf transformation yields the following closed-form expression of  $F$ :

$$F(t, x) = 1 - \frac{\mathcal{N}\left(\frac{t-x}{\sigma\sqrt{t}}\right)}{\mathcal{N}\left(\frac{t-x}{\sigma\sqrt{t}}\right) + \exp\left(\frac{2x-t}{2\sigma^2}\right) \mathcal{N}\left(\frac{x}{\sigma\sqrt{t}}\right)},$$

where  $\mathcal{N}(x) = \int_{-\infty}^x \frac{\exp(y^2/2)}{\sqrt{2\pi}} dy$ .

The drift coefficient of the  $i^{th}$  particle in the increasing order is then equal to  $\lambda^N(i) = 1 - \frac{2i-1}{2N}$  and the Euler discretization with step  $h \in (0, T]$  of the particle system is:

$$dX_t^{i,N,h} = \sigma dW_t^i + \left( 1 + \frac{1}{2N} - \frac{1}{N} \sum_{j=1}^N \mathbf{1}_{\left\{ X_{\tau_t^h}^{j,N,h} \leq X_{\tau_t^h}^{i,N,h} \right\}} \right) dt, \quad 1 \leq i \leq N, \quad t \in [0, T].$$

As  $F_0$  is the cumulative function of the Dirac mass centered at zero, we place the  $N$  particles at zero for their initialisation.

We seek to observe the dependence of the strong  $L^1$ -error  $\mathbb{E} \left[ \mathcal{W}_1 \left( \mu_T^{N,0}, \mu_T \right) \right]$  and the weak  $L^1$ -error  $\mathcal{W}_1 \left( \mathbb{E} \left[ \mu_T^{N,h} \right], \mu_T \right)$  at time  $T$  on the number  $N$  of particles and on the time step  $h$ . We recall (4.2.1) and (4.2.2) where the Wasserstein distance between a probability measure  $\nu$  and  $\mu_T$  can be expressed either using the quantile functions or the cumulative distribution functions:

$$\begin{aligned} \mathcal{W}_1(\nu, \mu_T) &= \int_0^1 |F_\nu^{-1}(u) - F_T^{-1}(u)| du \\ &= \int_{\mathbb{R}} |F_\nu(x) - F(T, x)| dx. \end{aligned}$$

We choose to use the second expression because we have an explicit formula for  $F(T, \cdot)$  unlike the inverse  $F_T^{-1}(\cdot)$  (which can still be numerically estimated but this is costly and induces additional numerical error).

When  $\nu$  is an empirical measure of the form  $\frac{1}{N} \sum_{i=1}^N \delta_{x^i}$ , we choose to approximate the  $\mathcal{W}_1$  distance not using a grid but in the following way. For  $(y^i)_{1 \leq i \leq N}$  denoting the increasing reordering of  $(x^i)_{1 \leq i \leq N}$ , we have:  $\mathcal{W}_1 \left( \frac{1}{N} \sum_{i=1}^N \delta_{x^i}, \mu_T \right) \simeq \Psi(y^1, y^2, \dots, y^N)$  where

$$\Psi(y^1, y^2, \dots, y^N) = \sum_{i=1}^{N-1} \frac{1}{2} (y^{i+1} - y^i) \left( \left| F(T, y^{i+1}) - \frac{i}{N} \right| + \left| F(T, y^i) - \frac{i}{N} \right| \right).$$

Therefore, for the strong  $L^1$ -error,  $\left(Y_{t,r}^{i,N,h}\right)_{i \in \llbracket 1, N \rrbracket}$  being the increasing reordering of the particles positions  $\left(X_{t,r}^{i,N,h}\right)_{i \in \llbracket 1, N \rrbracket}$ ,  $t > 0$  in the  $r^{th}$  out of  $R$  independent Monte-Carlo runs, we obtain the following approximation:

$$\mathbb{E} \left[ \mathcal{W}_1 \left( \mu_T^{N,h}, \mu_T \right) \right] \simeq \frac{1}{R} \sum_{r=1}^R \Psi \left( Y_{T,r}^{1,N,h}, \dots, Y_{T,r}^{N,N,h} \right).$$

We also define the precision of this estimation as half the width of the 95% confidence interval of the empirical error i.e. Precision =  $1.96 \times \sqrt{\text{Variance}/R}$  where Variance denotes the empirical variance over the runs of the empirical error over the particles.

