



HAL
open science

Probabilistic and statistical analysis of diffusion systems in presence of noise

Raphaël Maillet

► **To cite this version:**

Raphaël Maillet. Probabilistic and statistical analysis of diffusion systems in presence of noise. Analysis of PDEs [math.AP]. Université Paris sciences et lettres, 2024. English. NNT : 2024UPSLD025 . tel-04828513

HAL Id: tel-04828513

<https://theses.hal.science/tel-04828513v1>

Submitted on 10 Dec 2024

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.



THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à Université Paris-Dauphine

**Probabilistic and Statistical Analysis of Diffusion Systems
with Noise**

Soutenu par

Raphaël MAILLET

Le 22 octobre 2024

École doctorale n°543

SDOSE

Spécialité

Sciences

Préparée au

CEREMADE

Composition du jury :

Fabienne COMTE Université Paris Cité	<u>Rapporteuse</u>
Daniel LACKER Columbia University	<u>Rapporteur</u>
Jean-François CHASSAGNEUX Université Paris Cité	<u>Examineur</u>
Paul GASSIAT Université Paris-Dauphine	<u>Examineur</u>
Arnaud GLOTER Université d'Évry Val d'Essonne	<u>Examineur</u>
Eva LÖCHERBACH Université de Paris 1 Panthéon-Sorbonne	<u>Présidente du jury</u>
Pierre CARDALIAGUET Université Paris-Dauphine	<u>Directeur de thèse</u>
Marc HOFFMANN Université Paris-Dauphine	<u>Directeur de thèse</u>

A mon grand-père Philippe.

Remerciements

Après avoir passé trois ans à rabâcher les oreilles de mon entourage et de mes collègues avec mes histoires de bruit commun, je suis ravi d'avoir enfin l'opportunité de les remercier officiellement dans ces pages. Sans leur patience, leur soutien, leurs encouragements ainsi que leurs précieux conseils cette thèse n'aurait probablement jamais vu le jour. Je vais certainement en faire trop, mais comme vous allez rapidement vous en rendre compte, j'apprécie les envolées lyriques.

Merci tout d'abord, à mon étonnante - *je ne compte plus les fois où j'ai dû expliquer mon sujet de thèse à des regards interloqués (bon finalement les stats sur les MFG attendront ...)* - mais tout aussi exceptionnelle direction de thèse. J'ai été extrêmement chanceux - *en tant que Dauphinois de longue date* - de bénéficier de vos enseignements respectifs alors que je n'étais encore qu'un enfant. Grâce à vous deux, j'ai pu évoluer dans un environnement de travail stimulant, tout en ayant la liberté de laisser vagabonder mon esprit et d'explorer de nouvelles idées. Merci de m'avoir guidé avec autant de patience et de passion tout au long de ces trois années.

Merci à Marc, grâce à qui j'ai découvert (et aimé !) les statistiques en Licence. Merci d'avoir ensuite accepté d'encadrer chaque mémoire que j'ai dû faire tout au long de ma scolarité, ce qui m'a mené jusqu'à cette thèse. Merci pour ton écoute attentive et ton soutien depuis de nombreuses années. Tes conseils, habilement dissimulés derrière des métaphores militaires et/ou d'alpinisme, m'ont ouvert de nouvelles perspectives et m'ont aidé à gagner en confiance.

Merci Pierre pour ton encadrement scientifique irréprochable et ta bienveillance constante. Ton expertise et ta rigueur m'ont permis de progresser et de me dépasser tout au long de cette thèse. Un jeune chercheur ne peut rêver d'un meilleur accompagnement que celui que tu m'as offert. Je suis extrêmement heureux et fier d'avoir pu travailler avec toi.

I am profoundly grateful to Fabienne Comte and Daniel Lacker for their kindness in reviewing this manuscript and enabling its defense. Their respective works have been a tremendous source of inspiration for me, and I am deeply touched by their attention to my research. Un grand merci également à Eva Löcherbach, Arnaud Gloter, Jean-François Chassagneux et Paul Gassiat, qui ont accepté de faire partie du jury de ma thèse. C'est un grand honneur qu'ils me font et je leur en suis très reconnaissant.

Je ne peux pas passer à côté de l'occasion de remercier chaleureusement François Delarue et Etienne Tanré, avec qui j'ai eu l'immense plaisir de collaborer.

Merci à François de m'avoir ouvert les portes du LJAD ¹, merci pour tout le temps que tu m'as accordé et ton intérêt pour la question qui me taraudait l'esprit sans que je ne puisse y répondre. Travailler avec toi a été une grande source d'inspiration, j'admire profondément tes qualités humaines et scientifiques. Merci !

Merci à Etienne pour tous nos échanges passionnants, pour tes précieux conseils et ton humour qui a égayé chaque discussion. Merci d'avoir partagé avec moi ta connaissance des équations de McKean-Vlasov, me poussant à me poser de nouvelles questions et à explorer de nouveaux horizons.

¹et même celles de ton bureau !

Merci à tous les enseignants chercheurs du Ceremade, tout d'abord pour tous les cours que j'ai eu la chance de suivre au sein de l'Université mais ensuite pour m'avoir accueilli au sein du laboratoire. Merci aussi à César, Isabelle et Anne-Laure pour leur aide et leur disponibilité.

Merci à tous les doctorants qui m'ont accompagné durant ces trois années de thèse:

- A mes demi-frères de thèse du côté MFG, Samuel et Adrien, nos discussions en début de thèse m'ont permis de prendre confiance en moi et surtout de prendre du plaisir à parler de maths ;
- Du côté Stat, merci à Laetitia pour tous tes conseils et ton souci permanent à mon égard. Merci à Antoine et Claire pour nos discussions toujours très intéressantes. Bon courage à tous les deux, je suis certain que vous allez savoir profiter du reste de vos thèses respectives ; Merci aussi à Théophile avec qui j'ai eu l'occasion d'échanger depuis son arrivée récente !
- Merci à mon autre demi-frère de thèse, Grégoire. Je suis très heureux que nous ayons pu travailler ensemble et partager nos idées. Je te souhaite le meilleur pour la suite au Luxembourg ;
- Merci aussi à mes co-bureaux officiels, Chanqching, Xiaofei, Xiaozhen, Rodrigue et Marc-Antoine. Marc-Antoine, merci pour nos parties de tennis et m'avoir laissé quelques jeux avec tes double-fautes ;
- Merci à tous les jeunes-chercheurs que j'ai rencontré en conférence ou en école d'été et avec qui j'ai eu des échanges. Je voudrais en particulier mentionner Guillaume Le Mailloux et Alice l'Huillier pour nos discussions sur le Bayésien qui ont changé ma façon de penser aux stats ;
- Merci à Lucas, Charly, Théo, Emma, Tommy et Antoine pour nos nombreuses discussions sur des sujets variés ainsi que pour les bons moments passés. Un merci spécial pour Antoine qui n'a jamais arrêté de me relancer pour jouer au tennis;
- Merci à Nicolas avec qui j'ai passé énormément de temps durant cette thèse. Merci pour nos cafés-défense du matin et pour nos déjeuners à 11h40, évitant habilement la queue. J'espère que nous aurons l'occasion de continuer de nous voir à l'avenir et de partager de bons moments.

Merci enfin à Antoine Sihrener. À mon arrivée en prépa ECS, j'étais incertain de mes aspirations et même de mes centres d'intérêt. Rencontrer un enseignant aussi passionné et engagé a été une chance inestimable. Votre influence a, sans aucun doute, joué un rôle clé dans le chemin qui m'a conduit jusqu'à cette thèse. Je suis extrêmement heureux de vous avoir rencontré.

Après avoir suffisamment parlé de mes aventures académiques, il est grand temps de passer aux remerciements de mes proches. Ceux qui, malgré mes disparitions régulières, mes temps de réponse infinis et mon indisponibilité chronique (pour ne pas dire plus), ont continué de me supporter – au sens propre du terme.

Merci Emeric pour ton soutien constant. Sans toi, je n'aurais certainement pas eu le courage de me lancer dans cette thèse. Je suis extrêmement reconnaissant de tout ce que tu fais pour moi depuis des années maintenant. Merci de toujours te soucier de moi et d'être toujours présent quand j'en ai besoin.

Merci à Amaury, Robinson et Adrien d'être là depuis le lycée – et même avant. Chaque moment passé avec vous a été une vraie bouffée d'air pendant cette thèse. Merci de ne jamais abandonner l'idée

de me proposer de sortir, et d'arriver parfois à me traîner dehors de force – même si je cède rarement... j'apprécie toujours quand ça finit par arriver ! Je suis vraiment heureux de vous avoir toujours dans ma vie.

Merci à Anthony, Olivier, Jarod, Joséphine, Ariane et Sixtine d'avoir été là tout au long de cette thèse et d'avoir partagé tant de bons moments avec moi. Votre soutien a été essentiel, autant pour ma thèse que pour ma vie personnelle. Mention spéciale pour Ariane : nous avons commencé cette thèse en même temps et la terminons à une semaine d'intervalle. T'avoir comme amie tout au long de ce parcours m'a fait un bien fou.

Merci à Virginie, Alexandra et Alain pour toute leur affection. Merci d'avoir toujours pris des nouvelles de l'avancement de ma thèse et, surtout, merci à Alain de m'avoir souvent conseillé de "demander à Jean-Mi" quand j'étais bloqué. Vous m'avez vraiment été d'une aide immense. Un merci particulier à ma bofj préférée pour toutes nos brochettes qui m'ont redonné de l'énergie. J'en profite également pour remercier les familles Treuil, Bourrelly, Löfler et Sales pour leur gentillesse à mon égard.

Merci à mes grands-mères pour tout leur amour et leur soutien. Merci à Denise pour son attention constante, ses encouragements et pour les moments précieux que nous avons partagés ensemble. Merci à ma Dany, toujours prête à m'écouter, pour sa bonne humeur à chaque instant partagé, et surtout pour tout ce qu'elle a fait pour moi durant ma vie. Je ne serais pas la personne que je suis aujourd'hui sans elle.

Merci à mes parents. Il est difficile de résumer en quelques lignes tout ce que vous avez fait pour moi et qui m'a mené jusqu'à ce jour. Vous avez été présents à chaque instant. Vous m'avez toujours laissé choisir ce que je voulais faire et m'avez donné les moyens de réaliser tout ce que j'entreprenais. Je suis extrêmement chanceux et fier d'être votre fils, et la personne que je suis aujourd'hui vous doit absolument tout. Votre amour et votre soutien ont été, et sont encore, des piliers dans ma vie, et je ne pourrai jamais assez vous remercier pour tout ce que vous avez fait pour moi. Je vous serai à jamais reconnaissant d'avoir construit ce lien si fort entre nous quatre, cette famille, qui m'a toujours permis de me sentir écouté, soutenu et aimé. Merci pour tout.

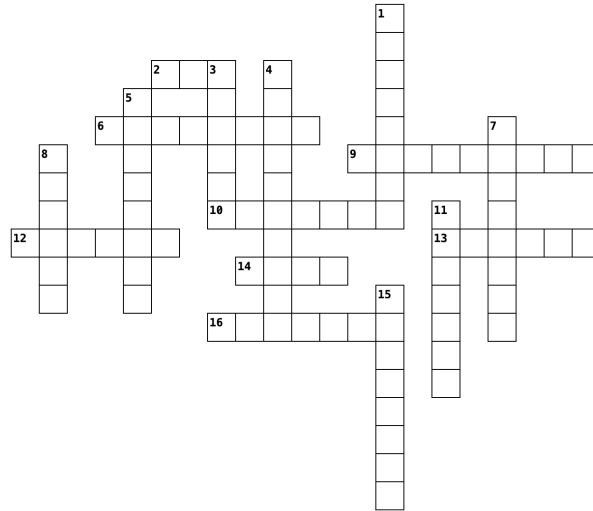
Merci à mon frère, Lucas. Notre relation est sans doute ce qui compte le plus dans ma vie. Merci d'être toujours là quand j'ai besoin de toi, de m'avoir soutenu tout au long de cette thèse, et d'avoir écouté mes histoires de maths à répétition, même quand ça ne devait pas toujours être passionnant. J'espère que notre petite compétition fraternelle continuera encore longtemps, parce qu'elle fait partie de ce qui nous rend si proches. Un grand Merci aussi à Maëva, bon courage pour supporter notre boule de nerf préférée ...

Je profite aussi de cette occasion pour remercier également mes deux autres frères, Chamuel et Skeffu, dont le rôle dans cette thèse n'est pas du tout négligeable.

Merci enfin à Marie-Julie. J'ai repoussé l'écriture de ce paragraphe aussi longtemps que possible, car il m'est difficile d'exprimer à quel point tu comptes pour moi. Merci d'avoir traversé ces trois années de thèse à mes côtés, de m'avoir écouté dans mes – nombreux – moments de déprime, d'angoisse, ou d'euphorie, souvent sans trop savoir pourquoi. Merci de m'accompagner à chaque étape, de m'aider à garder le cap quand tout me semble difficile. En si peu de temps, tu es devenue le centre autour duquel

mon monde gravite et cela ne changera pas de sitôt. Je te l'ai déjà dit mais c'est - je pense - la meilleure façon de résumer tout ça : Merci d'exister.

Maintenant que les remerciements sont terminés, certains d'entre vous trouveront peut-être le temps un peu long pendant cette soutenance ...



Horizontal

2. Elle codait avant l'ordinateur
6. Il pointe vers la vallée
9. Dans un ensemble mais tout seul
10. Père de la géométrie
12. Certainement la pire étape du manuscrit
13. Il possède son triangle
14. Groupe d'experts redouté
16. Se dit d'un nombre qui n'est divisible que par lui-même et par 1

Vertical

1. Effet secondaire de la lecture de certains papiers
3. Base d'un raisonnement
4. Proposition non démontrée mais supposée vraie
5. Processus de diffusion
7. Pionnier de l'intégration moderne
8. Ruban infini
11. Le plus petit des symboles
15. Sa structure est similaire à différentes échelles

Résumé

Cette thèse traite du comportement en temps long des équations stochastiques de Fokker-Planck en présence d'un bruit commun additif et présente des méthodes statistiques pour estimer la mesure invariante des processus de diffusion ergodiques multidimensionnels à partir de données bruitées. Dans la première partie, nous analysons les équations différentielles partielles stochastiques de type Fokker-Planck non linéaires, obtenues comme la limite du champ moyen de systèmes de particules en interaction dirigés par des bruits browniens idiosyncrasiques et en présence de bruit commun. Nous établissons des conditions sous lesquelles l'ajout d'un bruit commun permet de restaurer l'unicité de la mesure invariante. La principale difficulté provient de la dimension finie du bruit commun, alors que la variable d'état - *interprétée comme la loi marginale conditionnelle du système compte tenu du bruit commun* - opère dans un espace de dimension infinie. Nous démontrons que l'unicité est rétablie dès lors que le terme d'interaction du champ moyen attire le système vers sa moyenne conditionnelle (par rapport au bruit commun), en particulier lorsque l'intensité du bruit idiosyncrasique est faible, qui sont des cas typiques de perte d'unicité en l'absence de bruit commun.

Dans la deuxième partie, nous développons une méthodologie statistique afin d'approximer la mesure invariante d'un processus de diffusion à partir d'observations bruitées et discrètes de ce même processus. Cette méthode implique une technique de pré-moyennage des données qui réduit l'intensité du bruit tout en conservant les caractéristiques analytiques et les propriétés asymptotiques du signal sous-jacent. Nous étudions le taux de convergence de cet estimateur, qui dépend de la régularité anisotrope de la densité et de l'intensité du bruit. Nous établissons ensuite des conditions sur l'intensité du bruit qui permettent d'obtenir des taux de convergence comparables à ceux des cas sans bruit. Enfin, nous démontrons une inégalité de concentration de type Bernstein pour notre estimateur, ce qui permet de mettre en place une procédure adaptative pour la sélection de la fenêtre du noyau.

Mots clés : Équations de Fokker-Planck Stochastiques, Bruit Commun, Mesure Invariante, Estimation de Densité par Noyau, Processus de Diffusion Ergodique, Réduction du Bruit, Estimation Non-paramétrique

Abstract

This thesis deals with the long-time behavior of stochastic Fokker-Planck equations with additive common noise and presents statistical methods for estimating the invariant measure of multidimensional ergodic diffusion processes from noisy data. In the first part, we analyze stochastic Fokker-Planck Partial Differential Equations (SPDEs), obtained as the mean-field limit of interacting particle systems influenced by both idiosyncratic and common Brownian noises. We establish conditions under which the addition of common noise restores uniqueness of the invariant measure. The main challenge arises from the finite-dimensional nature of the common noise, while the state variable — *interpreted as the conditional marginal law of the system given the common noise* — operates within an infinite-dimensional space. We demonstrate that uniqueness is restored if the mean field interaction term attracts the system towards its conditional mean given the common noise, particularly when the intensity of the idiosyncratic noise is small. In the second part, we develop a new statistical methodology using kernel density estimation to effectively approximate the invariant measure from noisy observations, highlighting the crucial role of the underlying Markov structure in the denoising process. This method involves a pre-averaging technique that proficiently reduces the intensity of the noise while maintaining the analytical characteristics and asymptotic properties of the underlying signal. We investigate the convergence rate of our estimator, which depends on the anisotropic regularity of the density and the intensity of the noise. We establish noise intensity conditions that allow for convergence rates comparable to those in noise-free environments. Additionally, we demonstrate a Bernstein concentration inequality for our estimator, leading to an adaptive procedure for selecting the kernel bandwidth.

Keywords : Stochastic Fokker-Planck, Common Noise, Invariant Measure, Kernel Density Estimation, Ergodic Diffusion Processes, Noise Reduction, Non-parametric Estimation

Contents

Résumé succinct en Français	1
1 Introduction	9
1 The space of probability measures	10
1.1 Wasserstein metrics	10
1.2 Differential calculus on $\mathcal{P}_2(\mathbb{R}^d)$.	11
2 Non-linear stochastic Fokker-Planck equation	13
2.1 Between macroscopic and microscopic scales: interacting particle systems, the origin of the model	14
2.1.1 Introduction of the interacting particle system with common noise	16
2.1.2 On the importance of common noise	17
2.1.3 Conditional propagation of chaos	19
2.2 McKean-Vlasov/Fokker-Planck ecosystem	21
2.3 Long time behaviour: reminders on the case without common noise	23
2.3.1 Long time behaviour	24
2.4 Contributions of the first part	26
3 Statistical Analysis of diffusion models	32
3.0.1 Ergodicity for diffusion models with constant volatility	33
3.1 Presentation of the model and literature review	34
3.2 Contributions	35
4 Composition of the manuscript	39
I Probabilistic analysis of interacting particle systems with common noise	41
2 Long time behavior of non-linear stochastic Fokker-Planck Equation	43
1 Introduction	45
1.1 The process m as a mean field limit	50
2 Existence of an invariant measure for the stochastic flow of measures.	50
2.1 The existence result	53
3 The case of a uniformly convex confinement potential	55
3.1 Uniform in time propagation of chaos	56
3.2 Uniqueness of the invariant measure	58
3.3 Example	60
4 Non-convex potential without idiosyncratic noise	61
4.1 Existence of an invariant measure	62
4.2 Uniqueness of the invariant measure and exponential decay	63
5 Appendix	70

5.1	Proof of Proposition 2.8	70
5.2	Proof of Lemma 2.9	70
5.3	A measurable selection result	71
5.4	Proof of Proposition 2.22	72
5.5	Proof of Lemma 2.23	73
3	Ergodicity of some stochastic Fokker-Planck equations with additive common noise	75
1	Introduction	77
1.1	Deterministic Fokker-Planck equations	77
1.2	Common noise	78
1.3	Our contribution	79
1.4	Further connections with the literature on mean field models	80
1.5	Organisation of the paper	80
1.6	Notation	80
2	Main result: uniqueness recovery thanks to common noise	82
2.1	Stochastic Fokker-Planck PDE and Conditional McKean-Vlasov SDE	82
2.2	Assumptions, existence and uniqueness of solutions	83
2.3	Restoration of uniqueness	85
2.4	Discussions and examples	91
3	Proof of Theorem 3.8	95
3.1	First preliminary estimate: conditional variance	95
3.2	Second preliminary estimate: conditional exponential moments	97
3.3	Completion of the proof of Theorem 3.8	99
4	Proof of Proposition 3.7	100
4.1	Ansatz	100
4.2	Contraction for classical diffusion processes	100
4.3	Proof of Proposition 3.7	102
5	Appendix	112
5.1	Proof of Equation (3.38)	112
5.2	Proof of Equation (3.55)	113
II	Statistical analysis of diffusion models in presence of noise	115
4	Invariant density estimation from noisy data	117
1	Introduction	119
1.1	Setting	119
1.2	Motivation	119
1.3	Organization of the paper	121
2	Statistical and Probabilistic framework	122
2.1	Notation	122
2.2	The statistical diffusion model	122
3	Estimation procedure	124
4	Upper bounds and hyper-parameter choice	128
4.1	General results	129
4.2	Analysis of the convergence rate	131
4.3	Comments	133
5	Bernstein inequality and adaptive choice of the hyperparameters	133
5.1	Bernstein inequality	133
5.2	Choice of the hyperparameters	134

6	Numerical Analysis	135
7	Probabilistic tools: the Markovian structure of the preaveraged process	137
8	Invertibility of A	139
9	Proof of Proposition 4.3	139
10	Proof of Proposition 4.4	151
11	Proof of Proposition 4.5	152
	11.1 Structure and completion of the proof	152
	11.2 Proof of Lemma 4.26	158
	11.3 Proof of Lemma 4.27	159
	11.4 Proof of Lemma 4.28	160
	11.5 Proof of Lemma 4.29	161
12	Proof of the results of Section 4.1	163
	12.1 Preliminary results	163
	12.2 Proof of Proposition 4.7	164
	12.3 Proof of Proposition 4.8	165
	12.4 Proof of Proposition 4.10	166
	12.5 Proof of Proposition 4.11	167
13	Proof of the results of Section 5	168
	13.1 Proof of Theorem 4.12	168
	13.2 Proof of Proposition 4.14	170
14	Proof of the results of Section 7	175
	14.1 Proof of Lemma 4.17	176
	14.2 Proof of Lemma 4.18	179
	14.3 Proof of Corollary 4.19	179
	14.4 Proof of Lemma 4.20	179
Appendices		181
A Coupling by Reflection		183
1	Probabilistic context	183
2	The coupling method	184
B Kernel estimation - Adaptive methods		187
1	Framework and Kernel estimator	187
	1.1 Case of a known regularity	188
	1.2 Case of an unknown regularity	190
	1.3 The Goldenshulger-Lepski procedure	190
	1.3.1 Heuristic idea	190
	1.4 The procedure	191
References		195

Résumé succinct en français

This is the only chapter in French

Cette thèse est composée de deux parties :

- la première aborde les questions probabilistes en relation avec des systèmes de particules en interaction soumis à un bruit commun. Plus précisément, nous examinons dans cette première partie l'influence du bruit commun sur le comportement en temps long de la limite champ moyen de ces systèmes.
- Dans la seconde partie, nous nous penchons sur des questions d'ordre statistique liées à des modèles de diffusion ergodique. En particulier, on cherche à comprendre l'impact d'un bruit venant polluer les observations d'un processus de diffusion multidimensionnel sur la vitesse d'estimation de la mesure invariante du processus sous-jacent.

Partie 1 : Comportement en temps long des équations Fokker-Planck non linéaires stochastiques

La première partie est composée de deux papiers de recherche. L'un d'entre eux a été réalisé en collaboration avec François Delarue et Étienne Tanré. Ces deux travaux s'inscrivent dans une démarche visant à comprendre l'impact potentiel d'une corrélation entre les diffusions de différentes particules d'un système sur le comportement en temps long de celui-ci. Plus précisément, l'objectif est de répondre à la question suivante :

Comment un bruit commun affecte-t-il le comportement en temps long des systèmes en interaction avec un grand nombre de particules ?

Ainsi, dans la première partie de ce manuscrit, on considère un système de la forme

$$\begin{cases} dX_t^i = b(X_t^i, m_t^N)dt + \sigma dB_t^i + \sigma_0 dB_t^0 \\ \mathcal{L}(X_0^1, \dots, X_0^N) = \mu_0^{\otimes N}, \end{cases}$$

avec

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

Les $(B^i)_i$ sont des mouvements browniens indépendants en dimension d , ce sont les bruits *indiosyncratiques*, et B^0 est un mouvement brownien standard indépendant des B^i , appelé bruit *commun*. σ et σ_0 sont des coefficients positifs.

Chaque particule est représentée par une quantité $X^i \in \mathbb{R}^d$, qui correspond généralement à sa position (X_t^i indique alors la position de la particule i au temps t). La fonction $b : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ est appelée *drift* ou *transport*. Il est courant (et c'est ce que nous ferons dans la première partie de ce manuscrit), lorsque l'on considère le comportement en temps long de ce type de systèmes, de supposer que b s'écrit de la manière suivante :

$$b(x, \nu) = G(x) + F(x, \nu).$$

Dans ce cas, G est appelé force de *confinement*, c'est une force extérieure qui dirige le mouvement des particules. La fonction F est dite force d'*interaction* et définit la manière dont chaque particule interagit avec les autres.

À ce stade, il est important de noter que dans la suite, nous ne considérerons que des cas où l'interaction entre les particules du système est de type *champ moyen*, c'est-à-dire qu'elle intervient à travers la mesure empirique du système. Cela empêche deux particules d'interagir entre elles sans prendre en compte le comportement des autres: l'interaction ne peut intervenir qu'à travers le comportement moyen du système au temps donné. Ce type d'interaction est usuel et confère de bonnes propriétés asymptotiques (*propagation du chaos*), que nous détaillerons dans la suite. Lorsque le nombre de particules devient très grand, on peut essayer de caractériser le comportement du système en étudiant sa mesure empirique. On obtient alors par des arguments de compacité [CD18a, CD18b] que $(m_t^N)_{t \geq 0}$ admet une limite lorsque $N \rightarrow +\infty$, cette limite, notée $(m_t)_{t \geq 0}$, est solution (*en un certain sens*) de

$$dm_t = \frac{\sigma^2 + \sigma_0^2}{2} \Delta m_t dt - \nabla \cdot (m_t b(\cdot, m_t)) dt - \sigma_0 \nabla m_t \cdot dB_t^0. \quad (1)$$

Le flot $(m_t)_{t \geq 0}$ est ainsi un processus stochastique à valeur dans l'espace des mesures de probabilité, processus dont nous étudierons le comportement en temps long. Il est important de bien noter que pour étudier le comportement en temps long de ce processus, nous allons en réalité nous élever d'un espace et étudier le flot de mesure généré par sa loi. En particulier, nous allons nous intéresser au comportement en temps long de $(P_t)_{t \geq 0} := (\mathcal{L}(m_t))_{t \geq 0}$. La difficulté et l'originalité de cette question résident alors dans le fait que ce flot de mesures évolue dans $\mathcal{P}(\mathcal{P}(\mathbb{R}^d))$.

→ **Résumé du chapitre 2.**

Le chapitre 2 correspond au papier [Mai23]. Dans ce chapitre on considère le cas où le terme de transport b s'écrit de la manière suivante:

$$b(x, \mu) = -\nabla V(x) - \nabla W * \mu(x),$$

pour deux fonctions différentiables V et W et où $*$ désigne l'opérateur de convolution. L'équation (1) se réécrit alors

$$dm_t = \frac{\sigma^2 + \sigma_0^2}{2} \Delta m_t dt + \nabla \cdot (m_t (\nabla V + \nabla W * m_t)) dt - \sigma_0 \nabla m_t \cdot dB_t^0. \quad (2)$$

Dans le cas $\sigma_0 = 0$, le processus $(m_t)_t$ est simplement un flot de mesures déterministe et son comportement en temps long a été très étudié, aussi bien avec des approches EDP [CMV03, CMV06] que des approches probabilistes [Mal01, CGM08]. En particulier, il est bien connu (voir par exemple [Mal01]) que lorsque le potentiel V est strictement convexe et W est convexe, avec $\sigma_0 = 0$, l'équation (1) admet une unique solution stationnaire. L'idée sous jacente est que $(m_t)_{t \geq 0}$ peut- alors être vu comme le flot associé à la loi de X solution de

$$dX_t = -\nabla V(X_t) dt - \nabla W * \mathcal{L}(X_t) dt + \sigma d\beta_t, \quad (3)$$

pour un mouvement brownien d -dimensionnel β . Lorsque le potentiel V est uniformément convexe, le processus X est fortement attiré vers l'unique minimiseur de V assurant ainsi un comportement asymptotique unique. La première question traitée dans ce chapitre est alors de montrer que de tels résultats d'unicités sont préservés lorsqu'on ajoute du bruit commun dans le système. Plus précisément on montre dans [Mai23] que sous des hypothèses de convexité fortes, le flot de mesure $(P_t)_t$ converge toujours vers la même limite (en un certain sens), cette limite notée \bar{P} est alors l'unique mesure invariante associée au processus $(m_t)_{t \geq 0}$ avec dynamique (1). Plus précisément, on obtient

$$d_2^{\mathcal{P}(\mathbb{R}^d)}(P_t, \bar{P}) \leq e^{-ct} d_2^{\mathcal{P}(\mathbb{R}^d)}(P_0, \bar{P}),$$

pour toute condition initiale $P_0 \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$ et où $d_2^{\mathcal{P}(\mathbb{R}^d)}$ est la distance de Wasserstein-2 sur $\mathcal{P}_2(\mathbb{R}^d)$.

Ces résultats, assez naturels, sont les premiers concernant le comportement en temps long de l'équation (1) pour laquelle la source de bruit est fini-dimensionnelle. D'autres résultats, récents, existent dans le cas où le bruit commun est un mouvement Brownien cylindrique [ABKO23, ACKO24], et donc de dimension infinie.

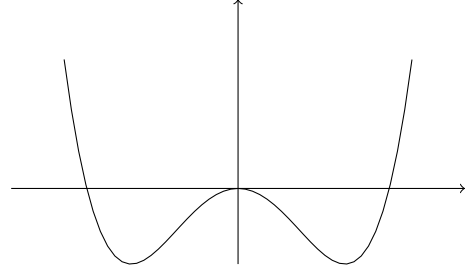
La suite du chapitre traite d'un cas particulier : l'absence de bruit idiosyncratique, $\sigma = 0$. Dans ce cas, l'équation d'intérêt se réécrit

$$dm_t = \frac{\sigma_0^2}{2} \Delta m_t dt + \nabla \cdot (m_t (\nabla V + \nabla W * m_t)) dt - \sigma_0 \nabla m_t \cdot dB_t^0.$$

Nous portons notre attention sur ce cas car il est alors clair que quand $\sigma_0 = 0$, le flot de mesures (déterministe) est solution de l'équation suivante

$$\partial_t m_t = \nabla \cdot (m_t (\nabla V + \nabla W * m_t)). \quad (4)$$

Lorsque V n'est pas supposé convexe, pensons par exemple à un double puit, il est facile de voir que l'équation (4) admet plusieurs solutions stationnaires. Par exemple, pour $V : x \mapsto |x|^4/4 - |x|^2/2$, on voit bien que δ_{-1} et δ_1 sont des solutions stationnaires.



Se pose alors naturellement la question suivante :

L'ajout de bruit commun peut-il permettre de récupérer l'unicité de la mesure invariante ? Si oui, dans quel(s) modèle(s) ?

La réponse à cette question peut sembler évidente au premier abord. En effet, il existe de nombreux exemples pour lesquels l'ajout de bruit restaure l'unicité du comportement asymptotique de particules en interaction. Cependant, la difficulté ici réside dans le fait que le bruit commun possède une structure particulière et vient modifier subtilement la dynamique : le bruit commun est un mouvement brownien dans \mathbb{R}^d agissant sur la dynamique d'un processus à valeurs dans $\mathcal{P}(\mathbb{R}^d)$. Cette asymétrie entre la dimension du processus d'intérêt et de la source de bruit est à l'origine de notre question.

Dans la suite du chapitre 2, nous démontrons l'existence et l'unicité de la mesure invariante lorsque $\sigma_0 > 0$, de plus on donne explicitement cette unique mesure invariante:

$$\bar{P} = \int_{\mathbb{R}^d} \delta_{\delta_a} m^*(da), \tag{5}$$

où m^* est la solution stationnaire de l'équation de Fokker-Planck linéaire avec terme de transport $-\nabla V$ et coefficient de diffusion σ_0 . On montre également que la vitesse de convergence vers le régime stationnaire est exponentielle en temps pour la distance $d_1^{\mathcal{P}(\mathbb{R}^d)}(\cdot, \cdot)$.

→ **Résumé du chapitre 3.**

Le chapitre 3 est issue d'une collaboration avec François Delarue et Etienne Tanré. L'idée provient d'une volonté de mieux comprendre le phénomène sous-jacent qui a mené au résultat de récupération d'unicité du chapitre 2. On considèrera néanmoins un cas un peu plus général. Plus précisément, nous nous intéressons au comportement en temps long du flot de mesure généré par l'équation

$$dm_t = \frac{\sigma_0^2 + \sigma^2}{2} \Delta m_t dt - \nabla \cdot (m_t (G + F(\cdot - \mu_1(m_t)))) dt - \sigma_0 \nabla m_t \cdot dB_t^0, \tag{6}$$

avec

$$\mu_1(m) = \int_{\mathbb{R}^d} xm(dx) \in \mathbb{R}^d,$$

pour tout $m \in \mathcal{P}_1(\mathbb{R}^d)$. Ce modèle est une généralisation de celui du chapitre précédent à de nombreux égards. La modification la plus importante est la prise en compte de modèles pour lesquels $\sigma > 0$. Ce cas est d'un intérêt primordial au regard des travaux [Daw83, HT10] qui exhibent des cas pour lesquels, même avec $\sigma > 0$, il existe plusieurs solutions invariantes pour (m_t) solution de (6) avec $\sigma_0 = 0$. On

se pose alors exactement la même question que dans le chapitre précédent, l'ajout de bruit commun peut-il permettre de récupérer l'unicité du comportement asymptotique ? Ce cas présente une difficulté supplémentaire par rapport au cas du chapitre 2. En effet, la présence de bruit idiosyncratique ($\sigma > 0$) nous empêche complètement d'identifier la mesure invariante de la même manière que dans (5).

Dans ce chapitre nous montrons que dans des régimes où σ est supposé assez petit, la dynamique du processus $(m_t)_{t \geq 0}$ est fortement dirigée par la dynamique de sa moyenne $(\mu_1(m_t))_{t \geq 0}$. Le point positif découlant de cette observation est que le processus $(\mu_1(m_t))_{t \geq 0}$ évolue dans un espace de dimension finie. On montrera alors que l'ajout de bruit commun permet de récupérer de l'ergodicité pour $(\mu_1(m_t))_{t \geq 0}$, ergodicité se propageant alors à tout le flot de mesures $(m_t)_{t \geq 0}$.

Une restriction importante néanmoins est que le modèle considéré dans ce chapitre est assez particulier car la non linéarité n'intervient qu'à travers la moyenne. Même s'il est clair que l'ajout de bruit commun ne peut pas permettre de restaurer l'unicité de la mesure invariante dans tous les modèles, il serait intéressant d'étudier des modèles plus généraux où la dynamique dépend de plusieurs moments. Par exemple, on pourrait envisager d'ajouter un bruit commun de dimension finie dans des dynamiques du type [Sch85a, Sch85b, Sch93], présentant des propriétés de périodicité.

Partie 2: Analyse statistique pour des modèles diffusifs ergodiques

→ Résumé du chapitre 4

Le dernier travail de cette thèse s'inscrit dans une volonté de bien appréhender le comportement en temps long des processus de diffusions. Il est tiré d'un article co-écrit avec Grégoire Szymanski. Dans ce papier on considère le processus de Markov suivant

$$dX_t = -\nabla V(X_t)dt + \sigma dB_t. \tag{7}$$

Sous des conditions de croissance pour le potentiel V , et dès lors que $\sigma > 0$, il est bien connu que le processus X est ergodique et admet une unique mesure invariante, absolument continue par rapport à la mesure de Lebesgue de densité

$$\bar{\mu}(x) \propto \exp\left(-\frac{2}{\sigma^2}V(x)\right).$$

Les questions de l'estimation du potentiel V et de la mesure invariante ont été le sujet de nombreuses recherches en statistique des processus ces dernières années [AG21, AG22, AG23, Str18, DR06, DR07]. Le chapitre 4 s'appuie pleinement sur ces nombreuses références en essayant d'étendre encore le cadre statistique. A l'image des papiers [Sch11, Sch12], le but de ce chapitre est de mettre en place des méthodes statistiques permettant de prendre en compte la présence d'un éventuel bruit dans les observations. Il est naturel que les données observées soient bruitées, c'est un phénomène courant dans de nombreuses applications pratiques telles que la finance, la biologie ou encore l'analyse de données de capteurs. Ce bruit peut provenir d'erreurs de mesure ou de perturbations extérieures, et peut avoir un impact significatif sur les résultats obtenus. Il est donc important de prendre en compte le bruit dans l'analyse et la modélisation de ces systèmes diffusifs. On répondra à la question suivante:

Comment la présence de bruit dans les données observées affecte-t-elle la vitesse d'estimation de la mesure invariante ?

Pour répondre à cette question nous considérerons le schéma d'observation suivant:

- observations discrètes à des temps fixés $k\Delta_n$ pour $k \in \{0, \dots, n\}$;
- observations haute fréquence : $\Delta_n \rightarrow 0$, et en temps long : $n\Delta_n \rightarrow +\infty$;
- observations bruitées.

En définitive, notre but est d'estimer la mesure invariant du processus X solution de l'équation (7) à partir de l'observation des $(Y_i)_{i=1}^n$ avec

$$Y_i = X_{i\Delta_n} + \tau_n \xi_i, \quad i \in \{1, \dots, n\}.$$

Nous sommes donc confrontés à un problème de déconvolution, bien connu et traité dans la littérature. Les modèles de déconvolution sont réputés pour être difficiles à résoudre en pratique et les taux de convergence sont souvent lents (logarithmiques) [CL13]. L'intérêt de ce travail est de montrer que dans ce cas, la structure markovienne du problème permet une meilleure extraction de l'information et donc d'obtenir des vitesses de convergence polynomiales, et ce même lorsque l'intensité du bruit est grande.

Pour obtenir ces vitesses, la stratégie adoptée dans ce chapitre est de procéder à un pré-traitement des données visant à réduire l'impact du bruit dans les observations, plus précisément, l'idée est de moyennner les observations par paquets de taille p , afin d'obtenir les observations suivantes:

$$\left(\frac{1}{p} \sum_{i=0}^{p-1} Y_{kp+i} \right)_{k=0}^{\lfloor n/p \rfloor - 1}. \quad (8)$$

À partir de ces données modifiées, nous mettons en place une procédure d'estimation par noyau classique combinée à une méthode de réduction du biais. Plus précisément, en considérant un noyau K , on définit pour toute fenêtre $\mathbf{h} = (h_1, \dots, h_d) \in (0, +\infty)^d$, for any $x \in \mathbb{R}^d$,

$$\nu_{n,\mathbf{h},p}(x) = \frac{1}{\lfloor n/p \rfloor} \sum_{k=0}^{\lfloor n/p \rfloor - 1} \mathbf{K}_{\mathbf{h}} \left(x - \frac{1}{p} \sum_{i=0}^{p-1} Y_{kp+i} \right),$$

avec la structure suivante : $\forall y \in \mathbb{R}^d$ and $\mathbf{h} \in (0, +\infty)^d$,

$$\mathbf{K}_{\mathbf{h}}(y) = \prod_{i=1}^d h_i^{-1} K(h_i^{-1} y_i).$$

La mise en place d'un tel estimateur génère un terme non usuel dans l'étude du biais. Pour se défaire de ce terme supplémentaire, on met en place une procédure de débiaisage qui sera détaillée dans le chapitre 4. Finalement, on obtient l'estimateur suivant

$$\hat{\mu}_{n,\mathbf{h},p}(x) := \sum_{\gamma} \mathbf{u}_{\gamma} \nu_{n,\mathbf{h},p}(x),$$

pour une suite $(u_\gamma)_\gamma$ construite dans le but de diminuer l'influence du bruit et du pré-averaging des données. On étudie ensuite la qualité d'estimation en démontrant des bornes supérieures pour le risque quadratique ponctuel. Les résultats obtenus sont cohérents avec les résultats minimax dans le cas où l'intensité du bruit est supposée faible. Dans les cas où l'intensité du bruit est trop importante, on voit apparaître de nouvelles vitesses d'estimation, polynomiales en l'intensité du bruit.

Chapter 1

Introduction

Abstract

This introductory chapter begins by outlining the primary objectives and key focus areas of this thesis. It provides context and background information to set the stage for the subsequent discussions. Following this, the chapter presents a summary of the main results achieved throughout the manuscript.

Contents

1	The space of probability measures	10
1.1	Wasserstein metrics	10
1.2	Differential calculus on $\mathcal{P}_2(\mathbb{R}^d)$	11
2	Non-linear stochastic Fokker-Planck equation	13
2.1	Between macroscopic and microscopic scales: interacting particle systems, the origin of the model	14
2.2	McKean-Vlasov/Fokker-Planck ecosystem	21
2.3	Long time behaviour: reminders on the case without common noise	23
2.4	Contributions of the first part	26
3	Statistical Analysis of diffusion models	32
3.1	Presentation of the model and literature review	34
3.2	Contributions	35
4	Composition of the manuscript	39

The structure of this thesis is mainly divided into two parts. The first part is dedicated to the analysis of the behaviour of solutions to some Stochastic Partial Differential Equations. The second part shifts its focus towards statistical methods, specifically dealing with estimation in diffusion models in the presence of noise.

- The focus of the first part is the long-time behaviour of the solutions to the non-linear stochastic Fokker-Planck equation. This equation can be thought of within an ecosystem consisting of a stochastic partial differential equation, a partial differential equation that is equivalent to it in some way, and a system of stochastic differential equations composed of N interacting diffusion processes with common noise that serve as a mean-field approximation in a large population limit.
- The second part deals with statistical estimation for multidimensional ergodic diffusion processes. More precisely, it consists of a new methodology based on Kernel density estimation to be able to approximate in a non-parametric way the invariant measure of an ergodic Markov process from noisy observations over time.

We begin this introduction chapter with reminders on the space of measures, which is a central space in the whole manuscript.

1 The space of probability measures

1.1 Wasserstein metrics

Throughout the manuscript, for any separable metric space (E, d) , we use $\mathcal{P}(E)$ to denote the space of Borel probability measures on E . There are multiple methods to define a metric on the space $\mathcal{P}(E)$, such as the Lévy-Prokhorov metric, the bounded-Lipschitz metric, or the Wasserstein metric, see [Zol76, Bog07, Dud02] for references on the subject. In this manuscript, we will focus solely on the latter. The definition of the Wasserstein metric is based on the idea of coupling. For $\mu, \nu \in \mathcal{P}(E)$, let $\Pi(\mu, \nu)$ stands for the transport plan between μ and ν . Precisely, $\Pi(\mu, \nu)$ is the set of Borel probability measures on $E \times E$ with first marginal μ and second marginal ν .

For $p \geq 1$, define $\mathcal{P}_p(E)$ to be the set of probability measures $\mu \in \mathcal{P}(E)$ satisfying

$$\int_E d(x, x_0)^p \mu(dx) < \infty,$$

for some point $x_0 \in E$, the choice of which does not influence the definition of the space. The p -Wasserstein metric on $\mathcal{P}_p(E)$ is defined by

$$d_p^E(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{E \times E} d(x, y)^p \pi(dx, dy) \right)^{1/p}.$$

Given a non-atomic probability space $(\Omega, \mathcal{F}, \mathbb{P})$, an equivalent definition reads

$$d_p^E(\mu, \nu) = \inf \mathbb{E} [d(X, Y)^p]^{1/p},$$

where the infimum runs over the set of random variables (X, Y) on $(\Omega, \mathcal{F}, \mathbb{P})$, such that $\mathcal{L}(X) = \mu$ and $\mathcal{L}(Y) = \nu$. It is well known, see for example [Vil09, AGS08], that d_p^E is a distance on $\mathcal{P}_p(E)$ which is compatible with weak convergence so that we can view $\mathcal{P}_p(E)$ itself as a metric space. Moreover, as soon as (E, d) is a complete and separable space, so is $(\mathcal{P}_p(E), d_p^E)$. Finally, we say that d_p^E is compatible with weak convergence in the sense that for a given a sequence $(m_n) \in \mathcal{P}_p(E)^\mathbb{N}$,

$$\lim_{n \rightarrow +\infty} d_p^E(m_n, m) = 0 \iff \begin{cases} m_n \text{ weakly converges to } m, \\ (m_n) \text{ has uniformly integrable } p\text{-moments.} \end{cases}$$

Also, for any probability measure $m \in \mathcal{P}(E)$ and any measurable $\varphi : E \rightarrow \mathbb{R}$ (so that the following quantity makes sense), we define the duality product $\langle \cdot ; \cdot \rangle_E$ as

$$\langle m ; \varphi \rangle_E = \int_E \varphi(x) m(dx).$$

1.2 Differential calculus on $\mathcal{P}_2(\mathbb{R}^d)$.

In this manuscript, we often work with a stochastic flow of probability measures. To study the behaviour of this stochastic process, we want to derive some properties from its law. To do so, we will repeatedly use the notion of derivative in the space of measures, mainly in the space of probability measures that admit a second moment of order 2. See however [LSZ23, Lac18] for a careful definition on the whole space of probability measures $\mathcal{P}(\mathbb{R}^d)$. Let us begin with two notions of differentiability in the space of measures.