Concerning the weak  $L^1$ -error, we approximate  $\mathbb{E} \left[ \mu_T^{N,h} \right]$  by  $\frac{1}{R \times N} \sum_{r=1}^R \sum_{i=1}^N \delta_{X_{T,r}^{i,N,h}}$ . But as  $R \times N$  will be as big as  $10^8$  in our simulations, rather than using the previous grid free approximation, we use the grid  $\left(F_T^{-1} \left( \frac{k}{K} \right)\right)_{1 \leq k \leq K-1}$  ( $K$  will be chosen equal to 5000) to compute the  $\mathcal{W}_1$  distance. For  $k \in \llbracket 0, K-1 \rrbracket$  and  $x \in \left[F_T^{-1} \left( \frac{k}{K} \right), F_T^{-1} \left( \frac{k+1}{K} \right)\right]$ , we make the following approximation  $F(T, x) \simeq \frac{2k+1}{2K}$ . We also define the function  $\varphi$  as:

$$\begin{aligned} & \varphi(u_0, u_1, \dots, u_{K-1}) \\ &= \sum_{k=1}^{K-2} \left| u_k - \frac{2k+1}{2K} \right| \left( F_T^{-1} \left( \frac{k+1}{K} \right) - F_T^{-1} \left( \frac{k}{K} \right) \right) + 2 \left| u_0 - \frac{1}{2K} \right| \left( F_T^{-1} \left( \frac{1}{K} \right) - F_T^{-1} \left( \frac{1}{2K} \right) \right) \\ &+ 2 \left| u_{K-1} - \left( 1 - \frac{1}{2K} \right) \right| \left( F_T^{-1} \left( 1 - \frac{1}{2K} \right) - F_T^{-1} \left( 1 - \frac{1}{K} \right) \right). \end{aligned}$$

Therefore, we can approach the weak  $L^1$ -error by

$$\mathcal{W}_1 \left( \mathbb{E} \left[ \mu_T^{N,h} \right], \mu_T \right) \simeq \varphi \left( \left( \frac{1}{R} \sum_{r=1}^R F_r^{N,h} \left( T, F_T^{-1} \left( \frac{2k+1}{2K} \right) \right) \right)_{0 \leq k \leq K-1} \right).$$

We divide the  $R$  runs into  $B$  batches of  $M = R/B$  independent simulations in order to estimate the associated precision. Indeed, we estimate the empirical variance over the batches while estimating the weak error for each independent simulation over the batches. And by the delta method, we may expect, denoting  $E = \mathbb{E} \left[ \left( F^{N,h} \left( T, F_T^{-1} \left( \frac{2k+1}{2K} \right) \right) \right)_{0 \leq k \leq K-1} \right]$ , that

$$\begin{aligned} & \sqrt{\rho} \left[ \varphi \left( \left( \frac{1}{\rho} \sum_{r=1}^{\rho} F_r^{N,h} \left( T, F_T^{-1} \left( \frac{2k+1}{2K} \right) \right) \right)_{0 \leq k \leq K-1} \right) - \varphi(E) \right] \\ & \xrightarrow{\mathcal{L}} \mathcal{N} \left( 0, \nabla \varphi^T(E) \text{Cov} \left[ \left( F^{N,h} \left( T, F_T^{-1} \left( \frac{2k+1}{2K} \right) \right) \right)_{0 \leq k \leq K-1} \right] \nabla \varphi(E) \right) \end{aligned}$$

when  $\rho \rightarrow +\infty$ . Applying this result with  $\rho = M$  and  $\rho = R$ , one expects that  $1/B$  times the empirical variance of  $\left( \varphi \left( \left( \frac{1}{M} \sum_{r=(b-1)M+1}^{bM} F_r^{N,h} \left( T, F_T^{-1} \left( \frac{2k+1}{2K} \right) \right) \right)_{0 \leq k \leq K-1} \right) \right)_{1 \leq b \leq B}$  over the batches provides an estimator of the variance of  $\varphi \left( \left( \frac{1}{R} \sum_{r=1}^R F_r^{N,h} \left( T, F_T^{-1} \left( \frac{2k+1}{2K} \right) \right) \right)_{0 \leq k \leq K-1} \right)$ . So the precision is computed as 1.96 times the square root of this estimator.

For both of the errors, we fix the time horizon  $T = 1$  and the diffusion coefficient  $\sigma^2 = 0.2$ .

### 4.6.1 Strong $L^1$ -error behaviour

We present numerical estimates of  $\mathbb{E} \left[ \mathcal{W}_1 \left( \mu_T^{N,h}, \mu_T \right) \right]$ , computed as described above.

► **Dependence on  $N$ :**

We fix the time-step  $h = 0.002$  small enough in order to observe the effect of the number  $N$  of particles on the error. The simulation is done with  $R = 100$  Monte-Carlo runs. We obtain the following results for the estimation of the error and the associated precision:

| Evolution of the strong $L^1$ -error w.r.t. $N$ |            |            |                   |
|-------------------------------------------------|------------|------------|-------------------|
| Number of particles $N$                         | Estimation | Precision  | Ratio of decrease |
| 250                                             | 0.03312361 | 0.00290442 | $\times$          |
| 1000                                            | 0.01598253 | 0.00133181 | 2.07              |
| 4000                                            | 0.00841976 | 0.00077491 | 1.90              |
| 16000                                           | 0.00358799 | 0.00028319 | 2.35              |
| 64000                                           | 0.00193416 | 0.00016111 | 1.86              |

We observe that the ratio of successive estimations  $\frac{\text{Estimation}(N/4)}{\text{Estimation}(N)}$  is around 2 when we multiply  $N$  by 4, which means that the strong  $L^1$ -error is roughly proportional to  $N^{-1/2}$ .