Definition 1.1. *We say that $\Phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ admits a L -derivative, whenever there exists a unique jointly continuous function $\delta_m \Phi : (m, x) \in \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \rightarrow \delta_m \Phi(m, x) \in \mathbb{R}^d$, at most of quadratic growth in x for any m , such that*

$$\lim_{h \rightarrow 0} \frac{\Phi(m + h(m' - m)) - \Phi(m)}{h} = \int_{\mathbb{R}^d} \delta_m \Phi(m, v)(m' - m)(dv),$$

for all $m, m' \in \mathcal{P}(\mathbb{R}^d)$ and

$$\int_{\mathbb{R}^d} \delta_m \Phi(m, v)m(dv) = 0, \quad m \in \mathcal{P}_2(\mathbb{R}^d).$$

This definition is also called *linear functional derivative*, see for example [JT21] for a nice introduction. One can also think of it as a derivative in the direction of δ_x more precisely, for any $m \in \mathcal{P}_2(\mathbb{R}^d)$,

$$\delta_m \Phi(m, x) = \lim_{\rho \rightarrow 0^+} \frac{\Phi((1 - \rho)m + \rho\delta_x) - \Phi(m)}{\rho},$$

and for any $m, \tilde{m} \in \mathcal{P}_2(\mathbb{R}^d)$,

$$\Phi(m) = \Phi(\tilde{m}) + \int_0^1 \int_{\mathbb{R}^d} \delta_m \Phi((1 - t)m + t\tilde{m}, y)(m - \tilde{m})(dy). \quad (1.1)$$

A way to consider this derivative is that it is the natural notion of derivative whenever the space

of probability measures is *equipped with the geometry of the straight paths*, i.e the paths of the form $((1-t)m + (1-t)\tilde{m})_{t \in [0,1]}$.

Let us now introduce another notion of differentiability called the intrinsic differentiability:

Definition 1.2. We say that $\Phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ is $\mathcal{C}^1(\mathcal{P}_2(\mathbb{R}^d))$ in the intrinsic sense whenever it admits an L -derivative and for any $m \in \mathcal{P}_2(\mathbb{R}^d)$, the mapping $x \mapsto \delta_m \Phi(m, x)$ is differentiable, with the gradient being denoted $D_m \Phi(m, x) := \nabla \delta_m \Phi(x)$; the mapping $(m, x) \mapsto D_m \Phi(m, x)$ is jointly continuous in (m, x) and at most of linear growth in x uniformly in m in bounded subsets of $\mathcal{P}_2(\mathbb{R}^d)$.

The natural way to understand this definition and to compare it with the one of Definition 1.1 is via displacement of measure. More precisely, the following proposition stands

Proposition 1.3. Let $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ be Borel measurable and bounded assume that Φ admits an intrinsic derivative. Then

$$\lim_{h \rightarrow 0^+} \frac{\Phi((id + h\varphi)\#m) - \Phi(m)}{h} = \int_{\mathbb{R}^d} D_m \Phi(m, x) \cdot \varphi(x) m(dx) \quad \forall m \in \mathcal{P}_2(\mathbb{R}^d),$$

where for any Borel set A of \mathbb{R}^d ,

$$\phi\#m(A) = m(\phi^{-1}(A)).$$

The previous proposition enables us to interpret the notion of intrinsic derivative as being associated with a particular geometry on the space of measures, similar to what we did for the L -derivative. However, in this case, we are not limited to the straight paths into the space of measures, but rather we can move along all paths of the form $((Id + t\phi)m)_{t \in [0,1]}$ for some Borelian and bounded function ϕ . This is just another point of view that appears as naturally as the one of Definition 1.1.

Another way to understand the intrinsic derivative is through the lifting of the map to a space of random variables: let us fix a generic probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and define $L^2(\Omega)$ as the set of random variables with a finite moment of order 2. Given a map $\Phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$, we define its lifting $\tilde{\Phi} : L^2(\Omega) \rightarrow \mathbb{R}$ by

$$\tilde{\Phi}(X) := \Phi(\mathcal{L}(X)) \quad \forall X \in L^2(\Omega).$$

Then we have the following proposition,

Proposition 1.4 ([CD18a, GT19]). Assume that $\Phi \in \mathcal{C}^1(\mathcal{P}_2(\mathbb{R}^d))$ (in the intrinsic sense). Then $\tilde{\Phi}$ is Fréchet differentiable with a continuous Fréchet derivative in $L^2(\Omega)$ with

$$\nabla \tilde{\Phi}(X) = D_m \Phi(\mathcal{L}(X), X) \quad \forall X \in L^2(\Omega).$$

In this manuscript, we will always use this notion, and will no longer be precise on the underlying sense of differentiability. We finally give the following definition which will be referred to in the whole chapter:

Definition 1.5. We say that $\Phi \in C_b^2(\mathcal{P}(\mathbb{R}^d))$ if $\Phi : \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}$ and such that:

- The function Φ belongs to $\mathcal{C}_1(\mathcal{P}(\mathbb{R}^d))$;
- For every fixed $x \in \mathbb{R}^d$, each component of the \mathbb{R}^d -valued function $m \mapsto D_m \Phi(m, x)$ satisfies the properties described in definition 1.1, resulting in some $D_m^2 \Phi(m, x, y) \in \mathbb{R}^{d \times d}$. The function $D_m^2 \Phi$ is bounded and continuous.
- For $m \in \mathcal{P}(\mathbb{R}^d)$, we write $D_x D_m \Phi(m, x)$ for the Jacobian of the function $x \mapsto D_m \Phi(m, x)$ and assume it to be continuous and bounded in (m, x) .

Remark 1.6. One may develop a similar theory of differentiation for functions on $\mathcal{P}_p(\mathbb{R}^d)$, for any $p \geq 1$, as long as one is careful to require that the derivative $\delta_m \Phi$, satisfies some growth conditions to ensure that the integral on the right-hand side of Equation (1.1) is well-defined.

The reader unfamiliar with this notion can refer to the following illuminating references for a detailed treatment of the subject [Car10], [CD18a, Chap. 5].

Example 1.7. In the sequel, we will frequently use the notion of derivative for functions with the following structure:

$$\Phi : m \mapsto \int_{\mathbb{R}^d} \phi(x) m(dx), \quad (1.2)$$

for some differentiable function $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ that smooth and bounded (or has controlled asymptotic growth, depending on the integrability assumptions of m so that this object makes sense). In this case, for all $x \in \mathbb{R}^d$,

$$D_m \Phi(m, x) = \nabla_x \phi(x).$$

2 Non-linear stochastic Fokker-Planck equation

In the first part of the manuscript, we are interested in the solutions of the non-linear stochastic Fokker-Planck Equation

$$d_t m_t = \nabla \cdot \left(\frac{\sigma^2 + \sigma_0^2}{2} \nabla m_t - m_t b(\cdot, m_t) \right) dt - \sigma_0 \nabla m_t \cdot dB_t^0, \quad \text{on } [0, +\infty) \times \mathbb{R}^d, \quad (1.3)$$

where $B^0 = (B_t^0)_{t \geq 0}$ is a d -dimensional Brownian motion, called *common noise*. Above, the unknown $(m_t)_{t \geq 0}$ is regarded as a stochastic process with values in the space of probability measures $\mathcal{P}(\mathbb{R}^d)$ on \mathbb{R}^d . Accordingly, the coefficient b is a function from $\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$ to \mathbb{R}^d depending on both space and measure arguments. The two parameters σ and σ_0 are non-negative scalars: the former is the intensity of the so-called *idiosyncratic noise* and the latter is the intensity of the *common noise*. In the whole manuscript, Equation (1.3) is understood in the weak sense, i.e., the process $(m_t)_{t \geq 0}$ satisfies for all $t \geq 0$ and $\varphi \in C_c^\infty(\mathbb{R}^d)$, the space of infinitely differentiable functions with compact support,

$$d \langle m_t, \varphi \rangle = \langle m_t, \mathcal{L}_{m_t} \varphi \rangle dt + \sigma_0 \langle m_t, (\nabla \varphi)^\top \rangle dB_t^0,$$

where for any probability measure $m \in \mathcal{P}(\mathbb{R}^d)$, the operator \mathcal{L}_m acts on a smooth function φ with compact support in the following manner:

$$\mathcal{L}_m \varphi = -b(\cdot, m) \cdot \nabla \varphi + \frac{\sigma_0^2 + \sigma^2}{2} \Delta \varphi.$$

To study the time evolution of this stochastic flow of measures, it is natural to study the flow of measures on the space of probability measure $(P_t)_{t \geq 0} = (\mathcal{L}(m_t))_{t \geq 0}$. Using Itô formula for stochastic flow of measures, [CD18a, Chap. 5] and [CD18b, Chap. 4], we obtain the following result

Proposition 1.8. *For any function $\Phi \in \mathcal{C}_b^2(\mathcal{P}_2(\mathbb{R}^d))$, for any $t \geq 0$,*

$$\langle P_t - P_0; \Phi \rangle = \int_0^t \langle P_s, \mathcal{M}\Phi \rangle ds, \quad (1.4)$$

where for any $m \in \mathcal{P}_2(\mathbb{R}^d)$,

$$\begin{aligned} \mathcal{M}\Phi(m) := & \int_{\mathbb{R}^d} \left[D_m \Phi(m, x) \cdot (G(x) + F(x - \mu_1(m))) + \frac{\sigma^2 + \sigma_0^2}{2} D_{xm}^2 \Phi(m, x) \right] m(dx) \\ & + \frac{\sigma_0^2}{2} \int_{\mathbb{R}^{2d}} \text{Tr} [D_{mm}^2 \Phi(m, x, y)] m(dx) m(dy). \end{aligned} \quad (1.5)$$

Proposition 1.8 gives a way to understand the flow of measures generated by the stochastic process $(m_t)_{t \geq 0}$. However, the equation for $(P_t)_{t \geq 0}$ is posed on $\mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ and then difficult to study because we lack some classical analysis tools. However, as previously mentioned, Equation (3.7) (and consequently Equation (1.3)) are part of a complete ecosystem of equations. This means that we can leverage probabilistic techniques to study the behaviour of the flows of measures $(m_t)_{t \geq 0}$ and $(P_t)_{t \geq 0}$. The next section is a step in this direction.

2.1 Between macroscopic and microscopic scales: interacting particle systems, the origin of the model

The solution to Equation (1.3) can be approximated in the mean-field limit by an interacting particle system as the number N of particles goes to infinity. It is important to clarify upfront that, in this manuscript, the term "particle" is used in a broad sense to mean an individual. Interacting individual systems of considerable scale is of paramount importance in a multitude of scientific fields, with the individuals ranging from biology (collective dynamics of animals), physics (elementary particles, ions, atoms, molecules, celestial bodies, galaxies), social science (crowd movements, interactions of financial agents, propagation of ideas) or financial markets (systemic risks). Let us do a quick historical recap on the introduction of interacting particle systems and mean field limits.

Let us consider a state space E , and a system of particles $\mathcal{Z}^N = (\mathcal{Z}^{1,N}, \dots, \mathcal{Z}^{N,N})$ where each particle is completely described by its state variable. The state variable can be anything, typically the position or speed of a particle, or the value of an asset (...). The concept of describing such systems can be traced back to the early works of Statistical Physics and Kinetic Theory of dilute gases [Bol70]. As in the Boltzmann example, when modelling particles in a gas the number of particles in $\mathcal{Z}^N = (\mathcal{Z}^{1,N}, \dots, \mathcal{Z}^{N,N})$ must be very large. It is known that there are approximately $N_A \approx 6.02 \cdot 10^{23}$

particles per mole [DHB⁺74]. Considering systems with such many particles imposes the following question:

Question 1.9. *Is there a limit, in some sense, when $N \rightarrow +\infty$?*

Answering Question 1.9 is central in several domains [LL63, Spo80, GKM⁺96, LL06a, LL06b, HJ15]. However, the answer really is a matter of scale. In order to describe such a system, three scales are possible (generating three different types of limit): **(1) The *microscopic scale*.** One possibility is to stay at the particle level, this is for example interesting when working with interacting particles systems on a graph or network. We follow the trajectory of each particle $\mathcal{Z}^{i,N}$ over time. **(2) The *macroscopic scale*.** We only have access to measurable quantities of the system, for example, the mean of the particles at each time. **(3) The *mesoscopic scale*.** In between the two previous scales is the mesoscopic or statistical scale. This corresponds to the scale where we are only interested in a statistical description of the system studying the density of particles in a region of the space.

The last point of view will be the only one studied in this manuscript. The concept of describing such systems in this manner is due to the so-called hypothesis of *molecular chaos*, presented by Boltzmann in [Bol70], giving birth to the equation bearing his name. This hypothesis posits that the particles in the system are uncorrelated. The molecular chaos introduced by Boltzmann mainly state that there exists a probability measure μ sur that for any $i \neq j$, $\mathcal{L}(\mathcal{Z}^{i,N}, \mathcal{Z}^{j,N}) = \mu \otimes \mu$. The rigorous formulation and proof of these is still an important subject of research in many cases. The ideas of Boltzmann quickly have been adopted for practical models: Currie-Weiss model for ferromagnetism [Wei07], Stellar-dynamic systems [Jea15, Jea19], Vlasov models [Vla45]. The underlying idea is that in the system, all the particles have the same role:

Definition 1.10. *We say that a system $\mathcal{Z}^N = (Z^{1,N}, \dots, Z^{N,N})$ is exchangeable (or that the N particles are indistinguishable), if for any permutation $\tau : \{1, \dots, N\} \rightarrow \{1, \dots, N\}$,*

$$\mathcal{L}(Z^{1,N}, \dots, Z^{N,N}) = \mathcal{L}(Z^{\tau(1),N}, \dots, Z^{\tau(N),N}).$$

In 1956, Kac published his seminal paper [Kac56] in which he derives the space homogeneous Boltzmann equation as the limit as the number of particles N goes to infinity of a system of N indistinguishable particles. The underlying idea is that when the number of particles N goes to infinity, two particles become statistically independent. This is the birth of Propagation of chaos:

Definition 1.11 (Kac's chaos, Propagation of chaos). *For a sequence*

$$\mathcal{Z}^N = (\mathcal{Z}_1^N, \dots, \mathcal{Z}_N^N)$$

of N -indistinguishable (in the sense of Definition 1.10) particle systems with $N \rightarrow \infty$, the stochastic chaos according to Kac means that there exists a probability measure μ such that

$$\mathcal{L}(\mathcal{Z}_i^N, \mathcal{Z}_j^N) \rightarrow \mu \otimes \mu \quad \text{as } N \rightarrow \infty, \quad \forall i \neq j \text{ fixed.}$$

Finally, McKean developed a stochastic version of the Vlasov equation, creating the McKean-Vlasov model, and proved propagation of chaos properties [McK66] linked to the latter. This is the origin of

the model of this manuscript.

Remark 1.12. *It is important to note that the definition of chaos propagation relies on a crucial assumption: indistinguishability. This assumption is essential because chaos propagation fundamentally implies that finally, we end up with a system composed of various versions of a typical particle.*

2.1.1 Introduction of the interacting particle system with common noise

Let us now introduce carefully the underlying interacting particle system of this manuscript. We consider another complete probability spaces $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1)$ endowed with a right-continuous and complete filtration $\mathbb{F}^1 = (\mathcal{F}_t^1)_{t \geq 0}$. We shall also assume that $(\Omega^0, \mathbb{F}^0, \mathbb{P}^0)$ is rich enough to support a Brownian motion B^0 playing the role of the common noise. Similarly, $(\Omega^1, \mathbb{F}^1, \mathbb{P}^1)$ is assumed to support a sequence of independent Brownian motions $(B^i)_{i \geq 1}$. By convention, in the whole manuscript, the index 0 always refers to the common noise and the index 1 to the idiosyncratic ones. We then define the product structure

$$\Omega = \Omega^0 \times \Omega^1, \quad \mathcal{F}, \quad \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, \quad \mathbb{P},$$

where $(\mathcal{F}, \mathbb{P})$ is the completion of $(\mathcal{F}^0 \otimes \mathcal{F}^1, \mathbb{P}^0 \otimes \mathbb{P}^1)$ and \mathbb{F} is the complete and right continuous augmentation of $(\mathcal{F}_t^0 \otimes \mathcal{F}_t^1)_{t \geq 0}$. Generic elements of Ω are denoted $\omega = (\omega^0, \omega^1)$ with $\omega^0 \in \Omega^0$ and $\omega^1 \in \Omega^1$. Given such a setup, let us introduce the following definition:

Definition 1.13. *Let us consider a random variable X defined on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, and $P \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$. According to Lemma 2.4 in [CD18b], for $\mathbb{P}^0 - a.e.$ $\omega_0 \in \Omega^0$, $X(\omega_0, \cdot)$ is a random variable on $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1)$. Then, $\mathcal{L}_1(X) : \Omega^0 \ni \omega_0 \mapsto \mathcal{L}(X(\omega_0, \cdot))$, we get a random variable from $(\Omega^0, \mathcal{F}^0, \mathbb{P}^0)$ into $\mathcal{P}(\mathbb{R}^d)$, providing a conditional law of X given \mathcal{F}^0 . Finally, we say that $\mathcal{L}(X) = P$ whenever $\mathcal{L}_1(X)$ is distributed with respect to P .*

The previous definition is a slight abuse of notation that allows us to consider *random initial conditions*. In the whole manuscript we denote by $\mathbb{E}_0, \mathbb{E}_1$ and \mathbb{E} the expectations on $(\Omega_0, \mathcal{F}^0, \mathbb{F}^0, \mathbb{P}_0)$, $(\Omega_1, \mathcal{F}^1, \mathbb{F}^1, \mathbb{P}_1)$ and $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ respectively. Analogously, for any random variable X defined on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, we denote by $\mathcal{L}_0(X)$ and $\mathcal{L}_1(X)$ the conditional laws of the random variable X given \mathcal{F}^1 and \mathcal{F}^0 respectively. Of course, $\mathcal{L}(X)$ stands for the law of X on the whole product space.

We now consider the interacting particle system $\mathbb{X}^N = (X^1, \dots, X^N)$ on the space $(\Omega, \mathcal{F}, \mathbb{P})$, with dynamic

$$\begin{cases} dX_t^i = b(X_t^i, m_t^N) dt + \sigma dB_t^i + \sigma_0 dB_t^0, \\ \mathcal{L}(\mathbb{X}_0^N) = P_0^{\otimes N}, \end{cases} \quad (1.6)$$

for some initial condition $P_0 \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$, and where

$$m_t^N := m^N(\mathbb{X}_t^N) : \mathcal{C}_d^N \ni \omega = (\omega^1, \dots, \omega^N) \mapsto N^{-1} \sum_{i=1}^N \delta_{\omega_t^i}$$

is the empirical measure of the particles. In the following of this section, let us consider the following set of assumptions.

Assumptions - Conditional propagation of chaos (CPoC)

1. The initial conditions $(X_0^i)_i$ are *i.i.d* and $\mathbb{E}[|X_0^i|^2] < +\infty$;
2. The map $b : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ is globally Lipschitz continuous, in the sense that there exists $L > 0$ such that

$$|b(x, \mu) - b(y, \nu)| \leq L \left(|x - y| + d_2^{\mathbb{R}^d}(\mu, \nu) \right),$$

for all $x, y \in \mathbb{R}^d$, and $(\mu, \nu) \in \mathcal{P}_2(\mathbb{R}^d)$.

Lemma 1.14 (Thm. 2.12 [CD18b]). *Under the set of Assumptions (CPoC), the interacting particle system (1.6) admits a unique solution for any $N \geq 1$.*

The unfamiliar reader may wonder at this point: *why do we consider the mesoscopic scale?* One could, in fact, consider the *microscopic* point of view and proceed as in Equation (1.12) to prove that the law of \mathbb{X}^N , denoted $(m_t^{N,N})_t$, is the solution of

$$\begin{cases} \partial_t m_t^{N,N} = \frac{\sigma^2 + \sigma_0^2}{2} \Delta m_t^{N,N} - \nabla \cdot (b^N m_t^{N,N}), \\ m_0^{N,N} = m_0^{\otimes N}, \end{cases}$$

where $b^N : \mathbb{R}^{Nd} \ni x = (x^1, \dots, x^N) \mapsto (b(x^i, m^N(x^1, \dots, x^N)))_{i=1}^N \in \mathbb{R}^N$. The flow of measures $m^{N,N}$ is the solution of a linear equation, but in high dimension. Such an equation (called the *Liouville equation* [Dar70, Ves42]) becomes increasingly difficult to manage as the number N grows. Given that this number is likely to be very large, such a description of the system proves challenging to use, thus justifying that the microscopic scale is not the most suitable for studying these types of interacting systems. One main objective here is to consider an object that "*remains of order one*" in N .

Models similar to (1.6) are very well known and studied in the case without common noise, see for example [Lac18] for a very nice course on the subject. Before going into more detail on the question of asymptotic behaviour of (1.6) in N , let us present a few examples that motivate the technical efforts made here to integrate common noise into the model.

2.1.2 On the importance of common noise

The system (1.6) has been extensively studied in the literature [Mél96, Szn91, CD18a, CD18b, CGM08, Mal01, JT21, CDL16]. In this section, we give some examples of systems that are similar to (1.6) in which the necessity of modelling interacting particle system is clear.

- Inter-bank transactions

We consider a simple model of inter-bank borrowing and lending where the evolution of the log-monetary reserves of N banks is described by a system of diffusion processes. The model is:

$$\begin{cases} dX_t^i = \frac{\alpha}{N} \sum_{j=1}^N (X_t^j - X_t^i) dt + \sigma dW_t^i + \sigma_0 dW_t^0, \\ \mathcal{L}(X_0^1, \dots, X_0^N) = \delta_0^{\otimes N}. \end{cases} \quad (1.7)$$

This model, proposed in [CFS15], highlights the importance of common noise in such systems. Here, each particle X^i represents the log-monetary reserve of bank i . We study the *probability of default* for the system, introducing a default level $D < 0$ where bank i defaults by time T if its log-monetary reserve hits D before T . The figures below show simulations of the log-monetary level for $N = 100$ banks, with one system including common noise and the other not. The dotted line indicates the default level.

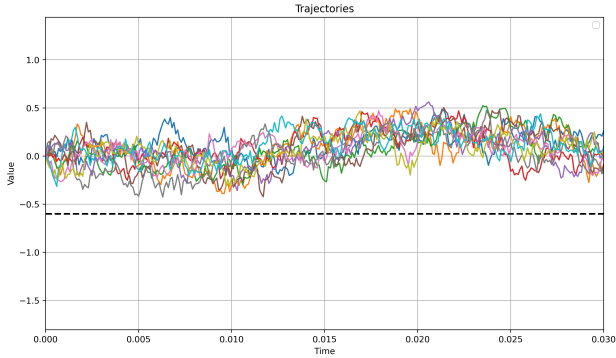


Figure 1.1: Without common noise, $\sigma_0 = 0$, $N = 100$.

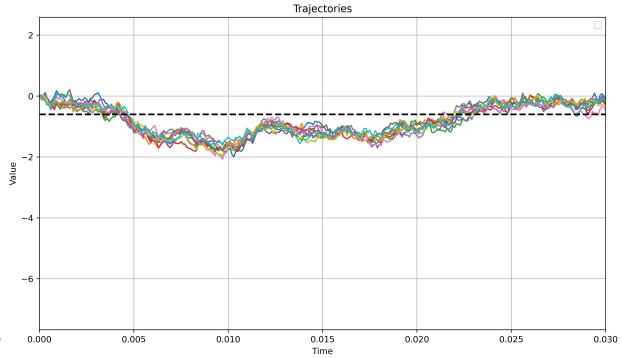


Figure 1.2: With common noise, $\sigma_0 > 0$, $N = 100$.

We can observe (see [CFS15] for a formal proof) that adding correlation between the Brownian makes the whole system more sensitive and allows for a non-vanishing probability of default.

- Mean-Field Game

Mean field game (MFG) theory is a mathematical framework for analyzing complex systems of many interacting agents. Initially proposed by Lasry and Lions [LL06a, LL06b, LL07] and by Huang, Malhamé, and Caines [HCM07, HMC06], MFG theory examines the actions of a representative agent assuming all other agents behave similarly, see [Car10] for a nice course on the subject. This leads to a set of partial differential equations describing the system's temporal development. Since its development, MFG theory has grown rapidly, with applications in economics, finance, engineering, and social sciences, including traffic flow, energy management, and crowd dynamics. Many MFG models now include common noise, particularly in economic models with aggregate shocks [BB92]. For a specific example, see [Tch18a].

- Climat risk integration

The impact of climate change on financial markets is a growing concern for insurers, financial institutions, and investors. They recognize the need to assess and manage this risk to ensure financial stability and sustainable practices. According to [Fra24], insurers could integrate climate risk into financial sector models by introducing a common noise, called carbon risk, affecting sectors like chemicals, electricity, gas, and real estate. The Black & Scholes model describes sector dynamics as:

$$dS_t^i = \sigma_t^i S_t^i dB_t^i + \sigma_0 S_t^i dZ_t, \tag{1.8}$$

where Z is the carbon risk martingale, and B^i are Brownian motions. Recently [Yag20] report that climate-related natural disasters and economic losses almost doubled from 1980-1999 to 2000-2019. These risks affect company valuations and share prices, and stringent policies under the Paris Agreements are expected to achieve carbon neutrality by 2050. This necessitates incorporating common noise into models of large systems.

2.1.3 Conditional propagation of chaos

As mentioned in the beginning of Section 2.1, it is important to understand the asymptotic behaviour (in N) of the particle system (1.6). Let us first note that for our system, the following Lemma stands

Lemma 1.15. *Let assumptions (CPoC) be in force. Then, the interacting particle system (1.6) is exchangeable in the sense of Definition 1.10.*

The previous lemma is the reason why we are going to study the system at a mesoscopic scale. To that extent, let us study the behaviour of the empirical measure $(m_t^N)_t$ of the system. Indeed, thanks to Lemma 1.15, we know that $(m_t^N)_t$ contains all the necessary information to establish the asymptotic behaviour of the system when the number of particles tends to infinity, see [Szn91] for more details.

As we are in the framework of an exchangeable model, we can use De Finetti's results (see e.g [Kin78]):

Theorem 1.16 (Conditionnal law of large numbers). *For any exchangeable sequence of random variables $(X_i)_{i \geq 1}$ define on $(\Omega, \mathcal{F}, \mathbb{P})$ with a finite moment of order one, we have,*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N X_i = \mathbb{E}[X_1 | \mathcal{F}_\infty],$$

where

$$\mathcal{F}_\infty := \bigcap_{n \geq 1} \sigma \{X_k, k \geq n\},$$

and the convergence holds in the almost sure sense.

Intuitively, this form of law of large numbers says that, when exchangeability holds, the empirical measure of the random variables $(X_i)_{i \geq 1}$ behaves asymptotically as if they were conditionally independent and identically distributed given the tail σ -algebra \mathcal{F}_∞ . In our setting, individuals are always linked through the common noise so there is no way to expect full independence in the limit. However, in the light of the previous paragraphs we expect conditional independence, given the common noise. That is why we need a conditional version of Kac's chaos. We can revisit the Definition 1.10

Definition 1.17 (Conditional Kac's chaos, Conditional Propagation of chaos). *For a sequence*

$$\mathcal{Z}^N = (\mathcal{Z}_1^N, \dots, \mathcal{Z}_N^N)$$

of N -indistinguishable particle systems with $N \rightarrow \infty$, the conditional stochastic chaos with respect to a σ -algebra \mathcal{G} , means that there exists a random probability measure μ , adapted to \mathcal{G} , such that

$$\mathcal{L}((\mathcal{Z}_i^N, \mathcal{Z}_j^N) | \mathcal{G}) \rightarrow \mu \otimes \mu \quad \text{as } N \rightarrow \infty, \quad \forall i \neq j \text{ fixed,}$$

where the previous limit holds in the almost sure sense.

Due to the particular form of the interaction and the presence of common noise, models as (1.6) exhibit conditional propagation of chaos properties under some reasonable set of conditions. These ideas are the cornerstone of the MFG theory. Mainly, the more general result on the subject is stated as follows:

Theorem 1.18 ([CD18b] Theorem 2.12). *Within the above framework and under the set of Assumptions (CPoC), the system of particles (1.6) has a unique solution for every $N \geq 1$. Moreover for any $T > 0$,*

$$\max_{1 \leq i \leq N} \mathbb{E} \left[\sup_{0 \leq t \leq T} |X_t^i - \bar{X}_t^i|^2 \right] + \sup_{0 \leq t \leq T} \mathbb{E} \left[d_2^{\mathbb{R}^d} \left(\frac{1}{N} \sum_{i=1}^N \delta_{X_t^{N_i}}, \mathcal{L}_1(\bar{X}_t^1) \right) \right] \rightarrow 0, \quad (1.9)$$

where $(\bar{X}^1, \dots, \bar{X}^N)$ is called the mean-field limit system of (1.6) and is given by

$$\begin{cases} d\bar{X}_t^i = b(X_t^i, \mathcal{L}_1(\bar{X}_t^i)) dt + \sigma dB_t^i + \sigma_0 dB_t^0, \\ \mathcal{L}(\mathbb{X}_0^N) = P_0^{\otimes N}. \end{cases} \quad (1.10)$$

Historically, one of the first tools used to prove propagation of chaos (in the case without common noise) is probabilistic as used by McKean in [McK66], and then popularised by Sznitman [Szn91], it is a coupling method. In the cases with common noise, the proofs can be done in the same manner using tools of DeFinetti's theory as Theorem 1.16, see [CDL16, CD18b].

Of course, the rate of convergence to 0 in Theorem 1.18 has been the subject of many studies. This is not a central question in our framework and we refer the interested reader to [CD22a, CD22b]. A question of huge interest however in our context is the following

Question 1.19. *How does (conditional) propagation of chaos behave in time? Can we prove uniform in time (conditional) propagation of chaos for the system (1.6)?*

Uniform in-time (Conditional) Propagation of Chaos is shown easily with coupling methods whenever strict space monotony assumptions are in force for the drift b (along with (CPoC)), see for example [Mal01] in the case without common noise or [Mai23] (Chapter 3 of this manuscript) in the case with common noise. When we lose the strict monotony assumptions, things become more complicated. It would be interesting to obtain uniform in-time propagation of chaos particularly because it would enable us to understand in a better way the long-time behaviour of the limit process.

In [DEGZ20] the authors demonstrate uniform in-time propagation of chaos (without common noise), without relying on strict monotonicity assumptions for b , using specific coupling methods that will be discussed in detail later in this manuscript. Additionally, they establish exponentially fast convergence to the stationary regime for the solutions of Equation (1.16). These methods can be extended to scenarios involving common noise, provided the intensity of the idiosyncratic noise $\sigma > 0$ is sufficiently large. More details on this phenomenon are given in Section ... If this condition is not met, achieving uniformity in time becomes challenging, opening a significant area for further research.

2.2 McKean-Vlasov/Fokker-Planck ecosystem

Let us consider a measure $P_0 \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$, a random variable m_0 constructed on the space $(\Omega_0, \mathcal{F}^0, \mathbb{F}^0, \mathbb{P}_0)$ such that $\mathcal{L}_0(m_0) = P_0$, and a random variable X_0 constructed on the product space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ such that $\mathcal{L}_1(X_0) = m_0$ almost surely. Now, we can construct the process X with dynamics given by the stochastic differential equation

$$\begin{cases} dX_t = b(X_t, \mathcal{L}_1(X_t))dt + \sigma dB_t + \sigma_0 dB_t, \\ X_{t|t=0} = X_0 \quad a.s \end{cases} \quad (1.11)$$

for some Brownian motion B adapted to \mathbb{F}^1 , and thus independent of B^0 . The particularity of the SDE (1.11) is that the process X is interacting with its own conditional law given the *common noise* B^0 . Such type of interaction is called *McKean* non-linearity.

The connection between the process X and Equation (1.3) is quite straightforward and was already mentioned in the paper of McKean [McK66] (in the case $\sigma_0 = 0$). Considering a function $\phi \in \mathcal{C}_c^\infty(\mathbb{R}^d)$, we write for any $0 \leq s \leq t$,

$$\begin{aligned} \phi(X_t) &= \phi(X_s) + \int_s^t \nabla \phi(X_r) b(X_r, m_r) dr + \sigma \int_s^t \nabla \phi(X_r) dB_r \\ &+ \sigma_0 \int_s^t \nabla \phi(X_r) dB_r^0 + \frac{\sigma^2 + \sigma_0^2}{2} \int_s^t \Delta \phi(X_r) dr. \end{aligned} \quad (1.12)$$

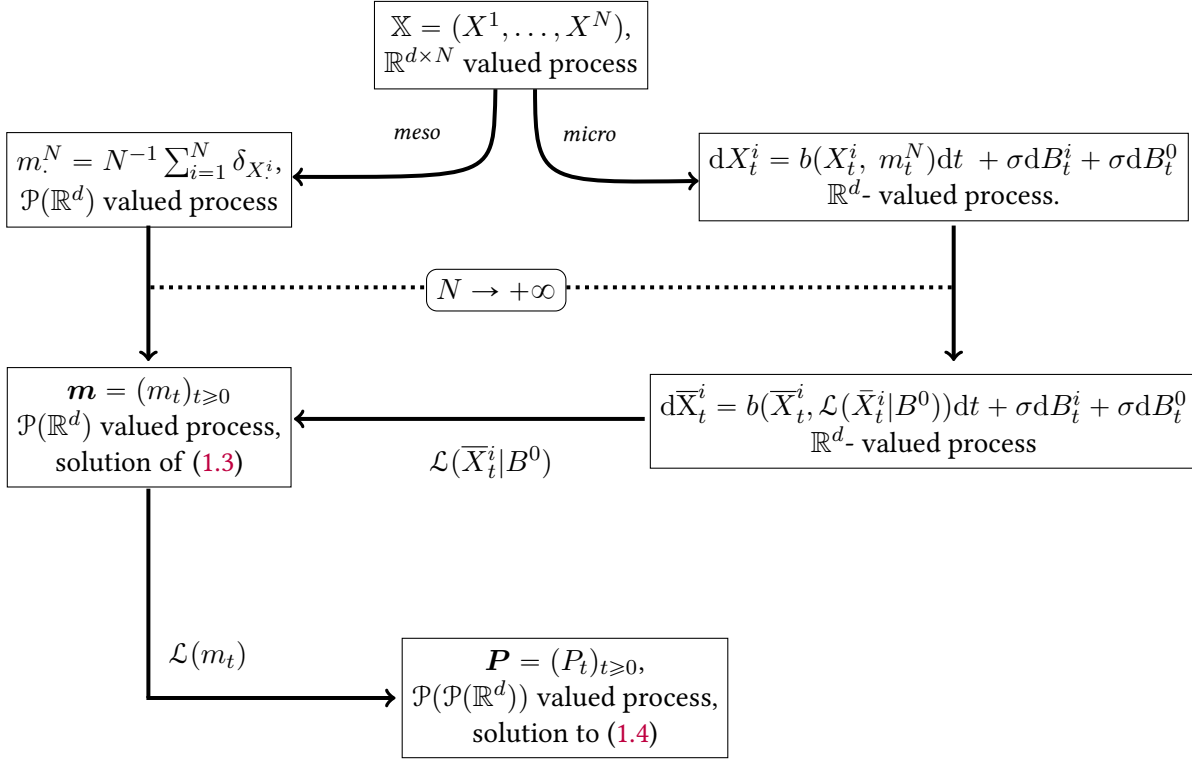
Taking the conditional expectation with respect to the common noise B^0 , we obtain for all $0 \leq s \leq t$,

$$\begin{aligned} \langle m_t; \phi \rangle &= \langle m_s; \phi \rangle + \frac{\sigma^2 + \sigma_0^2}{2} \int_s^t \langle m_r; \Delta \phi \rangle dr \\ &+ \int_s^t \langle m_r; \phi b(\cdot, m_r) \rangle dr + \sigma_0 \int_s^t \langle m_r; (\nabla \phi)^T \rangle dB_r^0. \end{aligned}$$

This being true for any $0 \leq s \leq t$ and $\phi \in \mathcal{C}_c^\infty(\mathbb{R}^d)$, we deduce that $(m_t)_{t \geq 0}$ satisfies Equation (1.3). Under various assumptions, existence and uniqueness results are known for Equation (1.11) (see [KX99, HvS21, CD18b]) and for the SPDE (1.3) (see [DV95, KX99, CG19, CDLL19]). For example, the following result stands

Proposition 1.20 (Prop. 2.8, [CD18b]). *Let the Assumption on the drift b in (CPoC) be in force. Assuming furthermore that $\mathbb{E}[|X_0|^2] < +\infty$, there exists a unique solution to (1.11) on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$.*

It is quite straightforward to observe that the existence of a solution to (1.11) implies the existence of a solution to (1.3) with the initial condition $\mathcal{L}_0(m_0) = P_0$. This result further implies the existence of a solution for the Fokker-Planck equation on $\mathcal{P}(\mathbb{R}^d)$ as given by (3.7). In [LSZ23], the authors even demonstrate that the converse implications are true, which allows for the verification of uniqueness results directly. These findings also lay the groundwork for characterising an invariant measure in Chapters 3 and 4. Ultimately, it becomes clear that our equation of interest is part of a broader ecosystem of equations that originates from the system of interacting particles described in (1.6).


 Figure 1.3: McKean-Vlasov/Fokker-Planck ecosystem of equations with *common noise*.

Let us now provide a concrete and simple example of McKean-Vlasov/Fokker-Planck ecosystem.

Example 1.21. *Let us now consider a particular case to make the things clear:*

$$\begin{cases} dX_t = -(X_t - \mathbb{E}_1[X_t])dt + \sigma dB_t + \sigma_0 dB_t^0 \\ X_{t|t=0} = 0 \quad a.s. \end{cases}$$

This process is uniquely determined and can be approximated in the mean-field limit by an interacting particle with

$$b(x, m) = -\left(x - \int_{\mathbb{R}^d} xm(dx)\right),$$

which satisfies (CPoC). This process can be treated exactly as an Orstein-Uhlenbeck process in \mathbb{R}^d . Then, we get an explicit solution

$$X_t = \sigma_0 B_t^0 + \sigma \int_0^t e^{-(t-s)} dB_s. \quad (1.13)$$

Then we obtain that for any $t \geq 0$,

$$m_t(x) = \frac{1}{\sigma \sqrt{\pi(1-e^{-t})}} \exp\left(-\frac{1}{\sigma^2(1-e^{-t})} |x - \sigma_0 B_t^0|^2\right). \quad (1.14)$$

In this example, we see explicitly the fact that the flow of measures $(m_t)_{t \geq 0}$ is random and adapted to the common noise. In this example, we can also explicitly write the deterministic flow of measures associated

to $(m_t)_{t \geq 0}$, for all $t > 0$,

$$P_t(dm) = \int_{\mathbb{R}^d} \delta_{\mathcal{N}(\sigma_0 \sqrt{t}x, \sigma^2 \pi(1-e^{-t}))}(dm) \gamma(dx),$$

where γ stands for the standard Gaussian density.

2.3 Long time behaviour: reminders on the case without common noise

Let us comment on the previously introduced system in the case $\sigma_0 = 0$. In this case, the particle dynamics are only dictated by independent Brownian motions, then the system reads

$$\begin{cases} dX_t^i = b(X_t^i, m_t^N) dt + \sigma dB_t^i, \\ \mathcal{L}(\mathbb{X}_0^N) = m_0^{\otimes N}, \end{cases} \quad (1.15)$$

for some probability measure $m_0 \in \mathcal{P}(\mathbb{R}^d)$. The particles would be i.i.d if not for the interaction coming through the empirical measure. In this case, under the same conditions on b (see (CPoC)), all the propagation of chaos results are true, see [Szn91, Mél96, CD18a, CD22a, CD22b]. Taking $\sigma_0 = 0$ in Equation (1.3), we obtain the following Fokker-Planck equation

$$\partial_t m_t = \frac{\sigma^2}{2} \Delta m_t - \nabla \cdot (m_t b(\cdot, m_t)). \quad (1.16)$$

This changes the nature of the objects considered, but all the probabilistic tools are still valid in both cases. The formalism used to describe these equations with common noise can be used to deal with problems without common noise. Indeed, Figure 1.3 remains true and it is then sufficient to note that $(m_t)_{t \geq 0}$ is a deterministic flow of measure. The flow of measure $(P_t)_{t \geq 0}$ then becomes trivial:

$$\forall t \geq 0, \quad P_t = \delta_{m_t}. \quad (1.17)$$

where for any measure $m \in \mathcal{P}(\mathbb{R}^d)$, δ_m stands for the Dirac measure centered at m . Moreover, the McKean-Vlasov process X , with dynamic

$$dX_t = b(X_t, \mathcal{L}(X_t))dt + \sigma dB_t \quad (1.18)$$

generates the flow of measures $(m_t)_{t \geq 0}$ solution of (1.16). Let us begin by presenting some classical existence and uniqueness results for the solutions of these equations. The aim of the following section is not to provide an exhaustive state-of-the-art, but rather to present a reading grid that will help to understand the fundamental results concerning notably the long-time behaviour of solutions of (1.3) in the rest of the manuscript. Naturally, as in the case with common noise, Equation (1.16) is understood in the weak sense. We say that (m_t) is a solution of (1.16) if for any smooth and bounded function ϕ , we have

$$\langle \phi; m_t \rangle = \langle \phi; m_s \rangle + \int_s^t \langle \phi b(\cdot, m_r); m_r \rangle dr + \frac{\sigma^2}{2} \int_s^t \langle \Delta \phi; m_r \rangle dr.$$

2.3.1 Long time behaviour

The question of the long-time behaviour of the solutions of (1.16) has been an important subject of interest for a long time. When discussing the long-time behaviour of solutions of (1.16), it is usual [CGM08, DEGZ20, BRV98, HT10, HIP08] to assume the following structure for the drift b :

$$b(x, \mu) = -\nabla V(x) - \nabla W * \mu(x). \quad (1.19)$$

The function V is then called the confinement potential and defines the landscape in which the associated process X evolves. The function W is called the interaction potential and defines how the process X interacts with its law. This is the source of nonlinearity in equation (1.16).

It is well known in the case where $W = 0$, that there exists a unique stationary solution to the equation (1.16), provided that V has good asymptotic properties (such as, coercivity) and $\sigma > 0$. Moreover, this unique invariant measure admits a density with respect to the Lebesgue measure, this density denoted \bar{m} is written as

$$\bar{m}(x) \propto e^{-\frac{2}{\sigma^2}V(x)}. \quad (1.20)$$

When $W \neq 0$, the equation of interest becomes non-linear, and the situation becomes more intricate. Similar results on the existence and uniqueness of the stationary solution have been shown under strict convexity assumptions on V , both with PDE [CMV03, CMV06] and probabilistic [CGM08, Mal01] techniques.

However, this is not the case anymore if we relax the convexity assumption as highlighted by [DV95] or [HT10]. In the latter, the authors show that any stationary solution of (1.16) is a solution to the following fixed-point problem:

$$\bar{m}(x) = \frac{1}{Z_{\bar{m}}} \exp \left(-\frac{2}{\sigma^2} (V(x) + W * \bar{m}(x) - W * \bar{m}(0)) \right), \quad (1.21)$$

for some normalising constant $Z_{\bar{m}}$. Let us consider an example to clarify the previous equation. Throughout this manuscript, a central case of study is given by: $V(x) = x^4/4 - x^2/2$ and $W(x) = \alpha x^2/2$, for $\alpha > 0$. This example is one-dimensional. In this case, the equation (1.21) satisfied by the stationary solutions rewrites:

$$\bar{m}(x) = \frac{1}{Z_{\bar{m}}} \exp \left(-\frac{2}{\sigma^2} (x^4/4 - x^2/2 + \alpha x^2/2 - \alpha x \mu_1(\bar{m})) \right), \quad (1.22)$$

where for any probability measure $m \in \mathcal{P}_1(\mathbb{R}^d)$, $\mu_1(m) = \int_{\mathbb{R}^d} xm(dx)$. In particular, it follows that \bar{m} is a solution of

$$\mu_1(m) = \Psi(\mu_1(m)),$$

where, for all $y \in \mathbb{R}$,

$$\Psi(y) = \frac{\int_{\mathbb{R}} x \exp \left(-\frac{2}{\sigma^2} (x^4/4 - x^2/2 + \alpha x^2/2 - \alpha xy) \right) dx}{\int_{\mathbb{R}} \exp \left(-\frac{2}{\sigma^2} (x^4/4 - x^2/2 + \alpha x^2/2 - \alpha xy) \right) dx}.$$

Let us first notice that $\mu_1(m) = 0$ is always a solution. Moreover, it has been shown in [Daw83]

that there exists a critical value $\sigma_c > 0$ such that, if $\sigma \geq \sigma_c$, $\mu_1(m) = 0$ is the only solution. However, when $\sigma < \sigma_c$, there exist two other solutions. In other words, when σ is too small, there are three stationary solutions to Equation (1.16).

A good way to understand this phenomenon, at least intuitively, is to go back to the study of a system with many particles:

$$dX_t^i = -\nabla V(X_t^i)dt - \alpha \left(X_t^i - \frac{1}{N} \sum_{j=1}^N X_t^j \right) dt + \sigma dB_t^i. \quad (1.23)$$

At this stage, it is natural that most particles are attracted by the minimizers of the potential $V : x \mapsto x^4/4 - x^2/2$. When σ is large enough, it dominates the interaction term, allowing particles to move from one minimizer to another, resulting in a distribution of particles between the two minimizers. On the other hand, when σ is too small, the interaction term dominates, and the particles will all gather near a minimum.