**Remark 4.6.1.** *This strong error was computed for the choice  $\sigma^2 = 0.2$ . When choosing the larger variance  $\sigma^2 = 20$  with the same time-step  $h = 0.002$ , the error is approximately multiplied by 10 and the  $N^{-1}$  behaviour is still observed. For the smaller variance  $\sigma^2 = 0.002$ , we need to choose the smaller time-step  $h = 0.001$  in order to recover the  $N^{-1}$  behaviour in the number of particles since the estimated error is very small.*

► **Dependence on  $h$ :**

We apply the same strategy to study the dependence of the error on  $h$  by choosing a large number  $N = 500000$  of particles. The following table presents numerical estimates of the  $L^1$ -norm of the error and its associated precision for  $R = 100$  runs.

| Evolution of the strong $L^1$ -error w.r.t. $h$ |            |                         |                   |
|-------------------------------------------------|------------|-------------------------|-------------------|
| Time-step $h$                                   | Estimation | Precision               | Ratio of decrease |
| 1/2                                             | 0.07963922 | $5.8279 \times 10^{-5}$ | $\times$          |
| 1/4                                             | 0.03550774 | $5.9151 \times 10^{-5}$ | 2.24              |
| 1/8                                             | 0.01682159 | $5.9598 \times 10^{-5}$ | 2.11              |
| 1/16                                            | 0.00817936 | $6.5275 \times 10^{-5}$ | 2.06              |
| 1/32                                            | 0.00409004 | $5.6482 \times 10^{-5}$ | 2.00              |
| 1/64                                            | 0.00206127 | $5.3539 \times 10^{-5}$ | 1.98              |
| 1/128                                           | 0.00116977 | $4.7468 \times 10^{-5}$ | 1.76              |
| 1/256                                           | 0.00084118 | $5.9673 \times 10^{-5}$ | 1.39              |

We observe that when the time step  $h$  is divided by 2, the ratio of decrease  $\frac{\text{Estimation}(h)}{\text{Estimation}(h/2)}$  is approximately equal to 2.

**Remark 4.6.2.** *This strong error was computed for the choice  $\sigma^2 = 0.2$ . When choosing the larger variance  $\sigma^2 = 20$  with the same number of particles  $N = 500000$ , the error is of the same order for the 3 first time-steps but deteriorates afterwards and becomes larger compared to errors obtained for  $\sigma^2 = 0.2$ . The behaviour of the error also deteriorates quickly and tends to be constant. The number of particles is no longer large enough in order to observe the effect of the time-step when this latter becomes small. For the smaller variance  $\sigma^2 = 0.002$ , the error is smaller than the ones obtained for greater  $\sigma^2$ . The  $h$ -behaviour is recovered when the time-steps are small.*

### 4.6.2 Weak $L^1$ -error behaviour

We present numerical estimates of  $\mathcal{W}_1\left(\mathbb{E}\left[\mu_T^{N,h}\right], \mu_T\right)$ , computed as described above.

► **Dependence on  $N$ :**

We fix the time-step  $h = 0.002$  small enough once again to observe the effect of the number  $N$  of particles on the weak error. The estimation is done with  $B = 100$  batches of  $M = 200$  independent simulations for a total of  $R = 20000$  Monte-Carlo runs and  $K = 5000$ . The results are shown in the following table:

| Evolution of the weak $L^1$ -error w.r.t. $N$ |            |                         |                   |
|-----------------------------------------------|------------|-------------------------|-------------------|
| Number of particles $N$                       | Estimation | Precision               | Ratio of decrease |
| 100                                           | 0.01018160 | $5.6947 \times 10^{-4}$ | $\times$          |
| 200                                           | 0.00483151 | $3.8455 \times 10^{-4}$ | 2.11              |
| 400                                           | 0.00248807 | $2.0485 \times 10^{-4}$ | 1.94              |
| 800                                           | 0.00136491 | $1.4707 \times 10^{-4}$ | 1.82              |
| 1600                                          | 0.00077723 | $1.0822 \times 10^{-4}$ | 1.76              |
| 3200                                          | 0.00038285 | $4.9747 \times 10^{-5}$ | 2.03              |

We observe that multiplying the number of particles by 2 implies a division of the error estimation by approximately 2 which proves that the weak  $L^1$ -error is roughly proportional to  $N^{-1}$ .