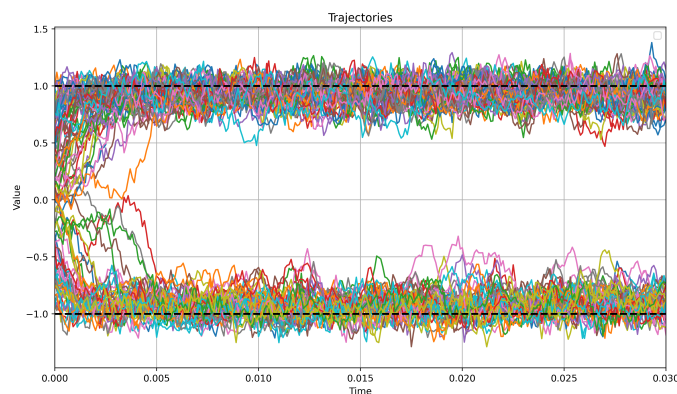


Figure 1.4: $N = 100$, $\sigma/\alpha = 600$, $V : x \mapsto x^4/4 - x^2/2$.

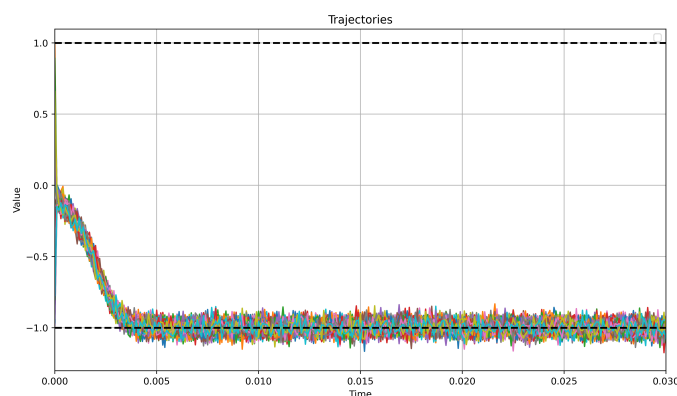


Figure 1.5: $N = 100$, $\sigma/\alpha = 6$, $V : x \mapsto x^4/4 - x^2/2$.

In the N -particle system for σ small, it is expected that particles will move around one of the points. If one of the N noises takes an extreme value and pushes the corresponding particle to the other stable state, then the dominant attractive force in the drift should quickly pull this particle back to the average particle location.

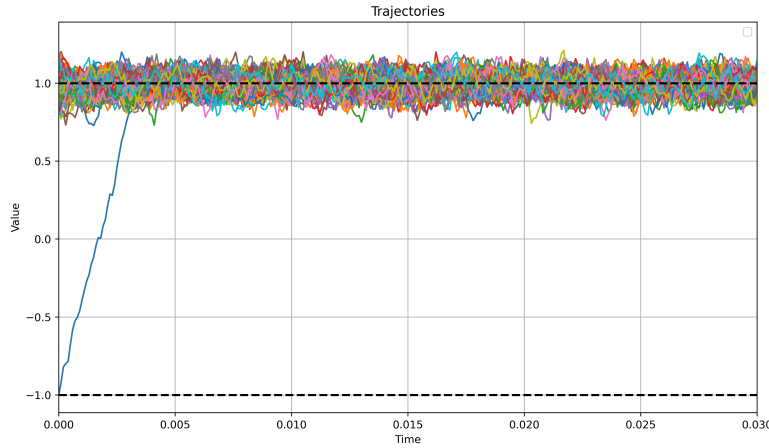


Figure 1.6: Particle attracted by the mean behaviour of the system - $N = 100$, $\sigma/\alpha = 6$,
 $V : x \mapsto x^4/4 - x^2/2$

Studying the probability that many particles, say $n \gg 1$, escape from the minimiser they are located in, within a time interval $\Delta > 0$ is of the order of

$$\mathbb{P}(\exists I = \{i_1, \dots, i_n\}, \quad \forall i \in I, B_\Delta^i > 1),$$

which is exponentially small in n . This roughly says that the probability that a large number of particles go from one minimiser to the other vanishes, creating multiplicity of the equilibria for $(m_t)_{t \geq 0}$. This result concerning the existence of several stationary solutions is extended to much more general contexts (however in dimension $d = 1$) in [HT10].

2.4 Contributions of the first part

The addition of common changes the asymptotic behaviour of the empirical measure of the system in N . As explained in Section 2.1, the limit when the number of particles $N \rightarrow +\infty$ is random. The long-time behaviour of this process has not been the subject of many studies. Indeed, apart from two papers [ABKO23, ACKO24] in the case where the common noise added to the system is a cylindrical Brownian motion, little is known. In the first part of this manuscript, we focus exclusively on studying the long-time behaviour of the solutions of this equation. We will address the following questions:

Question 1.22. *What is an invariant measure for the process $(m_t)_{t \geq 0}$? How can it be characterised?*

Question 1.23. *In the frameworks where uniqueness of the invariant measure holds without common noise, do we preserve uniqueness for the case without common noise, $\sigma_0 = 0$?*

Naturally, since the process $(m_t)_{t \geq 0}$ is a stochastic process taking values in the space of measures, we seek to understand its asymptotic behaviour in law. In the following, we then study the existence and uniqueness of invariant measures $\bar{P} \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$. Let us state the following definition:

Definition 1.24. *We say that $\bar{P} \in \mathcal{P}_2(\mathcal{P}_1(\mathbb{R}^d))$ is an invariant measure for the process $(m_t)_{t \geq 0}$ when the latter is regarded as taking values in $\mathcal{P}(\mathbb{R}^d)$, if the law of m_t is independent of t , when m_0 is distributed*

according to \bar{P} , i.e., for all continuous and bounded function $\phi \in \mathcal{C}_b(\mathcal{P}(\mathbb{R}^d))$

$$\mathbb{E}_0[\phi(m_t)] = \mathbb{E}_0[\phi(m_0)] = \int_{\mathcal{P}(\mathbb{R}^d)} \phi(m) \bar{P}(dm), \quad \forall t > 0.$$

This answers the first part of Question 1.22. In chapter 2, we show the following characterisation of an invariant measure, which is an adaptation of the previously mentioned result of Proposition 1.8:

Proposition 1.25 ([LSZ20, Mai23]). \bar{P} is an invariant measure for the process (m_t) if and only if

$$\langle \bar{P}, \mathcal{M}F \rangle = 0, \quad \forall F \in \mathcal{C}_b^2(\mathcal{P}_2(\mathbb{R}^d)), \quad (1.24)$$

where the generator \mathcal{M} is defined in Equation (1.5).

Let us now consider a concrete example of what constitutes an invariant measure in our context. To illustrate this, we will examine a linear process, making all the computations explicit.

Example 1.26. In this example, we consider an Ornstein-Uhlenbeck process with common noise, that is taking $V : x \mapsto |x|^2/2$ and $W = 0$. Let X be a real-valued stochastic process evolving in \mathbb{R}^d , driven by the following stochastic differential equation:

$$dX_t = -X_t dt + \sigma dB_t + \sigma_0 dB_t^0. \quad (1.25)$$

Then, the process associated process $(m_t)_{t \geq 0}$ is solution of

$$d_t m_t = \nabla \cdot \left[\frac{\sigma^2 + \sigma_0^2}{2} \nabla m_t + m_t x \right] dt - \sigma_0 \nabla m_t \cdot dB_t^0, \quad (1.26)$$

Moreover, we consider the following initial condition $\mathcal{L}(X_0) = P_0$ (in the sense of Definition 1.24), for some $P_0 \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$. In this particular setting, we are able to explicitly describe the invariant measure. The unique invariant measure \bar{P} of the process on the space $\mathcal{P}(\mathbb{R}^d)$ is the image, by the function $x \in \mathbb{R}^d \mapsto \mathcal{N}_d(-x, \sigma^2 Id) \in \mathcal{P}(\mathbb{R}^d)$ of the measure γ_{σ_0} ; where $\gamma_{\sigma_0}(dx) = (2\sigma_0^2)^{-1/2} e^{-|x|^2/2\sigma_0^2} dx$, where for all $\mu \in \mathbb{R}^d, \Sigma \in \mathcal{M}_d(\mathbb{R}), \mathcal{N}_d(\mu, \Sigma)$ denotes a gaussian distribution in dimension d centered in μ and with variance-covariance matrix Σ . We can rewrite

$$\bar{P}(dm) = \int_{\mathbb{R}^d} \delta_{\mathcal{N}_d(-x, \sigma^2 Id)}(dm) \gamma_{\sigma_0}(dx). \quad (1.27)$$

Thanks to Proposition 1.25, we can characterise an invariant measure. However, we can already notice that the equation satisfied by the invariant measure is difficult to study since we need the classic tools of differential calculus in the space of measures. This difficulty is the reason for the need to develop probabilistic approaches that we will not detail for the moment. Let us start by stating some assumptions under which we will work throughout the first part of this manuscript.

Assumptions - Long Time Behaviour (LTBa)

1. There exists a continuous function $\kappa : [0, +\infty) \rightarrow \mathbb{R}$, such that

$$\liminf_{r \rightarrow +\infty} \kappa(r) > 0,$$

and

$$(\nabla V(x) - \nabla V(y)) \cdot (x - y) \geq \kappa(|x - y|)|x - y|^2;$$

2. ∇V is L_V -Lipschitz continuous.
3. W is symmetric, i.e., $W(x) = W(-x)$ for all $x \in \mathbb{R}^d$;
4. ∇W is L_W -Lipschitz continuous.

The first result of this thesis states that under the set of assumptions (LTBa), there always exists at least one invariant measure for the process $(m_t)_{t \geq 0}$ solution of (1.3). This provides a first element of the answer to the Question 1.23. To complete the answer to this question, it is interesting to consider the classical cases of uniqueness in the case $\sigma_0 = 0$. For example, when the potential V is uniformly convex, which corresponds to the case $\kappa \geq \beta > 0$ in (LTBa). In Chapter 2, we show that this uniqueness result is preserved in the case with common noise. We prove this result by developing probabilistic methods. More specifically, we prove uniform in time propagation of chaos results that allow us to conclude the desired result. Ultimately, we obtain the following result,

Theorem 1.27. *Let us consider $P, Q \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$, and two particle systems $\mathbf{X} = (X^1, \dots, X^N)$ and $\mathbf{X}^N = (X^{1,N}, \dots, X^{N,N})$ with dynamics,*

$$dX_t^i = -\nabla V(X_t^i)dt - \nabla W * m_t(X_t^i)dt + \sigma dB_t^i + \sigma_0 dB_t^0, \quad \forall i \in \{1, \dots, N\}, \quad (1.28)$$

and

$$dX_t^{i,N} = -\nabla V(X_t^{i,N})dt - \frac{1}{N} \sum_{j=1}^N \nabla W(X_t^{i,N} - X_t^{j,N})dt + \sigma dB_t^i + \sigma_0 dB_t^0, \quad \forall i \in \{1, \dots, N\}, \quad (1.29)$$

where $(X_0^i)_{i \in \{1, \dots, N\}}$ are independent and identically distributed such that $\mathcal{L}(\bar{X}_0^1) = P$ in the sense of Definition 1.13, and where the same holds for the second system with $\mathcal{L}(X_0^{1,N}) = Q$. Under Assumptions (LTBa), and assuming that $\kappa \geq \beta > 0$, we get

$$\mathbb{E} \left[d_2^{\mathbb{R}^d} (m_{\mathbf{X}_t}^{P,N}, m_{\mathbf{X}_t}^{Q,N}) \right] \leq C \left(e^{-\beta t} + \frac{1}{2\beta\sqrt{N-1}} \right),$$

where $m_{\mathbf{X}_t}^{P,N} = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^i}$ and $m_{\mathbf{X}_t}^{Q,N} = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i,N}}$.

The following corollary provides an answer to the Question 1.23:

Corollary 1.28. *Under the set of Assumptions (LTBa), the stochastic process (m_t) admits a unique invariant measure $\bar{P} \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$. Moreover, for each $P_0 \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$, there is an exponential convergence*

to the invariant measure in Wasserstein-2 distance

$$d_2^{\mathcal{P}(\mathbb{R}^d)}(P_t, \bar{P}) \leq e^{-\beta t} d_2^{\mathcal{P}(\mathbb{R}^d)}(P_0, \bar{P}).$$

This consists in a restoration of uniqueness result, at least in $\mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$. Moreover, the following Lemma allows us to recover uniqueness among all the invariant measures (as defined in Definition 1.24):

Proposition 1.29. *Let Assumptions (LTBa) be in force. For any invariant measure \bar{P} in the sense of Definition 1.24, we get that $\bar{P} \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$.*

In Section 2.1, we saw that there are examples of loss of uniqueness in the case without common noise when the diffusion coefficient σ is too small. As the addition of common noise completely changes the structure of the problem, it is natural to ask the following question:

Question 1.30. *Are there any cases where the common noise allows for uniqueness recovery?*

Question 1.30 is the focus of Chapters 2 and 3. It is quite common that the addition of noise allows for uniqueness recovery of the asymptotic behaviour for stochastic processes or the uniqueness of the Nash equilibria in MFG [Del19, DFT20]. However, two points should be emphasised here, justifying the difficulty of the question:

1. For processes with nonlinear dynamics, there are examples where the introduction of noise creates periodic asymptotic behaviours [Sch93, Sch85b, Sch85a];
2. In this specific case, the common noise has a very particular structure since this finite-dimensional noise is the only source of randomness in the dynamics of the infinite-dimensional process $(m_t)_{t \geq 0}$.

The second point highlights the subtlety of the problem. It is clear that there is no chance to get the uniqueness of the invariant measure for a generic class of stochastic Fokker-Planck SPDEs forced by a *simple* noise like B^0 . For example, it is quite easy to find similar models for which the common noise does not allow at all to recover the uniqueness of the asymptotic behaviour. The thrust of the works [Mai23, DMT24] is more to identify one class of mean field dynamics for which the common noise helps in this matter.

Example 1.31 (A simple case without restoration of uniqueness). *A counter-example in dimension $d = 1$ is*

$$d_t m_t = \frac{\sigma(v(m_t))^2 + \sigma_0^2}{2} \partial_{xx}^2 m_t dt + \partial_x(x m_t) dt - \sigma_0 \partial_x m_t dB_t^0,$$

for some function $\sigma : [0, +\infty) \rightarrow (0, +\infty)$. Whenever $\sigma_0 = 0$ and the function $\sigma^2(\cdot)/4$ admits several fixed points, the stationary solutions to the above equation cannot be unique. In this case, $(m_t)_{t \geq 0}$ is indeed solution of

$$\partial_t m_t = \frac{\sigma(v(m_t))^2}{2} \partial_{xx}^2 (m_t) + \partial_x(x m_t), \quad t \geq 0. \quad (1.30)$$

Then, similar to the stationary solutions found in the previous example, the stationary solutions (equivalently, the invariant measures of (3.5)) are here solutions of

$$\frac{dm}{dx}(x) = \exp\left(-2\frac{x^2}{\sigma^2(v(m))}\right), \quad x \in \mathbb{R}.$$

Then, one can show that, if the function $x \mapsto \sigma^2(x)/4$ admits several fixed points, then there are several stationary solutions to (3.20).

When adding common noise, the situation does not get better. For instance, assuming that $X_0 = 0$ a.s., one can show that for all $t > 0$, \mathbb{P}_0 -almost surely, m_t is a (random) Gaussian probability measure with parameters

$$\begin{cases} \mu_1(m_t) = \sigma_0 \int_0^t e^{s-t} dB_s^0, \\ v(m_t) = \int_0^t e^{2(s-t)} \sigma^2(v(m_s)) ds. \end{cases}$$

Then, it is not difficult to see that $(\mu_1(m_t))_{t \geq 0}$ is an ergodic process (it is in fact an Ornstein-Uhlenbeck process with a negative mean-reverting parameter). As for the dynamics of the variance $(v(m_t))_{t \geq 0}$, we get

$$\partial_t v(m_t) = -2v(m_t) + \frac{1}{2}\sigma^2(v(m_t)), \quad t \geq 0.$$

Obviously, the above equation has several stationary solutions (say σ^*) if the function $x \mapsto \sigma^2(x)/4$ admits several fixed points. In the latter case, $(m_t)_{t \geq 0}$ has several invariant measures, whose form is similar to (3.18) except that σ therein is now replaced by any σ^* . In this specific example, it is noteworthy that the common noise has no influence on the mean field term $(v(m_t))_{t \geq 0}$ (which is completely deterministic). Clearly, this is a consequence of the additive structure of the noise, the effect of which is just to shift (or to translate) the measures $(m_t)_{t \geq 0}$.

Chapter 2 takes a first step in the direction of identifying the classes of mean field dynamics for which the common noise allows for uniqueness recovery by studying the case of a non-convex potential V , a linear interaction force and no idiosyncratic noise, i.e $\sigma = 0$. This case is not covered by [Daw83, HT10], however, it is quite easy to see that when V is not convex and $\sigma = \sigma_0 = 0$, the equation

$$\partial_t m_t = \nabla \cdot (\nabla V m_t + \alpha(x - \mu_1(m_t))m_t), \quad (1.31)$$

admits several stationary solutions located on the minimisers of V .

When we add common noise to the system, the mean-field limit gives the following equation

$$d_t m_t = \frac{\sigma_0^2}{2} \Delta m_t dt + \nabla \cdot (\nabla V m_t + \alpha(x - \mu_1(m_t))m_t) dt - \sigma_0 \nabla m_t \cdot dB_t^0. \quad (1.32)$$

We show in Chapter 2 that under assumption (LTBa) and assuming that $W(x) = \alpha|x|^2/2$, for some $\alpha > 0$, which allows us to consider double-well potentials and therefore non-convex, we may recover uniqueness of the invariant measure in the sense of the Definition 1.24. More precisely, we have

Theorem 1.32. *Under the set of Assumptions (LTBa) and assuming that $W(x) = \alpha|x|^2/2$, for some $\alpha > 0$, there exists $\bar{\alpha} > 0$, such that for any $\alpha \geq \bar{\alpha}$, there exists a unique invariant measure $\bar{P} \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$*

for the solutions of (1.32). Moreover,

$$\bar{P}(dm) = \int_{\mathbb{R}^d} \delta_{\delta_x}(dm) m^*(dx), \quad (1.33)$$

where m^* admits a density with respect to the Lebesgue measure, also denoted m^* and satisfying for any $x \in \mathbb{R}^d$,

$$m^*(x) \propto \exp\left(-\frac{2}{\sigma_0^2} V(x)\right) \quad (1.34)$$

Also, we show exponentially fast convergence to the invariant regime in Wasserstein distance.

Unfortunately, in this case, we are no longer able to prove uniform in time propagation of chaos results. In chapter 3, we prove similar results extended to a more general framework. The major difference being that we are now able to consider models with idiosyncratic noise, i.e. when $\sigma > 0$. More precisely, we consider classes of models of the form

$$d_t m_t = \frac{\sigma_0^2}{2} \Delta m_t dt - \nabla \cdot (G m_t + F(x - \mu_1(m_t)) m_t) dt - \sigma_0 \nabla m_t \cdot dB_t^0. \quad (1.35)$$

We then make the following assumptions on G and F , which are similar to the ones on $-\nabla V$ and the interaction:

Assumptions - Long Time Behaviour (LTBb)

1. There exists a continuous function $\kappa : [0, +\infty) \rightarrow \mathbb{R}$, such that

$$\liminf_{r \rightarrow +\infty} \kappa(r) > 0,$$

and

$$(G(x) - G(y)) \cdot (x - y) \leq -\kappa(|x - y|) |x - y|^2;$$

2. G is L_G -Lipschitz continuous.
3. F is L_F -Lipschitz continuous and there exists $\alpha > 0$ such that

$$(F(x) - F(y)) \cdot (x - y) \leq -\alpha |x - y|^2;$$

These assumptions are generalisations of the case treated in Chapter 2 as we only get ride of the gradient structure for the confinement and interaction forces. Nevertheless, the underlying intuition remains quite similar: **(1)** the confinement term G brings the particles back into a compact set thanks to its strict asymptotic monotonicity properties. However, we can consider forces G that are not monotonic near the origin, such as $x \mapsto x^3 - x$. **(2)** The interaction term F still involves the first moment. We do indeed separate from the linearity assumption, but we keep in mind the idea that F must help attract the process m towards its first moment.

In this case, we can still exhibit cases of uniqueness recovery thanks to common noise. More pre-

cisely, we have the following result:

Theorem 1.33. *There exists $\bar{\sigma} > 0$, $\bar{\alpha} > 0$, such that for all $\sigma \leq \bar{\sigma}$, and $\alpha \geq \bar{\alpha}$ in Assumption (LTBb), there is a unique invariant measure $\bar{P} \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$ for the process $(m_t)_{t \geq 0}$, solution of Equation (3.4).*

The study of this kind of models and the impact of a finite-dimensional common noise on the long-time behavior is far from being covered by the two papers [Mai23, DMT24] mentioned in this introductory chapter.

There are of course still many questions to be addressed on this topic, a few are listed here:

1. Can we extend the results obtained in this thesis to more general models? In particular, we would like to get rid of the assumption that interaction only occurs via the first-order moment.
2. Can we study the impact of the common noise on the propagation of chaos property?
3. Can we develop numerical schemes to approximate the invariant measure of the process and study their convergence properties?

3 Statistical Analysis of diffusion models

The second section of this thesis deals with non-parametric estimation for diffusion processes. The general framework is the following: we consider a stochastic process X solution of the Stochastic Differential Equation (SDE)

$$\begin{cases} dX_t^\nu = -\nabla V(X_t^\nu) + \sigma(t, X_t^\nu)dB_t \\ \mathcal{L}(X_0^\nu) = \nu, \end{cases} \quad (1.36)$$

for some initial condition $\nu \in \mathcal{P}_2(\mathbb{R}^d)$. We want to recover information on the dynamic of X from observations of this process. Under some mild conditions on V and σ , there exists a unique solution to (1.36). For decades, the study of this type of dynamics has attracted much attention in order to answer the following questions: *What can be estimated in this model? From what observations? What are the optimal estimation rates?* These natural questions have been the origin of the publication of many research papers, see [GC16] for a very nice introduction. We refrain here from presenting an extensive literature review on the subject. However, the following table contains a summary of the main references on the subject in different contexts.

	Est. of σ	Est. of ∇V	Est. of the invariant measure
Continuous obs.		[CGC20] [Bos97] [CM05] [Leb97]	[DR06] [DR07] [Str18]
Discrete obs.	[Hof97] [CGCR07] [GCJ94] [JLM ⁺ 09] [GCLP92]	[Hof99][CGCR07] [Sch13] [BP03] [GHR04]	[AG21] [AG22] [AG23] [Sch12]
Noisy obs.	[Sch12]	[Sch11]	[MS24]

This type of process appears in many fields [Pap95, Ber93, Hul03, Bai57, Ric77]. With the development of data collection methods, it seems necessary to implement statistical methods to (1) develop predictive models adapted to real-life data; (2) ensure the validity of the probabilistic models used based on observed data, see for example [ASJ10].

In Chapter 4, we focus on the case where the diffusion matrix is diagonal and constant. More precisely, there exists $\sigma > 0$ such that for all $x \in \mathbb{R}^d$, $\sigma(x) = \sigma \text{Id}$. In this case, the model (1.36) can be rewritten as

$$dX_t = -\nabla V(X_t) + \sigma dB_t, \quad (1.37)$$

where B is a standard Brownian motion in \mathbb{R}^d . It is easy to prove that we can characterise the flow of laws of the solutions of (1.37), $(\mu_t)_{t \geq 0} = (\mathcal{L}(X_t))_{t \geq 0}$ by

$$\partial_t \mu_t = \frac{\sigma^2}{2} \Delta \mu_t + \nabla \cdot (\mu_t \nabla V). \quad (1.38)$$

This linear equation has been widely studied in the literature, see [Ris89] and the references therein. Under fairly mild assumptions on V (linear behaviour at infinity), we know that there exists a unique invariant measure for (1.37) denoted by $\bar{\mu}$. This invariant measure is absolutely continuous with respect to the Lebesgue measure. This density is also denoted by $\bar{\mu}$, and is given by

$$\bar{\mu}(x) = \frac{1}{Z_V} \exp\left(-\frac{2}{\sigma^2} V(x)\right), \quad (1.39)$$

for some normalization constant Z_V .

3.0.1 Ergodicity for diffusion models with constant volatility

In this section, we make some quick reminders on the notion of ergodicity for diffusion models of the form (1.37). Let us introduce the following set of assumptions

Assumptions - Ergodicity (Ergo)

1. The function b is differentiable and satisfies

$$|b(0)| \leq b_0 \quad \text{and} \quad \forall i \in \{1, \dots, d\}, \quad \|\partial_i b\|_\infty \leq b_1/d,$$

where b_0 and b_1 are positive constants.

2. There exists a function $V : \mathbb{R}^d \rightarrow \mathbb{R}$ differentiable and bounded below by a constant V_0 such that $b = -\nabla V$ and $V(0) = 0$.

3. There exists $\tilde{C}_b > 0$ and $\tilde{\rho}_b > 0$ such that $\langle x, b(x) \rangle \leq -\tilde{C}_b |x|$, $\forall x : |x| \geq \tilde{\rho}_b$.

Moreover, when considering this kind of process, it is possible to show results of exponential convergence towards the invariant measure. This has been a very vast field of research in the past decades.

Such convergence results can be shown using functional inequalities see for example [ABC⁺00] or the very nice course [CL17]. More recently, the use of coupling techniques developed by [LR86] allowed

to prove with purely probabilistic approach some contraction properties resulting in exponentially fast convergence to the stationary regime under (Ergo).

Theorem 1.34. *Under Assumptions (Ergo), there exists a constant $c > 0$, such that for any initial conditions $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$,*

$$d_1^{\mathbb{R}^d}(\mathcal{L}(X_t^\mu), \mathcal{L}(X_t^\nu)) \leq d_1^{\mathbb{R}^d}(\mu, \nu)e^{-ct}. \quad (1.40)$$

This approach has since been widely use in various domain, particularly in control [Con23] or interacting particle system theory [DEGZ20]. The unfamiliar reader shall find details on coupling by reflection and the proof of Theorem 1.34 in the Appendix A.

3.1 Presentation of the model and literature review

In Chapter 4, we consider a large range of potential V . Mainly, we work under the set of Assumptions (Ergo). In this case, we know that X is ergodic and admits a unique invariant density denoted by $\bar{\mu}$. The topic of estimating the invariant measure for ergodic processes has been addressed in many contexts [DR06, DR07, Str18, AG21, AG22, AG23]. Several works have considered various structures of processes, including jump processes [AN22].

It is well-known and natural that the estimation rate of a function strongly depends on its regularity. The adapted notion of regularity in such a context is Hölder regularity:

Definition 1.35. *Let $\alpha = (\alpha_1, \dots, \alpha_d) \in (0, \infty)^d$ and $\mathcal{L} = (\mathcal{L}_1, \dots, \mathcal{L}_d) \in (0, \infty)^d$. A function $g : \mathbb{R}^d \rightarrow \mathbb{R}$ is said to belong to the anisotropic Hölder class $\mathcal{H}_d(\alpha, \mathcal{L})$ of functions if, for all $1 \leq i \leq d$, g is $\lfloor \alpha_i \rfloor$ -differentiable in the i -th variable and the partial derivatives satisfy for $0 \leq k \leq \lfloor \alpha_i \rfloor$*

$$\|\partial_i^k g\|_\infty \leq \mathcal{L}_i \quad \text{and} \quad \forall t \in \mathbb{R}, \|\partial_i^{\lfloor \alpha_i \rfloor} g(\cdot + te_i) - \partial_i^{\lfloor \alpha_i \rfloor} g(\cdot)\|_\infty \leq \mathcal{L}_i |t|^{\alpha_i - \lfloor \alpha_i \rfloor}$$

where (e_1, \dots, e_d) is the canonical basis of \mathbb{R}^d .

We now provide a more detailed review of the literature on our topic of interest: the estimation of the invariant measure for ergodic diffusion processes. Statistical studies on this subject focused on two different asymptotic regimes: continuous observations of the process $(X_t)_{t \leq T}$ with $T \rightarrow \infty$, or low frequency observations given by $(X_{i\Delta})_i$ with $0 \leq i \leq n$. Since X is exponentially β -mixing under mild assumptions on b , low-frequency observations naturally relate to the i.i.d. case and present similar convergence rates.

On the other hand, continuous observations can be studied using fine probabilistic tools for ergodic continuous-time Markovian dynamics [BCG08, CG08, Lez01, Pau15] and precise estimates on the transition densities [CW97, QRZ03, QZ04]. The seminal works of [DR06, DR07] first established convergence rates of kernel estimators under continuous observations of X over a time interval $[0, T]$. In these works, the invariant density is still estimated through a kernel based estimator as in the i.i.d. case. [DR06, DR07] also obtain the convergence rate

$$T^{-\alpha/(2\alpha+d-2)}$$

when $d \geq 3$ and $\bar{\mu}^b$ is α -Hölder in the sense of Definition 1.35, with $\alpha = (\alpha, \dots, \alpha)$, for some $\alpha \geq 2$. When considering an anisotropic framework, the convergence rate depends on the effective average smoothness [Str18] and minimax rates are derived in [AG21]. Two natural questions remain:

Question 1.36. *What happens if we access discrete high-frequency observations of X , i.e. we observe $(X_{i\Delta_n})_i$ for $0 \leq i \leq n$ when $\Delta_n \rightarrow 0$ and $n\Delta_n \rightarrow \infty$?*

Question 1.37. *Can we still estimate the invariant measure when the observations are blurred by some noise?*

The Question 1.36 naturally arises with the advent of high-frequency data collection in particular in finance [ASJ14]. Hence, understanding the estimation rates under different asymptotic conditions becomes crucial. This includes specifying conditions on Δ_n that determine when continuous or low-frequency observations are more analytically pertinent. While this topic has only recently received substantial attention, pioneering works like [GHR04] and [CT16], which explore random sampling times, stand out. However, this question has recently been addressed in full generalities in [AG23] where the breakeven point between the high-frequency observations similar to continuous observations and low-frequency observations similar to the i.i.d. case is identified and studied. The continuous rate is known to be optimal [AG21] in any dimension, but the question of the optimality of the low frequency rate is still an open question, with the exception of the one dimensional case, solved in [AG22].

Answering Question 1.37 is crucial in statistical analyses for several compelling reasons. The presence of noise in observations is often unavoidable, especially in practical applications such as financial modelling, biological experiments, or sensor data analysis. Addressing the impact of noise in the analysis can help develop models and methodologies that are more applicable to real-world scenarios, increasing the external validity. This is the purpose of Chapter 4.

3.2 Contributions

As in [Sch11, Sch12], the aim of chapter 4 is to develop a statistical method that can take into account the presence of noise in the observations, this time to estimate the invariant measure. When noise is assumed to have an additive structure, existing literature uses Fourier inversion and kernel-based methods to recover the distribution of interest [Dev89, LT89, SC90]. Later works [CH88, Fan91, Fan93] establish minimax optimality of this procedure under the assumptions that the noise distribution is known and has a non-vanishing Fourier transform. It is important to note that in this setup, the convergence rates are slow. For instance, when the distribution of interest is α -Hölder regular and the noises are independent standard Gaussian variables, the optimal rate of convergence for any estimator is only $\log(n)^{-\alpha/2}$. This leads us to the following questions in our context:

Question 1.38. *What is the impact of noise on the estimation rates?*

Question 1.39. *Can we use the Markovian structure of the underlying process to obtain better convergence rates than for the classical deconvolution problems?*

In Chapter 4, we answer both questions 1.38 and 1.39. More precisely, we obtain that unlike the i.i.d. case, the structure of (4.2) and (4.3) allows better extraction of the information hidden by the noise. The estimation procedure is composed of three steps

- Step 1: Pre-averaging. To reduce noise effects, data is split into groups of equal size p and empirical means are built into the batches. These techniques are usual in the literature of statistical estimation for diffusion processes [JLM⁺09, JPV10, HP13, JM15, Sch11, Sch12].

We obtain the following process

$$\left(\frac{1}{p} \sum_{l=0}^{p-1} Y_{kp+l} \right)_{k=0}^{\lfloor n/p \rfloor - 1}.$$

Even though we lose some information, this pre-processing of the data allows us to significantly reduce the intensity of the noise. Indeed, for any $k \in \{0, \dots, \lfloor n/p \rfloor - 1\}$,

$$\frac{1}{p} \sum_{l=0}^{p-1} Y_{kp+l} = \frac{1}{p} \sum_{l=0}^{p-1} X_{(kp+l)\Delta_n} + \frac{\tau_n}{p} \sum_{l=0}^{p-1} \xi_{kp+l}. \quad (1.41)$$

Of course, since $(\xi_i)_i$ are standard Gaussian random variables, we have

$$\frac{\tau_n}{p} \sum_{l=0}^{p-1} \xi_{kp+l} \stackrel{d}{=} \frac{\tau_n}{\sqrt{p}} \xi_1, \quad (1.42)$$

where equality holds in distribution.

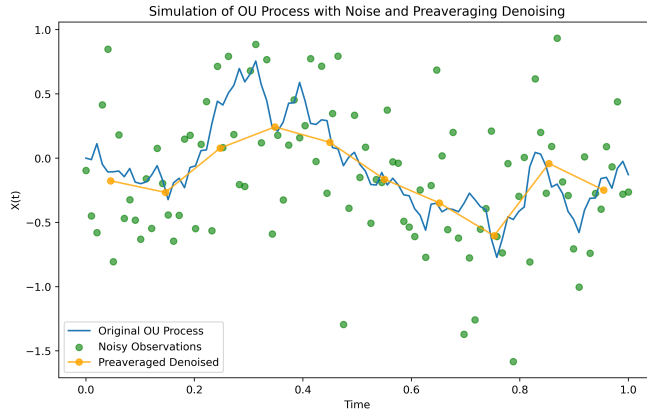


Figure 1.7: Pre-averaging of noisy observations of an Ornstein-Uhlenbeck process

- Step 2: Construction of a classical Kernel estimator. From the previously constructed process,

$$\hat{v}_{n,h,p}(x) = \frac{1}{\lfloor n/p \rfloor} \sum_{k=0}^{\lfloor n/p \rfloor - 1} \mathbf{K}_h \left(x - \frac{1}{p} \sum_{\ell=0}^{p-1} Y_{kp+\ell} \right),$$

where for any $y \in \mathbb{R}^d$, $\mathbf{K}_h(y) = \prod_{i=1}^d h_i^{-1} K(h_i^{-1} y_i)$, for some Kernel function $K : \mathbb{R} \rightarrow \mathbb{R}$. This estimation procedure is classical see [Tsy09] or Appendix B.

- Step 3: Debiasing procedure. Intuitively, the approach of Step 2 works because of the regularity

of X . Indeed, we have

$$\frac{1}{p} \sum_{\ell=0}^{p-1} Y_{kp+\ell,n} = X_{kp\Delta_n} + \frac{1}{p} \sum_{\ell=0}^{p-1} (X_{(kp+\ell)\Delta_n} - X_{kp\Delta_n}) + \frac{\tau_n}{p} \sum_{\ell=0}^{p-1} \xi_{kp+\ell,n}. \quad (1.43)$$

This expression can be seen as $X_{kp\Delta_n} + \text{noise}$ and the effective sampling frequency becomes $(p\Delta_n)^{-1}$. Moreover, the noise now comes from two sources: in addition to the noise $(\xi_{i,n})_{i,n}$, we now have a preaveraging error $p^{-1} \sum_{\ell=0}^{p-1} (X_{(kp+\ell)\Delta_n} - X_{kp\Delta_n})$. Using the pathwise regularity of X , this error should remain small when p is not too large. Moreover, since the random variables $(\xi_{i,n})_{i,n}$ are independent standard Gaussian variables, we can rewrite $p^{-1} \sum_{\ell=0}^{p-1} \xi_{kp+\ell,n} = \tau_n p^{-1/2} \tilde{\xi}_{k,n}$ where $\tilde{\xi}_{k,n}$ is also a standard Gaussian variable. Using finally the fact that

$$\frac{1}{p} \sum_{\ell=0}^{p-1} (X_{(kp+\ell)\Delta_n} - X_{kp\Delta_n}) \approx \frac{1}{p} \sum_{\ell=0}^{p-1} (B_{(kp+\ell)\Delta_n} - B_{kp\Delta_n}),$$

where the precise sense of the previous Equation is detailed in Chapter 4, we obtain that the effective noise intensity in the observation is now

$$\tilde{\tau}_{n,p} = \left(\frac{\tau_n^2}{p} + \frac{(p-1)(2p-1)\Delta_n}{12p} \right)^{1/2}.$$

Then we define,

$$\hat{\mu}_{n,\mathbf{h},p}(x) = \sum \mathbf{u}_\gamma \hat{\nu}_{n,\mathbf{h},p}(x + \gamma \tilde{\tau}_{n,p}), \quad (1.44)$$

for some sequence \mathbf{u}_γ and where the sum holds over all $\gamma = (\gamma_1, \dots, \gamma_d)$. Note that the choice of the sequence \mathbf{u}_γ is explicit in Chapter 4 and made in order to reduce the bias. Moreover, it is important to notice that such a combination of estimators does not change the variance (up to some constant independent of n, p, \mathbf{h}).

To study the rate of estimation for the estimator, we control the pointwise quadratic loss:

$$\mathbb{E}[|\hat{\mu}_{n,\mathbf{h},p}(x) - \mu(x)|^2].$$

Then we separately control the bias and the variance of $\hat{\mu}_{n,\mathbf{h},p}(x)$:

Proposition 1.40 (Control of the Bias). *Under (Ergo), assuming moreover that $\bar{\mu}$ is α -Hölder with $\alpha_1 \leq \dots \leq \alpha_d$, there exists a constant $C > 0$ so that for any $x \in \mathbb{R}^d$, any $p \in \{1, \dots, \lceil \Delta_n^{-1/2} \rceil\}$ and any $\mathbf{h} \in \mathbb{R}^d$, we have*

$$|\mathbb{E}_{\bar{\mu}}[\hat{\mu}_{n,\mathbf{h},p}(x)] - \bar{\mu}^b(x)| \leq C \begin{cases} \tau_n^{\alpha_1} + \sum_{i=1}^d h_i^{\alpha_i} & \text{if } p = 1, \\ \sqrt{p\Delta_n} + \frac{\tau_n^{\alpha_1}}{p^{\alpha_1/2}} + \sum_{i=1}^d h_i^{\alpha_i} & \text{if } p \geq 2. \end{cases}$$

The control of the bias term is somehow the main contribution of Chapter 4. In fact the presence of the residual terms coming from the noise are identified in the paper [MS24] for the first time. The upper bound for the variance term on the contrary is similar to the one of [AG23] even if the presence of noise makes the derivation more technical. The previous proposition allows to answer Question 1.38.

Proposition 1.41. *The best choice of pre-averaging is given by*

$$p^* := \left[(\tau_n^{2\alpha_1} \Delta_n^{-1})^{1/(1+\alpha_1)} \right] \vee 1,$$

and the impact of the noise on the estimation speed is then of order

$$\sqrt{p^* \Delta_n} + \frac{\tau_n^{\alpha_1}}{p^{*\alpha_1/2}}.$$

The optimal choice of the hyperparameters \mathbf{h} and p depends on the regularity of the quantity of interest. Of course, this regularity has no reason to be known. In Chapter 4, we set up an adaptive choice for the bandwidth \mathbf{h} inspired by the works [GL08, GL09, GL11], allowing to obtain Oracle inequalities, see [Cha13] for a nice introduction to these topics. The non-familiar reader shall also have a look to the appendix B where an example of an adaptive method is detailed. Finally, the choice of p is much more intricate and is an interesting research direction, more details are given in Chapter 4.

4 Composition of the manuscript

- ▶ Chapter 2 contains the paper

[A note on the Long-Time behaviour of Stochastic McKean-Vlasov Equations with common noise,](#)

which has been submitted for publication to Annals of Applied Probability.

- ▶ Chapter 3 contains the paper

[Ergodicity of some stochastic Fokker-Planck equations with additive common noise](#)

which is co-authored with François Delarue and Etienne Tanré, it has been submitted for publication to Annales de l'Institut Henri Poincaré.

- ▶ Chapter 4 contains the paper

[Estimation of the invariant measure of a multidimensional diffusion from noisy observations,](#)

which is co-authored with Grégoire Szymanski, it has been submitted for publication to Annals of Statistics.

Part I

Probabilistic analysis of interacting particle systems with common noise

Chapter 2

An introduction to long time behavior of non-linear stochastic Fokker-Planck equation

Abstract

This paper presents an investigation into the long-term behaviour of solutions to a non-linear stochastic McKean-Vlasov equation with common noise. The equation arises naturally in the mean-field limit of systems composed of interacting particles subject to both idiosyncratic and common noise. Initially, we demonstrate that the addition of common diffusion in each particle's dynamics does not disrupt the established stability results observed in the absence of common noise. However, our main objective is to understand how the presence of common noise can restore the uniqueness of equilibria. Specifically, in a non-convex landscape, we establish uniqueness and convergence towards equilibria when there is no idiosyncratic noise in the system.

Contents

1	Introduction	45
1.1	The process m as a mean field limit	50
2	Existence of an invariant measure for the stochastic flow of measures.	50
2.1	The existence result	53
3	The case of a uniformly convex confinement potential	55
3.1	Uniform in time propagation of chaos	56
3.2	Uniqueness of the invariant measure	58
3.3	Example	60
4	Non-convex potential without idiosyncratic noise	61
4.1	Existence of an invariant measure	62
4.2	Uniqueness of the invariant measure and exponential decay	63
5	Appendix	70
5.1	Proof of Proposition 2.8	70
5.2	Proof of Lemma 2.9	70
5.3	A measurable selection result	71

5.4	Proof of Proposition 2.22	72
5.5	Proof of Lemma 2.23	73

1 Introduction

We consider the following non-linear Stochastic Partial Differential Equation on $[0, +\infty) \times \mathbb{R}^d$,

$$d_t m_t = \nabla \cdot \left(\frac{\sigma^2 + \sigma_0^2}{2} \nabla m_t + m_t b(t, \cdot, m_t) \right) dt - \sigma_0 \nabla m_t \cdot dB_t^0. \quad (2.1)$$

This SPDE is posed on a filtered probability space $(\Omega_0, \mathcal{F}^0, \mathbb{F}^0, \mathbb{P}_0)$, B^0 is a d -dimensional \mathbb{F}^0 -Brownian motion, the drift $b : [0, +\infty) \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ depends in time, space and measure, σ and σ_0 are two non-negative constants. This paper aims to study the long-time behavior of solutions of Equation (3.1), and more precisely, to understand the effect of the common noise on the asymptotic stability. We assume that the drift term b has a specific linear structure with respect to the measure variable, with two continuously differentiable functions V and W such that $b(t, x, \mu) = -\nabla V(x) - \nabla W * \mu(x)$, where $*$ stands for the convolution operator. This assumption is typical when studying long-time behavior in McKean-Vlasov type equations, see *e.g* [Bas20, BGG13, DMT19, HT10, Tug13]. In the following, we are interested in getting existence and uniqueness results for the invariant measure of the probability measure valued process (m_t) . As the problem under consideration falls within the realm of McKean-Vlasov type, one may expect existence of an invariant measure and uniqueness at least in some specific cases. In this paper, we verify whether, in the case where W is convex and V is uniformly convex, the introduction of common noise does not compromise the classical uniqueness results of [BRV98]. When V is not convex, the matter becomes considerably more intricate, studied so far without the presence of common noise, only partial results are known. There exist cases in which the uniqueness of the invariant measure is not satisfied. Unlike linear elliptic equations, the presence of nonlinearity leads to the existence of multiple invariant measures. Specifically, it has been proven in [HT10] that when the confinement potential uniformly convex outside of a ball centred in the origine, admits a double-well and the diffusion coefficient σ is sufficiently small, there exist exactly three invariant solutions of the following equation:

$$\partial_t m_t = \frac{\sigma^2}{2} \Delta m_t + \nabla \cdot (m_t (\nabla V + \nabla W * m_t)).$$

Since 2019, several papers [Del19], [DFT20] or [DV21] investigate the restoration of uniqueness in mean-field games derived from deterministic differential games with a large number of players by introducing an external noise. In a similar manner, this paper explores the restoration of uniqueness of the invariant measure by introducing common noise to the system. More precisely, we prove existence and uniqueness of the invariant measure for process (m_t) in the following cases: (1) when the confinement potential V is uniformly convex and the interaction potential is convex; (2) when the potential V is not convex and there is no idiosyncratic noise in the system, *i.e* $\sigma = 0$. More precisely, in the previously mention case, we get uniqueness of the invariant measure for solutions of

$$d_t m_t = \nabla \cdot \left(\frac{\sigma_0^2}{2} \nabla m_t + m_t b(t, \cdot, m_t) \right) dt - \sigma_0 \nabla m_t \cdot dB_t^0. \quad (2.2)$$

In this case, we also get exponential rates of convergence to the invariant measure. More precisely, we show the existence of $\bar{P} \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$, and a constant $\eta > 0$, such that for all initial condition P_0 , we get the existence of a constant $C > 0$, such that for each time $t > 0$,

$$d_1^{\mathcal{P}(\mathbb{R}^d)}(P_t, \bar{P}) \leq C e^{-\eta t},$$

where $P_t = \mathcal{L}(m_t)$, and $d_1^{\mathcal{P}(\mathbb{R}^d)}(\cdot, \cdot)$ stands for the Wasserstein distance.