**Remark 4.6.3.** *This weak error was computed for the choice  $\sigma^2 = 0.2$ . When choosing the larger variance  $\sigma^2 = 20$  with the time-step  $h = 0.002$ , the  $N^{-1}$  behaviour is still observed. When choosing the smaller variance  $\sigma^2 = 0.05$  and the smaller time-step  $h = 0.001$ , the  $N^{-1}$  behaviour is recovered. The weak error estimates only vary very slightly when varying  $\sigma$ . However, for smaller values of  $\sigma$ , the computation of the quantile function becomes difficult.*

► **Dependence on  $h$ :**

Once again, we do the same to study the dependence of the weak error on  $h$  by choosing a large number  $N = 100000$  of particles,  $B = 20$  batches of  $M = 50$  independent simulations for a total of  $R = 1000$  Monte-Carlo runs and  $K = 5000$ .

| Evolution of the weak $L^1$ -error w.r.t. $h$ |            |                         |                   |
|-----------------------------------------------|------------|-------------------------|-------------------|
| Time-step $h$                                 | Estimation | Precision               | Ratio of decrease |
| 1/2                                           | 0.07954397 | $4.7356 \times 10^{-5}$ | $\times$          |
| 1/4                                           | 0.03546112 | $4.7932 \times 10^{-5}$ | 2.24              |
| 1/8                                           | 0.01681185 | $4.0437 \times 10^{-5}$ | 2.11              |
| 1/16                                          | 0.00816986 | $4.1616 \times 10^{-5}$ | 2.06              |
| 1/32                                          | 0.00407191 | $3.9306 \times 10^{-6}$ | 2.01              |
| 1/64                                          | 0.00199744 | $3.0719 \times 10^{-5}$ | 2.04              |
| 1/128                                         | 0.00096767 | $5.6043 \times 10^{-5}$ | 2.06              |
| 1/256                                         | 0.00048294 | $3.6172 \times 10^{-5}$ | 2.00              |

We observe that dividing the time step  $h$  by 2 implies a ratio of decrease  $\frac{\text{Estimation}(h)}{\text{Estimation}(h/2)}$  greater or equal to 2 which shows an  $L^1$ -weak error roughly proportional to  $h$ .

**Remark 4.6.4.** *This weak error was computed for the choice  $\sigma^2 = 0.2$ . When choosing the smaller variance  $\sigma^2 = 0.05$  with the same number of particles  $N = 100000$ , the  $h$  behaviour is recovered. When choosing the larger variance  $\sigma^2 = 20$ , the  $h$  behaviour is observed for the first time-steps until 1/32. However, for smaller time-steps, the error becomes larger, compared to the values obtained for smaller  $\sigma$ , and tends to be constant. In order to only observe the effect of the time-step on the weak error, we need to increase the number of particles when  $\sigma$  is large.*

## B.2 Appendix

The first lemma gives a condition under which we can interchange a Lebesgue and a stochastic integral. It is called the stochastic Fubini theorem and is a consequence of Theorem 2.2 proved by Veraar in [86].

**Lemma B.2.1.** *Let  $V : [0, T] \times \mathbb{R} \times \Omega \rightarrow \mathbb{R}$  be a progressively measurable function.*

*If  $\int_{\mathbb{R}} \left( \int_0^T |V(t, x)|^2 dt \right)^{1/2} dx < \infty$  almost surely then one has:*

$$\forall t \in [0, T], \quad a.s., \quad \int_{\mathbb{R}} \left( \int_0^t V(s, x) dW_s \right) dx = \int_0^t \left( \int_{\mathbb{R}} V(s, x) dx \right) dW_s.$$

For  $t > 0$ , let  $G_t$  denote the probability density function of the normal law  $\mathcal{N}(0, \sigma^2 t)$ :

$$G_t(x) = \exp\left(-\frac{x^2}{2\sigma^2 t}\right) / \sqrt{2\pi\sigma^2 t}.$$

The following lemma provides a set of estimates that are very useful:

**Lemma B.2.2.** *The function  $G_t(x)$  solves the heat equation:*

$$\partial_t G_t(x) - \frac{\sigma^2}{2} \partial_{xx} G_t(x) = 0, \quad (t, x) \in [0, +\infty) \times \mathbb{R}. \quad (\text{B.2.1})$$

*We can express the square of the first order spatial derivative as:*

$$(\partial_x G_t)^2(x) = \frac{x^2}{2\sigma^5 t^{5/2} \sqrt{\pi}} G_{t/2}(x), \quad (\text{B.2.2})$$

*and deduce the  $L^1$ -norm of  $\partial_x G^2$ :*

$$\int_{\mathbb{R}} (\partial_x G_t)^2(x) dx = \frac{1}{4\sigma^3 t^{3/2} \sqrt{\pi}}. \quad (\text{B.2.3})$$