Obtaining uniqueness of the invariant measure in the general case of Equation (3.1) without strong convexity assumptions on the confinement potential appears to be a challenging problem that is not completely solved in this paper. The results obtained in the present paper are a step towards understanding the long-time behavior of the solutions to Equation (3.1), but are far from providing a complete understanding of the asymptotic stability for that kind of Stochastic Fokker-Planck Equations with additive common noise.

Probabilistic setting & Motivation. Let us consider a filtered probability space $(\Omega^1, \mathcal{F}^1, \mathbb{F}^1, \mathbb{P}^1)$. Then, we define the following product structure

$$\Omega = \Omega^0 \times \Omega^1, \quad \mathcal{F}, \quad \mathbb{F}, \quad \mathbb{P},$$

where $(\mathcal{F}, \mathbb{P})$ is the completion of the set $(\mathcal{F}^0 \otimes \mathcal{F}^1, \mathbb{P}^0 \otimes \mathbb{P}^1)$ and \mathbb{F} is the right continuous augmentation of $(\mathcal{F}_t^0 \otimes \mathcal{F}_t^1)_{t \geq 0}$. We also consider a d -dimensional Brownian motion B^0 supported by $(\Omega^0, \mathcal{F}^0, \mathbb{P}^0)$, adapted to \mathbb{F}^0 and another Brownian motion B supported by $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1)$, adapted to \mathbb{F}^1 and independent of \mathbb{F}^0 . Let us now consider a probability measure on the space of probability measures $P_0 \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$, we are able to define m_0 , a \mathcal{F}_0^0 -measurable random variable with value in the space of probability measure $\mathcal{P}(\mathbb{R}^d)$ and such that $\mathcal{L}(m_0) = P_0$, in the sense that for any bounded measurable function $F : \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}$, $\mathbb{E}_{\mathbb{P}^0}[F(m_0)] = \langle P_0; F \rangle$. We can now define on the whole probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ a random variable X_0 such that $\mathcal{L}(X_0 | \mathcal{F}_0^0) = m_0$ almost surely. Let us define the stochastic process X evolving in \mathbb{R}^d , supported by $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, which dynamic is given by

$$\begin{cases} dX_t = -\nabla V(X_t)dt - \nabla W * m_t(X_t)dt + \sigma dB_t + \sigma_0 dB_t^0 \\ X_{|t=0} = X_0, \end{cases} \quad (2.3)$$

where m_t stands for the conditional law of the random variable X_t , with respect to the σ -algebra \mathcal{F}_t^0 . Precisely, $m_t = \mathcal{L}(X_t | \mathcal{F}_t^0)$ almost surely, and B is a d -dimensional \mathbb{F}^1 -Brownian motion independent of \mathbb{F}^0 . The dynamic of the process $(m_t)_{t \geq 0}$ is well known and given by the following Lemma (see for example [CD18b] among other references).

Lemma 2.1. *The measure valued process $(m_t)_{t \geq 0}$ is solution in the weak sense of the following Stochastic Partial Differential Equation:*

$$d_t m_t = \nabla \cdot \left(\frac{\sigma^2 + \sigma_0^2}{2} \nabla m_t + m_t b(t, \cdot, m_t) \right) dt - \sigma_0 \nabla m_t \cdot dB_t^0,$$

with initial condition $\mathcal{L}(m_0) = P_0$.

Equation (3.1) connects closely to the McKean-Vlasov Equation with common noise. It shows how large systems of interacting particles evolve. In mathematical finance, this model is particularly useful for situations like inter-bank borrowing and lending systems (see [CFS15] or [GSSS15]). Studying the long-term behavior of solutions to Equation (3.1) is important because it gives us insight of the behavior of the solution to Equation (3.5).

Literature. Stochastic Partial Differential Equations in the more general form,

$$d_t m_t = \left[\sum_{i,j} \nabla_{ij}^2 (a_{ij}(t, \cdot, m_t) m_t) + \operatorname{div} (m_t b(t, \cdot, m_t)) \right] dt - \sigma_0(t, \cdot, m_t) \nabla m_t \cdot dB_t^0, \quad (2.4)$$

have been extensively studied in recent decades as it naturally arises in several applications. Equation (2.4) is linked to the stochastic scalar conservations law of the form

$$d_t m_t + \nabla \cdot (\sigma_0(\cdot, u_t) u_t) \circ dW_t = 0, \quad (2.5)$$

where \circ stands for the Stratonovich stochastic integral. In the case where $\sigma_0(x, \mu) = \sigma_0(\mu(x))$, meaning that the diffusion coefficient depends in the measure in a local way, this class of equations has been introduced in [LPS13] paving the way to several papers dealing with well-posedness of solutions of (2.5) in various frameworks [LPS14, GS15, FG16, GS17, FG19]. Uniqueness of the solutions to (3.1) is a well known result in the class of solutions admitting a square integrable density with respect to the Lebesgue measure, see [KX99], and has been shown recently without any further moments assumptions in [CG19].

In a slightly different context, a series of papers demonstrated the well-posedness for a large class of Stochastic Differential Equations similar to (3.1), called *Mean Reflected Stochastic Differential Equations*, see [BEH18], [BH21] and [BCCdRH20]. More precisely, [BH21] and [BCCdRH20] state conditional propagation of chaos under regularity conditions on the drift and diffusion terms. This equations naturally appears when considering interacting particle systems with constraint on the empirical measure of the systems and then the study of such equations is particularly important for example for applications to Mean Field Games.

In this paper, we focus on the following specific Stochastic McKean-Vlasov Equation with common noise:

$$d_t m_t = \nabla \cdot \left(\frac{\sigma_0^2}{2} \nabla m_t + m_t b(t, \cdot, m_t) \right) dt - \sigma_0 \nabla m_t \cdot dB_t^0. \quad (2.6)$$

As mentioned before, extensive research has been conducted on the equation in question, and recent studies have made notable contributions to understanding its properties. For example [HvS21] explores the existence and uniqueness of solutions for McKean-Vlasov Stochastic Differential Equations (SDEs) with common noise. Similarly, [Mar21] proposes a regularization approach in an infinite-dimensional setting for the McKean-Vlasov equation with Wasserstein diffusion, enhancing the understanding of solutions' regularity properties. Additionally, [KNRS22] investigate the well-posedness and numerical

methods for McKean-Vlasov equations with common noise, providing valuable insights on the stability and convergence of computational approaches for solving these equations.

However, to the best of our knowledge, little is known about the asymptotic behavior of the solutions of equations of the form (2.6). In the case without common noise, $\sigma_0 = 0$, where m is a deterministic flow of measures, past research has focused on various aspects of the solutions of (3.5), including existence, uniqueness ([McK66], [Fun84], [GKM⁺96]), and stability. Over the past two decades, significant advancements have been made in understanding the convergence to equilibrium for solutions of the deterministic McKean-Vlasov equation. For example, see [CMV03] or [CMV06] for proofs of an exponential convergence rate to equilibrium under strict convexity conditions on the potentials V and W . The case without strict convexity assumptions is more intricate. Nevertheless, through a thorough examination of the dissipation of the Wasserstein distance, [BGG13] showed an exponential convergence to equilibrium in a weakly-convex case. Recently, involving a coupling method issued from [LR86], it has been shown using nice concentration properties from [Ebe16] that the convergence to equilibrium holds with an exponential speed in the case of a confinement potential that is only convex far from the origin, as seen in [DEGZ20]. The latter shows uniform in time propagation of chaos property, as introduced in [Kac56] and [Szn91], allowing one to conclude the uniqueness of the invariant measure and provide a rate of convergence to equilibrium.

Organisation of the paper. This paper has three main parts in which we adopt a probabilistic approach. In Section 2, we study the existence of an invariant measure for the process $(m_t)_{t \geq 0}$, with dynamic given by Equation (3.1), and provide conditions on the potentials V and W for the existence of such an invariant measure. We also give moment estimates for the invariant measures. In section 3, we study the uniqueness of invariant measures with a uniformly convex confinement potential V , adapting known results without common noise. We also discuss the Ornstein-Uhlenbeck process with common noise, where the invariant measure can be explicitly described. Moreover, we demonstrate uniform-in-time propagation of chaos and convergence to equilibrium. Section 4 explores the same topic for non-convex potential V when $\sigma = 0$ and we prove that common noise can help to restore uniqueness of the invariant measure. Technical proofs are provided in Appendix 5.

Definition and notation. Throughout the paper, for a Polish space E we write $\mathcal{P}(E)$ for the space of Borel probability measures on E equipped with the topology of weak convergence and the corresponding Borel σ -algebra. We also denote by $\langle \cdot ; \cdot \rangle$ the duality product on $\mathcal{P}(E)$. In this paper, we consider a stochastic process $(m_t)_{t \geq 0}$ with value in the space of probability measures $\mathcal{P}(\mathbb{R}^d)$. We denote by P_t the law $\mathcal{L}(m_t)$ of m_t , for $t \geq 0$, which is a probability measure on the space of probability measure. Then $(m_t)_{t \geq 0}$ is a continuous $\mathcal{P}(\mathbb{R}^d)$ -valued process, and $P = (P_t)_{t \geq 0}$ belongs to $C([0, +\infty[; \mathcal{P}(\mathcal{P}(\mathbb{R}^d)))$, the space of continuous functions from $[0, +\infty[$ to $\mathcal{P}(\mathcal{P}(\mathbb{R}^d))$.

Whenever there exists a distance d so that (E, d) is a metric space, we call $\mathcal{P}_p(E)$, for any $p > 0$, the collection of elements $\mu \in \mathcal{P}_p(E)$ such that

$$\exists x_0 \in E : \int_E d(x_0, x)^p \mu(dx) < +\infty.$$

In fact, the integral above is finite or not, whatever the choice of x_0 . We then define for any $p \geq 1$, the p -Wasserstein distance on $\mathcal{P}_p(E)$ as

$$d_p^E(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \left(\int_{E \times E} d(x, y)^p \pi(dx, dy) \right)^{1/p}, \quad \mu, \nu \in \mathcal{P}_p(E),$$

where $\Pi(\mu, \nu)$ stands for the set of all couplings of μ and ν (i.e., all the joint probability measures on $E \times E$ with μ and ν as first and second marginals).

In this paper, we mainly use a probability-based approach, often switching between the measure-valued stochastic process m and its probabilistic counterpart X , which solves Equation (3.5). At this point, it is worth noting that we are concerned with a stochastic process $(m_t)_{t \geq 0}$ that takes values in the space of probability measures $\mathcal{P}(\mathbb{R}^d)$. This means that at each time $t > 0$, we are dealing with measures on the space of probability, rather than on the underlying space \mathbb{R}^d . To study this process, we use a probability-based approach and introduce several definitions that are specific to this setting. For instance, we define the notion of an invariant measure in $\mathcal{P}(\mathcal{P}(\mathbb{R}^d))$, which is a probability measure that remains invariant under the evolution of the stochastic process. These definitions are essential for our analysis and detailed below.

Definition 2.2. *Let us consider a random variable X defined on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, and $P \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$. According to Lemma 2.4 in [CD18b], for $\mathbb{P}^0 - a.e.$ $\omega_0 \in \Omega^0$, $X(\omega_0, \cdot)$ is a random variable on $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1)$. By defining $\mathcal{L}^1(X) : \Omega^0 \ni \omega_0 \mapsto \mathcal{L}(X(\omega_0, \cdot))$, we get a random variable from $(\Omega^0, \mathcal{F}^0, \mathbb{P}^0)$ into $\mathcal{P}(\mathbb{R}^d)$, providing a conditional law of X given \mathcal{F}^0 . Finally, we say that $\mathcal{L}(X) = P$ whenever $\mathcal{L}^1(X)$ is distributed with respect to P .*

Definition 2.3 (Invariant measure). • *We say that $\bar{P} \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$, is an invariant measure for the process $(m_t)_{t \geq 0}$, if whenever m_0 is distributed according to \bar{P} , then at each time, the law of m_t is independent of t . More precisely, we say that \bar{P} is an invariant measure if and only if*

$$\mathcal{L}(m_0) = \bar{P} \quad \Rightarrow \quad \mathcal{L}(m_t) = \bar{P}, \quad \forall t > 0.$$

- *We say that a stochastic process X on the probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ admits an invariant measure in $\mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$ if and only if, the measure valued stochastic process $m_t = \mathcal{L}(X_t | \mathcal{F}_t^0)$ admits an invariant measure.*

Definition 2.4. *For any function f , with at most linear growth, such that $(x, y) \mapsto f(|x - y|)$ defines a distance on \mathbb{R}^d , we define the following distance on $\mathcal{P}_1(\mathbb{R}^d)$*

$$d_f^{\mathbb{R}^d}(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} f(|x - y|) \pi(dx, dy) \right). \quad (2.7)$$

Moreover, we define on $\mathcal{P}_1(\mathcal{P}_1(\mathbb{R}^d))$,

$$d_f^{\mathcal{P}(\mathbb{R}^d)}(P, Q) = \inf_{\Gamma \in \Pi(P, Q)} \int_{\mathcal{P}_2(\mathbb{R}^d)} d_f^{\mathbb{R}^d}(\mu, \nu) \Gamma(d\mu, d\nu),$$

for any $P, Q \in \mathcal{P}_1(\mathcal{P}_1(\mathbb{R}^d))$.

1.1 The process m as a mean field limit

Thanks to the definition of the previous subsection, we are now ready to give an interpretation of the process (m_t) in terms of mean field limit for interacting particle system. Let us consider $P_0 \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$, let also $N \geq 1$ be an integer, and $(X_0^{1,N}, \dots, X_0^{N,N})$, N random variables which are conditionally independent and identically distributed with respect to \mathcal{F}_0^0 , such that $\mathcal{L}(X_0^{i,N}) = P_0$ for all $i \in \{1, \dots, N\}$. We now define the following interacting particle system

$$\begin{cases} dX_t^{i,N} = -\nabla V(X_t^{i,N}) - N^{-1} \sum_{j=1}^N \nabla W(X_t^{i,N} - X_t^{j,N}) dt + \sigma dB_t^i + \sigma_0 dB_t^0, \\ X_{|t=0}^{i,N} = X_0^{i,N}, \quad \forall i \in \{1, \dots, N\}, \end{cases}$$

where the B^i are independent d -dimensional \mathbb{F}^1 -Brownian motion which are independent of \mathbb{F}^0 . Then, we consider the mean-field limit system $(\bar{X}^1, \dots, \bar{X}^N)$ driven by

$$\begin{cases} d\bar{X}_t^i = -\nabla V(\bar{X}_t^i) - \nabla W * m_t(\bar{X}_t^i) dt + \sigma dB_t^i + \sigma_0 dB_t^0, \\ \bar{X}_{|t=0}^i = \bar{X}_0^i, \quad \forall i \in \{1, \dots, N\}, \end{cases}$$

where $(\bar{X}_0^i)_i$ are conditionally *iid* random variables with respect to \mathcal{F}_0^0 , such that $\mathcal{L}(\bar{X}_0^i) = P_0$, for all $i \in \{1, \dots, N\}$. Our framework is exactly the same as the classical one for mean field games system with common noise. However here, the law of the initial conditions is random. Then, conditioning with respect to the σ -algebra \mathcal{F}_0^0 , we get back to a more classical framework where the initial condition is a deterministic measure. More precisely, as stated in [CD18b], under sufficient regularity conditions on the transport part b mainly Lipschitz continuity with respect to the space and measure variables, we get that for any fixed $t \geq 0$,

$$\lim_{N \rightarrow +\infty} \mathbb{E}[|\bar{X}_t^i - X_t^{i,N}|^2] + \mathbb{E}[d_2^{\mathbb{R}^d}(m_t^N, m_t)] = 0,$$

where $d_2^{\mathbb{R}^d}(\cdot, \cdot)$ stands for the classical Wasserstein distance on \mathbb{R}^d , and $m_t^N := N^{-1} \sum_{i=1}^N \delta_{X_t^{i,N}}$ is the empirical measure of the interacting particle system.

2 Existence of an invariant measure for the stochastic flow of measures.

Let us consider a stochastic process $(m_t)_{t \geq 0}$ with value in the space of probability measures $\mathcal{P}(\mathbb{R}^d)$, and with dynamic given by (3.1). Then, m is a weak solution of

$$d_t m_t = \nabla \cdot \left(\frac{\sigma^2 + \sigma_0^2}{2} \nabla m_t + m_t (\nabla V + \nabla W * m_t) \right) dt - \sigma_0 \nabla m_t \cdot dB_t^0. \quad (2.8)$$

More precisely, for all $t \geq 0$ and $\varphi \in C_c^\infty(\mathbb{R}^d)$,

$$d\langle m_t, \varphi \rangle = \langle m_t, L_{m_t} \varphi \rangle dt + \sigma_0 \langle m_t, (\nabla \varphi)^\top \rangle dB_t^0,$$

where for any probability measure $m \in \mathcal{P}(\mathbb{R}^d)$, the operator L_m acts on a smooth function φ of compact support by

$$L_m \varphi = -(\nabla V + \nabla W * m) \cdot \nabla \varphi + \frac{\sigma_0^2 + \sigma^2}{2} \Delta \varphi,$$

where ∇, Δ respectively stands for the gradient and laplacian operator, while \cdot denotes the usual inner product in \mathbb{R}^d , and for any $\varphi \in C_c^\infty(\mathbb{R}^d)$ and any probability measure m ,

$$\langle m; \varphi \rangle = \int_{\mathbb{R}^d} \varphi dm.$$

In this section, we aim at giving conditions on the potentials V and W to ensure existence of an invariant measure for the process (m_t) . Let us now consider the following assumptions:

Assumption 1. 1. *There exists a continuous function $\kappa : [0, +\infty) \rightarrow \mathbb{R}$, such that*

$$\liminf_{r \rightarrow +\infty} \kappa(r) > 0,$$

and

$$(\nabla V(x) - \nabla V(y)) \cdot (x - y) \geq \kappa(|x - y|)|x - y|^2.$$

2. ∇V is L_V -Lipschitz continuous.

Assumption 2. 1. *W is symmetric, i.e., $W(x) = W(-x)$ for all $x \in \mathbb{R}^d$.*

2. ∇W is L_W -Lipschitz continuous.

In the following, we work under the set of assumptions 1&2. The assumption regarding confinement potential V primarily ensures convexity at infinity, which helps keep the process within a compact set with a high probability. We can moreover note that this implies the existence of $m_V > 0$ and $M_V \geq 0$, such that

$$(\nabla V(x) - \nabla V(y)) \cdot (x - y) \geq m_V |x - y|^2 - M_V.$$

The two following propositions stand.

Proposition 2.5. *Under Assumptions 1 and 2 and for an initial condition X_0 satisfying $\mathbb{E}[|X_0|^2] < \infty$, the conditional McKean-Vlasov equation (3.5) has a unique \mathbb{F} -progressively measurable solution $(X_t)_{t \geq 0}$, with continuous trajectories, such that, for all $T > 0$, $\mathbb{E}[\sup_{0 \leq t \leq T} |X_t|^2] < \infty$.*

Proof. The proof consists in a straightforward fixed point argument, using the Lipschitz properties of the two coefficients ∇V and ∇W . We refer to [CD18b, Chap. 2]. \square

By the superposition principle for conditional McKean-Vlasov equations, see [LSZ20], we deduce that existence and uniqueness also hold true for the stochastic Fokker-Planck equation:

Proposition 2.6. *Let Assumptions 1 and 2 be in force. For an initial condition m_0 such that*

$$\mathbb{E}_{\mathbb{P}^0} \int_{\mathbb{R}^d} |x| m_0(dx) < \infty,$$

the stochastic Fokker-Planck equation (2.8) has a unique solution $(m_t)_{t \geq 0}$ in the space of \mathbb{F}^0 -progressively measurable processes with values in $\mathcal{P}_2(\mathbb{R}^d)$. Moreover, this solution satisfies

$$\mathbb{E}_{\mathbb{P}^0} \left[\sup_{0 \leq t \leq T} \int_{\mathbb{R}^d} |x|^2 m_t(dx) \right] < \infty.$$

In this section, we start with a result regarding uniform-in-time control of the process $(X_t)_{t \geq 0}$ with initial condition in $\mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$ dynamic given by (3.5). To prove the existence of an invariant measure for $(m_t)_{t \geq 0}$ solution of (3.1), we will use the concept of intrinsic derivative for a functional defined on a space of measure, as seen in [CDLL19].

Definition 2.7. *Let us define $\mathcal{C}_b^2(\mathcal{P}(\mathbb{R}^d))$ as the collection of continuous, bounded functions $F : \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}$ with the following properties:*

- *There exists a unique continuous and bounded function $\partial_m F : \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \rightarrow \mathbb{R}$ such that*

$$\lim_{h \rightarrow 0} \frac{F(m + h(m' - m)) - F(m)}{h} = \int_{\mathbb{R}^d} \partial_m F(m, v)(m' - m)(dv),$$

for all $m, m' \in \mathcal{P}(\mathbb{R}^d)$ and

$$\int_{\mathbb{R}^d} \partial_m F(m, v) m(dv) = 0, \quad m \in \mathcal{P}(\mathbb{R}^d);$$

- *The mapping $x \mapsto \partial_m F(m, x)$ is continuously differentiable with uniformly bounded gradient $D_m F(m, x)$ in (m, x) ;*
- *For any fixed $x \in \mathbb{R}^d$, every component of the \mathbb{R}^d -valued function $m \mapsto D_m F(m, x)$ satisfies the same conditions as the first two bullet points, resulting in a continuous and bounded $D_m^2 F(m, x, y) \in \mathbb{R}^{d \times d}$;*
- *For $m \in \mathcal{P}(\mathbb{R}^d)$, we use $D_x D_m F(m, v)$ to denote the Jacobian of the function $x \mapsto D_m F(m, x)$, which is assumed to be continuous and bounded in (m, x) .*

Proposition 2.8. *Considering a measure valued process (m_t) , which dynamic is given by (3.1), and defining $P_t = \mathcal{L}(m_t)$, we have that for any bounded and twice differentiable function $F \in \mathcal{C}_b^2(\mathcal{P}(\mathbb{R}^d), \mathbb{R})$:*

$$\langle P_t - P_0, F \rangle = \int_0^t \langle P_s, \mathcal{M}F \rangle ds, \quad \forall t > 0, \quad (2.9)$$

where we define, for $m \in \mathcal{P}(\mathbb{R}^d)$,

$$\begin{aligned} \mathcal{M}F(m) &:= \int_{\mathbb{R}^d} \left[D_m F(m, x) \cdot b(x, m) + \frac{\sigma^2 + \sigma_0^2}{2} \nabla \cdot (D_m F(m, x)) \right] m(dx) \\ &\quad + \frac{\sigma_0^2}{2} \int_{\mathbb{R}^{2d}} \text{Tr} [D_{mm}^2 F(m, x, y)] m(dx) m(dy), \end{aligned}$$

and where for all measurable and bounded function $\Phi : \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}$, and all $P \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$,

$$\langle P; \Phi \rangle = \int_{\mathcal{P}(\mathbb{R}^d)} \Phi(m) P(dm).$$

Moreover, \bar{P} is an invariant measure if and only if

$$\langle \bar{P}, \mathcal{M}F \rangle = 0, \quad \forall F \in \mathcal{C}_b^2(\mathcal{P}(\mathbb{R}^d), \mathbb{R}). \quad (2.10)$$

The proof of Proposition 2.8 is based on [LSZ20, Section 1.2] and postponed to the Appendix 5.1.