*Moreover, we have estimates of the  $L^1$ -norm of the spatial derivatives of  $G$ :*

$$\|\partial_x G_t\|_{L^1} = \sqrt{\frac{2}{\pi\sigma^2 t}}, \quad (\text{B.2.4})$$

$$\|\partial_{xx} G_t\|_{L^1} \leq \frac{2}{\sigma^2 t}. \quad (\text{B.2.5})$$

*We also have estimates of the  $L^1$ -norm of  $G^2$ :*

$$\int_{\mathbb{R}} G_t^2(x) dx = \frac{1}{2\sigma \sqrt{\pi t}}, \quad (\text{B.2.6})$$

*which implies that for every measurable function  $y : [0, T] \rightarrow \mathbb{R}$ ,*

$$\int_{\mathbb{R}} \int_0^t G_{t-s}^2(y(s) - x) ds dx = \frac{1}{\sigma} \sqrt{\frac{t}{\pi}}. \quad (\text{B.2.7})$$

*Proof.* The second estimate is obtained by rewriting  $\partial_{xx} G_t(x)$  as  $\partial_{xx} G_t(x) = -\frac{1}{\sigma^2 t} G_t(x) + \frac{1}{\sigma^2 t} (-x \partial_x G_t(x))$ .

We apply an integration by parts for the second term and obtain:  $\int_{\mathbb{R}} \partial_{xx} G_t(x) dx \leq \frac{2}{\sigma^2 t} \|G_t\|_{L^1} = \frac{2}{\sigma^2 t}$ .

As for the estimates of  $\|G_t^2\|_{L^1}$  and  $\int_{\mathbb{R}} \int_0^t G_{t-s}^2(y(s) - x) ds dx$ , we use the fact that

$$G_t^2(x) = G_{t/2}(x) / 2\sigma \sqrt{\pi t}.$$

□





- [1] L. Ambrosio, N. Fusco and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Mathematical Monographs, pp.119–120, 2000.
- [2] F. Antonelli and A. Kohatsu-higa, *Rate of convergence of a particle method to the solution of the McKean-Vlasov's equation*. The Annals of Applied Probability, Vol.12(2), pp. 423–476, 2002.
- [3] V. Bally and C. Rey, *Approximation of Markov semigroups in total variation distance*. Electronic Journal of Probability, Vol.21(12), 2016.
- [4] A.D. Banner, E.R. Fernholz and I. Karatzas, *Atlas models of equity markets*. Annals of Applied Probability, Vol.15(4), pp.2296–2330, 2005.
- [5] J. Bao, X. Huang and C. Yuan, *Convergence Rate of Euler-Maruyama Scheme for SDEs with Hölder-Dini Continuous Drifts*. Journal of Theoretical Probability, Vol.32, pp.848–871, 2019.
- [6] O. Bencheikh and B. Jourdain, *Bias behaviour and antithetic sampling in mean-field particle approximations of SDEs nonlinear in the sense of McKean*. ESAIM:ProcS, Vol.65, pp.219–235, 2019.
- [7] S. Bobkov and M. Ledoux, *One-Dimensional Empirical Measures, Order Statistics, and Kantorovich Transport Distances*. Memoirs of the American Mathematical Society, Vol.265(1259), 2019.
- [8] M. Bossy, *Optimal rate of convergence of a stochastic particle method to solutions of 1D viscous scalar conservation laws*. Mathematics of Computation, Vol.73(246), pp.777–812, 2003.
- [9] M. Bossy and D. Talay, *Convergence rate for the approximation of the limit law of weakly interacting particles: Application to the Burgers equation*. Annals of Applied Probability, Vol.6(3), pp.818–861, 1996.
- [10] H. Brezis, *Analyse Fonctionnelle: Théorie et applications*. Collection Mathématiques appliquées pour la maîtrise - Masson, 2<sup>e</sup> tirage, 1987.
- [11] K. Bujok, B. Hambly and C. Reisinger, *Multilevel Simulation of Bernoulli Random Variables with Applications to Basket Credit Derivatives*. Methodology and Computing in Applied Probability, Vol.17, pp.579–604, 2015.
- [12] J.-F. Chassagneux, D. Crisan and F. Delarue, *A Probabilistic approach to classical solutions of the master equation for large population equilibria*. Preprint arXiv:1411.3009, 2015.
- [13] J.-F. Chassagneux, L. Szpruch and A. Tse, *Weak quantitative propagation of chaos via differential calculus on the space of measures*. Preprint arXiv:1901.02556, 2019.
- [14] P.-E. Chaudru de Raynal and N. Frikha, *From the backward Kolmogorov PDE on the Wasserstein space to propagation of chaos for McKean-Vlasov SDEs*. Preprint arXiv:1907.01410, 2019.

- [15] N. Chen, P. Glynn and Y. Liu, *Estimating expectations of functionals of conditional expectations via multilevel nested simulation*. Presentation at MCQMC, 2012.
- [16] K. Dareiotis and M. Gerencsér, *On the regularisation of the noise for the Euler-Maruyama scheme with irregular drift*. Preprint arXiv:1812.04583, 2020.
- [17] T. Daun, *On the randomized solution of initial value problems*. Journal of Complexity, Vol.27, pp.300–311, 2011.
- [18] P. Del Moral, F. Patras and S. Rubenthaler, *Tree based functional expansions for Feynman-Kac particle models*. Annals of Applied Probability, Vol.10(2), pp.778–825, 2009.
- [19] P. Eto and M. Martinez, *Exact simulation for solutions of one-dimensional Stochastic Differential Equations with discontinuous drift*. ESAIM: Probability and Statistics, EDP Sciences, Vol.18, pp.686–702, 2014.
- [20] E.R. Fernholz, *Stochastic Portfolio Theory*. Applications of Mathematics - Springer, Vol.48, pp.1–24, 2002.
- [21] A. Friedman, *Stochastic Differential Equations and Applications*. Academic Press, Vol.1, 1975.
- [22] N. Frikha, *On the weak approximation of a skew diffusion by an Euler-type scheme*. Bernoulli, Vol.24(3), pp.1653–1691, 2018.
- [23] J.-G. Gaines and T.-J. Lyons, *Random generation of stochastic area integrals*. SIAM Journal on Applied Mathematics, Vol.54(4), pp.1132–1146, 1994.
- [24] M. Giles, *Multilevel Monte Carlo Path Simulation*. Operations Research, Vol.56(3), pp.607–617, 2008.
- [25] M. Giles, *Multilevel Monte Carlo methods*. Acta Numerica, Vol.24, pp.259–328, 2015.
- [26] M. Giles and L. Szpruch, *Antithetic multilevel Monte Carlo estimation for multidimensional SDES without Lévy area simulation*. The Annals of Applied Probability, Vol.24(4), pp.1585–1620, 2014.
- [27] E. Gobet and C. Labart, *Sharp estimates for the convergence of the density of the Euler scheme in small time*. Electronic Communications in Probability, Institute of Mathematical Statistics, Vol.13, pp.352–363, 2008.
- [28] S. Göttlich, K. Lux and A. Neuenkirch, *The Euler scheme for stochastic differential equations with discontinuous drift coefficient: A numerical study of the convergence rate*. Advances in Difference Equation, Vol.429, 2019.
- [29] C. Graham and S. Méléard, *Stochastic Particle Approximations for Generalized Boltzmann Models and Convergence Estimates*. Annals of Probability, Vol.25(1), pp.115–132, 1997.
- [30] J. Guyon, *Euler scheme and tempered distributions*. Stochastic Processes and their Applications, Vol.116(6), pp.877–904, 2006.
- [31] I. Gyöngy and N. Krylov, *Existence of strong solutions for Ito’s stochastic equations via approximations*. Probability Theory and Related Fields, Vol.105, pp.143–158, 1996.
- [32] I. Gyöngy, *A note on Euler’s approximations*. Potential Analysis, Vol.8, pp.205–216, 1998.
- [33] I. Gyöngy and M. Rásonyi, *A note on Euler approximations for SDEs with Hölder continuous diffusion coefficients*. Stochastic Processes and their Applications, Vol.121, pp.2189–2200, 2011.
- [34] A.-L. Haji-Ali, *Pedestrian flow in the mean-field limit*. MSc thesis, KAUST, 2012.
- [35] A.-L. Haji-Ali and R. Tempone, *Multilevel and Multi-index Monte Carlo methods for the McKean-Vlasov equation*. Statistics and Computing, Vol.28, pp.923–935, 2017.

- [36] N. Halidias and P.E. Kloeden, *A note on the Euler-Maruyama scheme for stochastic differential equations with a discontinuous monotone drift coefficient*. BIT Numerical Mathematics, Vol.48, pp.51–59, 2008.
- [37] S. Heinrich and B. Milla, *The randomized complexity of initial value problems*. Journal of Complexity, Vol.24, pp.77–88, 2008.
- [38] J. M. Holte, *Discrete Gronwall Lemma And Applications*. Mathematical Association of America North Central Section Meeting at UND, 24 October 2009.
- [39] K. Itô, *On stochastic differential equations*. Mathematika, Vol.1(1), pp.1–57, 1951.
- [40] J.-F. Jabir, *Rate of propagation of chaos for diffusive stochastic particle systems via Girsanov transformation*. Preprint arXiv:1907.09096, 2019.
- [41] A. Jentzen and A. Neuenkirch, *A random euler scheme for carathéodory differential equations*. Journal of computational and applied mathematics, Vol.224(1), pp.346–359, 2009.
- [42] B. Jourdain, *Diffusion Processes Associated with Nonlinear Evolution Equations for Signed Measures*. Methodology and Computing in Applied Probability, Vol.2(1), pp.69–91, 2000.
- [43] B. Jourdain, *Diffusions with a nonlinear irregular drift coefficient and probabilistic interpretation of generalized Burger’s equations*. ESAIM : P&S, Vol.1, pp.339–355, 1997.
- [44] B. Jourdain, *Probabilistic approximation for a porous medium equation*. Stochastic Processes and their Applications, Vol.89(1), pp.81–99, 2000.
- [45] B. Jourdain, *Probabilistic characteristics method for a 1D scalar conservation law*. Annals of Applied Probability, Vol.12(1), pp.334–360, 2002.
- [46] B. Jourdain and F. Malrieu, *Propagation of chaos and Poincaré inequalities for a system of particles interacting through their cdf*. Annals of Applied Probability, Vol.18(5), pp.1706–1736, 2008.
- [47] B. Jourdain, S. Méléard and W. Wołczyński, *Nonlinear SDEs driven by Lévy processes and related PDEs*. ALEA Latin American Journal of Probability and Mathematical Statistics, Vol.4, pp.1–29, 2008.
- [48] B. Jourdain, S. Méléard and W.A. Wołczyński, *Probabilistic approximation and inviscid limits for 1-D fractional conservation laws*. Bernoulli, Vol.11(4), pp.689–714, 2005.
- [49] B. Jourdain and J. Reygner, *Optimal convergence rate of the multitype sticky particle approximation of one-dimensional diagonal hyperbolic systems with monotonic initial data*. Discrete & Continuous Dynamical Systems - A, Vol.36(9), pp. 4963–4996, 2016.
- [50] B. Jourdain and J. Reygner, *Propagation of chaos for rank-based interacting diffusions and long time behaviour of a scalar quasilinear parabolic equation*. Stochastic PDE: Analysis & Computations, Vol.1(3), pp.455–506, 2013.
- [51] B. Jourdain and J. Reygner, *The small noise limit of order-based diffusion processes*. Electronic Journal of Probability, Vol.19, pp.29–36, 2014.
- [52] B. Jourdain and R. Roux, *Convergence of a stochastic particle approximation for fractional scalar conservation laws*. Stochastic Processes and their Applications, Vol.121(5), pp.957–988, 2011.
- [53] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics, Vol.113(2), p.441, 1998.
- [54] S. Kanagawa, *On the rate of convergence for Maruyama’s approximate solutions of stochastic differential equations*. Yokohama Mathematical Journal, Vol.36(1), pp.79–86, 1988.
- [55] P. Kloeden and T. Shardlow, *Gauss-quadrature method for one-dimensional mean-field SDEs*. Preprint arXiv:1608.06741, 2016.

- [56] P. Kolli and M. Shkolnikov, *SPDE limit of the global fluctuations in rank-based models*. The Annals of Probability, Vol.46(2), pp.1042–1069, 2018.
- [57] V. Kolokoltsov and M. Troeva, *On the mean filed games with common noise and the McKean-Vlasov SPDEs*. Stochastic Analysis and Applications, Vol.37(4), pp.522–549, 2019.
- [58] A. Kohatsu-Higa, A. Lejay and K. Yasuda, *On Weak Approximation of Stochastic Differential Equations with Discontinuous Drift Coefficient*. Mathematical Economics, Kyoto, Japan, pp.94–106, 2012.
- [59] A. Kohatsu-Higa, A. Lejay and K. Yasuda, *Weak rate of convergence of the Euler-Maruyama scheme for stochastic differential equations with non-regular drift*. Journal of Computational and Applied Mathematics, Vol.326, pp.138–158, 2017.
- [60] V. Konakov and S. Menozzi, *Weak error for the Euler scheme approximation of diffusions with non-smooth coefficients*. Electronic Journal of Probability, Vol.22(46), pp.1–47, 2017.
- [61] R. Kruse and Y. Wu, *A randomized Milstein method for stochastic differential equations with non-differentiable drift coefficients*. Discrete and Continuous Dynamical Systems, Vol.24(8), pp.3475–3502, 2019.
- [62] H. Kunita, *Stochastic differential equations and stochastic flows of diffeomorphisms*. École d'été de Probabilités de Saint-Flour XII - 1982. Lecture Notes in Mathematics - Springer Verlag, pp.144–305, 1984.
- [63] G. Leobacher and M. Szölgyenyi, *A numerical method for SDEs with discontinuous drift*. BIT Numerical Mathematics, Vol.56, pp.151–162, 2016.
- [64] G. Leobacher and M. Szölgyenyi, *A strong order 1/2 method for multidimensional SDEs with discontinuous drift*. Annals of Applied Probability, Vol.27(4), pp.2382–2418, 2017.
- [65] G. Leobacher and M. Szölgyenyi, *Convergence of the Euler-Maruyama method for multidimensional SDEs with discontinuous drift and degenerate diffusion coefficient*. Numerische Mathematik, Vol.138(1), pp.219–239, 2017.
- [66] P.-L. Lions, *Théorie des jeux à champs moyens*. Cours au Collège de France, 2014.
- [67] S. Mischler, C. Mouhot and B. Wennberg, *A new approach to quantitative propagation of chaos for drift, diffusion and jump processes*. Probability Theory and related fields - Springer, Vol.161(1-2), pp.1–59, 2015.
- [68] R. Mikulevicius and E. Platen, *Rate of convergence of the Euler approximation for diffusion processes*. Mathematische Nachrichten, Vol.151, pp.233–239, 1991.
- [69] R. Mikulevicius, *On the rate of convergence of simple and jump-adapted weak Euler schemes for Lévy driven SDEs*. Stochastic Processes and their Applications, Vol.122(7), pp.2730–2757, 2012.
- [70] R. Mikulevicius and C. Zhang, *Weak Euler approximation for Itô's diffusion and jump processes*. Stochastic Analysis and Applications, Vol.33(3), pp.549–571, 2015.
- [71] T. Müller-Gronbach and L. Yaroslavl'tseva, *On the performance of the Euler-Maruyama scheme for SDEs with discontinuous drift coefficient*. Annales de l'Institut Henri Poincaré - Probabilités et Statistiques, Vol.56(2), pp.1162–1178, 2020.
- [72] T. Müller-Gronbach and L. Yaroslavl'tseva, *A strong order 3/4 method for SDEs with discontinuous drift coefficient*. Preprint arXiv:1904.09178, 2019.
- [73] A. Neuenkirch, M. Szölgyenyi and L. Szpruch, *An Adaptive Euler-Maruyama Scheme for Stochastic Differential Equations with Discontinuous Drift and its Convergence Analysis*. SIAM Journal on Numerical Analysis, Vol.57(1), pp.378–403, 2019.
- [74] A. Neuenkirch and M. Szölgyenyi, *The Euler-Maruyama Scheme for SDEs with Irregular Drift: Convergence Rates via Reduction to a Quadrature Problem*. Preprint arXiv:1904.07784, 2020.

- [75] H. Ngo and D. Taguchi, *Strong rate of convergence for the Euler-Maruyama approximation of stochastic differential equations with irregular coefficients*. Mathematics of Computation, Vol.85, pp.1793–1819, 2016.
- [76] H. Ngo and D. Taguchi, *On the Euler-Maruyama approximation for one-dimensional stochastic differential equations with irregular coefficients*. IMA Journal of Numerical Analysis, Vol.37(4), pp.1864–1883, 2017.
- [77] S. Rachev and L. Rüschendorf, *Mass Transportation Problems, Volume I: Theory*. Probability and its Applications, p.45, 1998.
- [78] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*. Vol.293(3), p.224, 2005.
- [79] L. Szpruch and A. Tse *Antithetic multilevel particle system sampling method for McKean-Vlasov SDEs*. Preprint arXiv:1903.07063, 2019.
- [80] B. W. Silverman, *Density estimation for statistics and data analysis*. London: Chapman & Hall/CRC. p.45, 1998.
- [81] G. Stengle, *Numerical Methods for Systems with Measurable Coefficients*. Applied Mathematics Letters, Vol.3(4), pp.25–29, 1990.
- [82] G. Stengle, *Error analysis of a randomized numerical method*. Numerische Mathematik, Vol.70, pp.119–128, 1995.
- [83] Y. Suo, C. Yuan and S. Zhang, *Weak convergence of Euler scheme for SDEs with singular drift*. Preprint arXiv:2005.04631, 2020.
- [84] A.-S. Snitzman, *Topics in propagation of chaos*. École d’été de Probabilités de Saint-Flour XIX - 1989. Lecture Notes in Mathematics - Springer Berlin Heidelberg, pp.165–251, 1991.
- [85] D. Talay and L. Tubaro, *Expansion of the global error for numerical schemes solving stochastic differential equations*. Stochastic Analysis and Applications, Vol.8(4), pp. 483–509, 1990.
- [86] M. Veraar, *The Stochastic Fubini theorem revisited*. Stochastics: An International Journal of Probability and Stochastic Processes, Vol.84(4), pp.1–9, 2011.
- [87] A. J. Veretennikov, *On strong solutions and explicit formulas for solutions of stochastic integral equations*. Mathematics of the USSR-Sbornik, Vol.39(3), pp.387–403, 1981.
- [88] A. J. Veretennikov, *On the criteria for existence of a strong solution of a stochastic equation*. Theory of Probability and its Applications, Vol.27(3), pp.441–449, 1982.
- [89] A. J. Veretennikov, *On stochastic equations with degenerate diffusion with respect to some of the variables*. Mathematics of the USSR-Izvestiya, Vol.22(1), pp.173–180, 1984.
- [90] L. Yan, *The Euler scheme with irregular coefficients*. The Annals of Probability, Vol.30(3), pp.1172–1194, 2002.
- [91] A. K. Zvonkin, *A transformation of the phase space of a diffusion process that removes the drift*. Mathematics of the USSR-Sbornik, Vol.22(129), pp.129–149, 1974.