2.1 The existence result

We now present a result about the existence of an invariant measure for the equation of interest, adapting classical results for McKean-Vlasov equation without common noise where the flow of probability measures, represented as $(m_t)_{t \geq 0}$ is deterministic. We begin this section with the following Lemma, the proof of which is classical and postponed to Appendix 5.2.

Lemma 2.9. *Under Assumptions 1 & 2, let us consider $P_0 \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$. Then, denoting by $(X_t)_t$ the associated stochastic process with initial condition $\mathcal{L}(X_0) = P_0$ in the sense of Definition 3.1, and dynamic given by (3.5), we have the following uniform in time moment control:*

$$\sup_{t > 0} \mathbb{E} [|X_t|^2] < +\infty.$$

Now, we are ready to state the following Proposition:

Proposition 2.10. *Under Assumptions 1 & 2, the dynamical system given by (3.1) admits at least an invariant measure $\bar{P} \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$, i.e such that*

$$\int_{\mathcal{P}(\mathbb{R}^d)} \int_{\mathbb{R}^d} |x|^2 m(dx) \bar{P}(dm) < +\infty.$$

Proof of Proposition 2.10. Let us fix $P_0 \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$, i.e such that:

$$\int_{\mathcal{P}(\mathbb{R}^d)} \int_{\mathbb{R}^d} |x|^2 m(dx) P_0(dm) < +\infty.$$

Step 1. Let us now consider a random flow of probability measures $(m_t)_{t \geq 0}$ with dynamic given by (3.1) and initial condition $P_0 \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$. At each time $t \geq 0$, we denote by P_t the law of this

measure valued process. For $T > 0$, we define the process $(Q_T)_{T>0}$ with value in $\mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$ by

$$Q_T = T^{-1} \int_0^T P_t dt.$$

This defines a sequence of probability measures on $\mathcal{P}(\mathbb{R}^d)$. Let us show that $(Q_T)_{T>0}$ admits at least a convergent subsequence. Thanks to Prohorov theorem, we only need to show that this sequence is tight. For $R > 0$, let us consider the set

$$K_R = \left\{ m \in \mathcal{P}(\mathbb{R}^d), \int_{\mathbb{R}^d} |x|^2 m(dx) \leq R \right\}.$$

This set is compact for the topology of weak convergence in $\mathcal{P}(\mathbb{R}^d)$. Now, for $T > 0$,

$$\begin{aligned} Q_T(K_R) &= T^{-1} \int_0^T P_t(K_R) dt \\ &= T^{-1} \int_0^T \int_{\mathcal{P}(\mathbb{R}^d)} \mathbb{1}_{\{m \in K_R\}} P_t(dm) dt \\ &= T^{-1} \int_0^T \mathbb{E} [\mathbb{1}_{\{m_t \in K_R\}}] dt \\ &\geq 1 - \frac{c}{R}, \end{aligned}$$

where $c = \sup_{t>0} \mathbb{E} [|X_t|^2] < +\infty$, thanks to Lemma 2.9. Hence, for any $\varepsilon > 0$, there exists $R_\varepsilon > 0$ such that $Q_T(K_{R_\varepsilon}) > 1 - \varepsilon$, for all $T > 0$. This gives tightness of the sequence and then existence of a converging subsequence that we keep denoting by $(Q_T)_{T>0}$ in the following.

Step 2. Let us denote by Q the limit of this converging subsequence, and show that Q is an invariant measure. Let $T > 0$, and $F \in C_b^2(\mathcal{P}(\mathbb{R}^d), \mathbb{R})$, thanks to Proposition 2.8, we get that

$$\begin{aligned} \langle Q_T, \mathcal{M}F \rangle &= \int_{\mathcal{P}(\mathbb{R}^d)} \left(\int_{\mathbb{R}^d} [D_m F(m, x) \cdot b(x, m) + \frac{\sigma^2 + \sigma_0^2}{2} \nabla \cdot (D_m F(m, x))] m(dx) \right. \\ &\quad \left. + \frac{\sigma_0^2}{2} \int_{\mathbb{R}^{2d}} \text{Tr}[D_{mm}^2 F(m, x, y)] m(dx) m(dy) \right) Q_T(dm) \\ &= T^{-1} \int_0^T \int_{\mathcal{P}(\mathbb{R}^d)} \left(\int_{\mathbb{R}^d} [D_m F(m, x) \cdot b(x, m) + \frac{\sigma^2 + \sigma_0^2}{2} \nabla \cdot (D_m F(m, x))] m(dx) \right. \\ &\quad \left. + \frac{\sigma_0^2}{2} \int_{\mathbb{R}^{2d}} \text{Tr}[D_{mm}^2 F(m, x, y)] m(dx) m(dy) \right) P_t(dm) \\ &= T^{-1} \langle P_T - P_0, F \rangle. \end{aligned}$$

Hence, for all $F \in C_b^2(\mathcal{P}(\mathbb{R}^d))$, $\langle Q, \mathcal{M}F \rangle = 0$. This ensures that Q is an invariant measure.

Step 3. Finally we move on to the moment estimate. We know that there exists a subsequence of

Q_T which converges weakly to \bar{P} . Moreover,

$$\int_{\mathcal{P}(\mathbb{R}^d)} \int_{\mathbb{R}^d} |x|^2 m(dx) Q_T(dm) = \frac{1}{T} \int_0^T \int_{\mathcal{P}(\mathbb{R}^d)} \int_{\mathbb{R}^d} |x|^2 m(dx) P_t(dm) dt,$$

and

$$\sup_{T>0} \frac{1}{T} \int_0^T \int_{\mathcal{P}(\mathbb{R}^d)} \int_{\mathbb{R}^d} |x|^2 m(dx) P_t(dm) dt < +\infty,$$

as $\sup_{t>0} \mathbb{E}[|X_t|^2] < +\infty$. Moreover, the function $m \in \mathcal{P}_2(\mathbb{R}^d) \mapsto \int_{\mathbb{R}^d} |x|^2 m(dx)$ is lower semi continuous, and then,

$$\int_{\mathcal{P}(\mathbb{R}^d)} \int_{\mathbb{R}^d} |x|^2 m(dx) \bar{P}(dm) \leq \liminf_{T \rightarrow +\infty} \frac{1}{T} \int_0^T \int_{\mathcal{P}(\mathbb{R}^d)} \int_{\mathbb{R}^d} |x|^2 m(dx) P_t(dm) dt < +\infty.$$

□

From the previous result we get the following Corollary

Corollary 2.11. *For $P \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$, we have that $d_2^{\mathcal{P}(\mathbb{R}^d)}(P, \bar{P}) < +\infty$.*

3 The case of a uniformly convex confinement potential

In this section, we show that the process $(m_t)_{t \geq 0}$, driven by equation (2.8), has a unique invariant measure under strong convexity assumptions on the confinement potential. Moreover, our method allows us to find exponential rates of convergence toward the invariant measure for a specific set of initial conditions. We again consider $P_0 \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$ and a random variable X_0 such that $\mathcal{L}(X_0) = P_0$, following Definition 3.1. Next, we study the stochastic process (X_t) driven by equation (3.5) with the initial condition X_0 . In this part of the paper, we consider strict convexity assumptions on the confinement potential V . To be specific, we adopt the following assumptions throughout this section:

Assumption 3. • V is uniformly convex, more precisely, there exists $\beta > 0$ such that:

$$\nabla^2 V \geq \beta \text{Id}.$$

- ∇V is Lipschitz continuous.
- W is even, convex, and ∇W is globally Lipschitz continuous, with Lipschitz constant L_W .

Thanks to the previous section, under Assumption 3, a process X driven by equation (3.5) has an invariant measure. Specifically, there exists $\bar{P} \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$ such that the process $(m_t)_{t \geq 0}$, governed by the dynamic in Equation (3.1) and with an initial condition of $\mathcal{L}(m_0) = \bar{P}$, is invariant. This section starts with a key result that states uniform propagation of chaos uniformly in time. This result will then help us establish the uniqueness of the invariant measure and to give a rate of convergence to equilibrium.

3.1 Uniform in time propagation of chaos

In the case without common noise, the uniqueness of the invariant measure for the process X has been already been established see *e.g* [CGM08], [Mal01], and [BGG13]. In this section, our aim is to adapt this result to the case with common noise and obtain the uniqueness of the invariant measure for the probability measure valued stochastic process (m_t) , and exponentially fast convergence to the equilibria.

Theorem 2.12. *Let us consider $P, Q \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$, and two particle systems $\mathbf{X} = (X^1, \dots, X^N)$ and $\mathbf{X}^N = (X^{1,N}, \dots, X^{N,N})$ with dynamics,*

$$dX_t^i = -\nabla V(X_t^i)dt - \nabla W * m_t(X_t^i)dt + \sigma dB_t^i + \sigma_0 dB_t^0, \quad \forall i \in \{1, \dots, N\}, \quad (2.11)$$

and

$$dX_t^{i,N} = -\nabla V(X_t^{i,N})dt - \frac{1}{N} \sum_{j=1}^N \nabla W(X_t^{i,N} - X_t^{j,N})dt + \sigma dB_t^i + \sigma_0 dB_t^0, \quad \forall i \in \{1, \dots, N\}. \quad (2.12)$$

where $(X_0^i)_{i \in \{1, \dots, N\}}$ are independent and identically distributed such that $\mathcal{L}(\bar{X}_0^1) = P$ in the sense of Definition 3.1, and where the same holds for the second system with $\mathcal{L}(X_0^{1,N}) = Q$. Then, under assumptions 3, there exists a constant $C > 0$ depending only on the dimension d and the probability measures P and Q , such that

$$\mathbb{E} \left[d_2^{\mathbb{R}^d}(m_{\mathbf{X}_t}^{P,N}, m_{\mathbf{X}_t^N}^{Q,N}) \right] \leq C \left(e^{-\beta t} + \frac{1}{2\beta\sqrt{N-1}} \right),$$

where $m_{\mathbf{X}_t}^{P,N} = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^i}$ and $m_{\mathbf{X}_t^N}^{Q,N} = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i,N}}$.

The proof of this result closely follows the approach presented in [BRV98] and [Mal01], which shows the propagation of chaos for particle systems in cases without common noise. However, in our situation, we need to be careful with the interaction term and its dependency on the common noise.

Proof. We only sketch the proof, as it follows closely the proof of [Mal01, Thm 3.3]. Let $i \in \{1, \dots, N\}$, using Itô formula, we get

$$\begin{aligned} \frac{d}{dt} |X_t^{i,N} - X_t^i|^2 &= -2(X_t^{i,N} - X_t^i) \cdot (\nabla V(X_t^{i,N}) - \nabla V(X_t^i)) \\ &\quad - \frac{2}{N} \sum_{j=1}^N (X_t^{i,N} - X_t^i) \cdot (\nabla W(X_t^{i,N} - X_t^{j,N}) - \nabla W * m_t(X_t^i)). \end{aligned}$$

In order to control the second term, we make the following decomposition:

$$\begin{aligned}
 & -2(X_t^{i,N} - X_t^i) \left(N^{-1} \sum_{j=1}^N \nabla W(X_t^{i,N} - X_t^{j,N}) - \nabla W * m_t(X_t^i) \right) \\
 &= -2(X_t^{i,N} - X_t^i) \left(N^{-1} \sum_{j=1}^N \nabla W(X_t^{i,N} - X_t^{j,N}) - N^{-1} \sum_{j=1}^N \nabla W(X_t^i - X_t^j) \right) \\
 & - 2(X_t^{i,N} - X_t^i) \left(N^{-1} \sum_{j=1}^N \nabla W(X_t^i - X_t^j) - \nabla W * m_t(X_t^i) \right) \\
 &= \Xi_t^{i,N} + \Upsilon_t^{i,N}.
 \end{aligned}$$

Summing the first term over i shows and using the convexity of the interaction potential, we get

$$\frac{1}{N} \sum_{i=1}^N \Xi_t^{i,N} = -2N^{-2} \sum_{i,j=1}^N (X_t^{i,N} - X_t^{j,N} - X_t^i + X_t^j) (\nabla W(X_t^i - X_t^j) - \nabla W(X_t^i - X_t^j)) \leq 0.$$

Moreover, the second term can be decomposed into two terms

$$\begin{aligned}
 \Upsilon_t^{i,N} &= -2(X_t^{i,N} - X_t^i) \cdot \left((N-1)^{-1} \sum_{j=1}^N \nabla W(X_t^i - X_t^j) - \nabla W * m_t(X_t^i) \right) \\
 & - 2(X_t^{i,N} - X_t^i) \cdot \left(\left(\frac{1}{N} - \frac{1}{N-1} \right) \sum_{j=1}^N \nabla W(X_t^i - X_t^j) \right)
 \end{aligned}$$

For the first one

$$\begin{aligned}
 & \mathbb{E}[|\nabla W * m_t(X_t^i) - \frac{1}{N-1} \sum_{j=1}^N \nabla W(X_t^i - X_t^j)|^2] \\
 &= \mathbb{E}\left[\mathbb{E}\left[\left|\nabla W * m_t(X_t^i) - \frac{1}{N-1} \sum_{j=1}^N \nabla W(X_t^i - X_t^j)\right|^2 \middle| X_t^i, \mathcal{F}_t^0\right]\right] \\
 &= \mathbb{E}\left[\text{Var}\left[\frac{1}{N-1} \sum_{j=1}^N \nabla W(X_t^i - X_t^j) \middle| X_t^i, \mathcal{F}_t^0\right]\right] \\
 &\leq \frac{1}{N-1} \mathbb{E}[\mathbb{E}[|\nabla W(X_t^i - X_t^j)|^2 | X_t^i, \mathcal{F}_t^0]], \quad \text{for some } j \neq i \\
 &\leq \frac{L_W^2}{N-1} \mathbb{E}[\mathbb{E}[|X_t^i - X_t^j|^2 | X_t^i, \mathcal{F}_t^0]] \\
 &\leq \frac{2L_W^2}{N-1} \mathbb{E}[|X_t^1|^2],
 \end{aligned}$$

where the first inequality comes from the fact that $\mathbb{E}[\nabla W(X_t^i - X_t^j) | X_t^i, \mathcal{F}_t^0] = \nabla W * m_t(X_t^i)$. Then,

thanks to Lemma 2.9, there exists a constant $C > 0$, such that

$$\begin{aligned} & \mathbb{E}\left[\left(X_t^{i,N} - X_t^i\right) \cdot \left((N-1)^{-1} \sum_{j=1}^N \nabla W(X_t^i - X_t^j) - \nabla W * m_t(X_t^i)\right)\right] \\ & \leq \mathbb{E}[|X_t^{i,N} - X_t^i|^2]^{1/2} \mathbb{E}\left[\left|\nabla W * m_t(X_t^i) - \frac{1}{N-1} \sum_{j=1}^N \nabla W(X_t^i - X_t^j)\right|^2\right]^{1/2} \\ & \leq \frac{C}{\sqrt{N-1}} \mathbb{E}[|X_t^{i,N} - X_t^i|^2]^{1/2}. \end{aligned}$$

Now,

$$\mathbb{E}\left[\left|\frac{1}{N-1} \cdot \frac{1}{N} \sum_{j=1}^N \nabla W(X_t^i - X_t^j)\right|^2\right] \leq \frac{2L_W^2}{(N-1)^2} \mathbb{E}[|X_t^1|^2].$$

Finally, using once again Lemma 2.9 we get the existence of a constant C , such that:

$$\mathbb{E}[|\Upsilon_t^{i,N}|] \leq C \left(\frac{1}{N-1} + \frac{1}{\sqrt{N-1}} \right) \mathbb{E}[|X_t^{i,N} - X_t^i|^2]^{1/2}.$$

Finally, we get that

$$\frac{1}{N} \sum_{i=1}^N \frac{d}{dt} \mathbb{E}[|X_t^i - X_t^{i,N}|^2] \leq -\frac{2\beta}{N} \sum_{i=1}^N \mathbb{E}[|X_t^i - X_t^{i,N}|^2] + \frac{C}{\sqrt{N-1}} \left(N^{-1} \sum_{i=1}^N \mathbb{E}[|X_t^{i,N} - X_t^i|^2] \right)^{1/2},$$

for some constant $C > 0$. Let us now denote $v_N(t) = N^{-1} \sum_{i=1}^N \mathbb{E}[|X_t^i - X_t^{i,N}|^2]$, then we have

$$v_N'(t) \leq -2\beta v_N(t) + \frac{C}{\sqrt{N-1}} v_N(t)^{1/2}.$$

This gives, using Grönwall Lemma:

$$v_N(t)^{1/2} \leq e^{-\beta t} v_N(0) + \frac{C}{2\beta\sqrt{N-1}}. \quad (2.13)$$

Moreover, $v_N(0) = N^{-1} \sum_{i=1}^N \mathbb{E}[|X_0^i - X_0^{i,N}|^2] \leq \mathbb{E}[|X_0^i|^2 + |X_0^{i,N}|^2]$ is bounded uniformly in N , and as $\mathbb{E}[d_2^{\mathbb{R}^d}(m_{\mathbf{X}_t}^N, m_{\mathbf{X}_t^N}^N)] \leq v_N(t)^{1/2}$, we conclude the proof of Theorem 2.12. \square

3.2 Uniqueness of the invariant measure

The main consequence of the previous result is the uniqueness of the invariant measure for the process $(m_t)_{t \geq 0}$ driven by (3.1). As recalled at the beginning of the section, we have already shown that under Assumptions 1 and 2, there exists an invariant measure \bar{P} . From Theorem 2.12, we get the following Corollary

Corollary 2.13. *Under Assumptions 3, the stochastic process (m_t) admits a unique invariant measure $\bar{P} \in \mathcal{P}_2(\mathcal{P}(\mathbb{R}^d))$. Moreover, for each $P_0 \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$, there is an exponential convergence to the*

invariant measure:

$$d_2^{\mathcal{P}(\mathbb{R}^d)}(P_t, \bar{P}) \leq e^{-\beta t} d_2^{\mathcal{P}(\mathbb{R}^d)}(P_0, \bar{P})^2.$$

This implies uniqueness of the invariant measure and the convergence to this equilibria for a large class of initial conditions P_0 . In order to prove the previous result, we begin with a technical Lemma:

Lemma 2.14. *Let m and ρ be two probability measures valued random variables which are \mathcal{F}_0^0 -measurable. Then, there exists a random variable ξ defined on the space $(\Omega, \mathcal{F}, \mathbb{P})$ and with value in $\mathcal{P}(\mathbb{R}^d)$, such that almost surely:*

$$\xi \in \arg \min_{\pi \in \Pi(m, \rho)} \int_{\mathbb{R}^d} |x - y|^2 \pi(dx, dy).$$

The proof of this Lemma is postponed to Appendix 5.3 and relies on mesurability arguments for set valued functions issued from [SV79].

Proof of Corollary 2.13. The proof of this result relies on the result and the proof of Theorem 2.12, but the important difference is the choice of the initial conditions. More precisely, for $P_0 \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$, we pick $\Gamma \in \Pi(P_0, \bar{P})$ which is not empty. Let us consider a couple of probability measure valued random variables (m_0, \bar{m}_0) , and such that $\mathcal{L}((m_0, \bar{m}_0)) = \Gamma$. It means that $\mathcal{L}(m_0) = P_0$ and $\mathcal{L}(\bar{m}_0) = \bar{P}_0$. Thanks to Lemma 2.14, we know that there exists ξ random variable such that almost surely,

$$\xi \in \arg \min_{\pi \in \Pi(m_0, \bar{m}_0)} \int_{\mathbb{R}^d} |x - y|^2 \pi(dx, dy).$$

We consider once again the particle system:

$$dX_t^i = -\nabla(X_t^i)dt - \nabla W * m_t(X_t^i)dt + \sigma dB_t^i + \sigma_0 dB_t^0, \quad \forall i \in \{1, \dots, N\},$$

and

$$dX_t^{i,N} = -\nabla(X_t^{i,N})dt - \frac{1}{N} \sum_{j=1}^N \nabla W(X_t^{i,N} - X_t^{j,N})dt + \sigma dB_t^i + \sigma_0 dB_t^0, \quad \forall i \in \{1, \dots, N\},$$

where the $(X_0^i, X_0^{i,N})$ for $i \in \{1, \dots, N\}$ are independent and such that $\mathcal{L}((X_0^i, X_0^{i,N}) | \mathcal{F}_0^0) = \xi$. From Theorem 2.12, and more precisely Equation (2.13), we know that there exists a constant $C > 0$, such that:

$$\mathbb{E}[d_2^{\mathbb{R}^d}(m_{\mathbf{X}_t}^N, m_{\mathbf{X}_t^N}^N)^2] \leq e^{-\beta t} \mathbb{E}[|X_0^1 - X_0^{1,N}|^2] + CN^{-1/2}. \quad (2.14)$$

Moreover, with the particular choice we made for the initial conditions, we get that

$$\begin{aligned} \mathbb{E}[|X_0^1 - X_0^{1,N}|^2] &= \mathbb{E}[\mathbb{E}[|X_0^1 - X_0^{1,N}|^2 | \mathcal{F}_0^0]] \\ &= \mathbb{E} \left[\int_{\mathbb{R}^d} |x - y|^2 \xi(dx, dy) \right] \\ &= \int_{\mathcal{P}(\mathbb{R}^d)} d_2^{\mathbb{R}^d}(\mu, \nu) \Gamma(d\mu, d\nu). \end{aligned}$$

Taking the infimum over all the transport plan Γ between P_0 and \bar{P} in (2.14), we get that

$$\mathbb{E}\left[d_2^{\mathbb{R}^d}(m_{\mathbf{X}_t}^N, m_{\mathbf{X}_t^N}^N)^2\right] \leq e^{-\beta t} d_2^{\mathcal{P}(\mathbb{R}^d)}(P_0, \bar{P}) + \frac{C}{\sqrt{N}}. \quad (2.15)$$

Moreover,

$$d_2^{\mathcal{P}(\mathbb{R}^d)}(P_t, \bar{P}) \leq \mathbb{E}\left[d_2^{\mathbb{R}^d}(\bar{m}_t, m_t)^2\right], \quad (2.16)$$

where (\bar{m}_t) is the stochastic flow of measure driven by (3.1) with initial condition \bar{P} and (m_t) has the same dynamic with initial condition P_0 . Now combining (2.15) and (2.16), taking the limit N to infinity, we get

$$d_2^{\mathcal{P}(\mathbb{R}^d)}(P_t, \bar{P}) \leq e^{-\beta t} d_2^{\mathcal{P}(\mathbb{R}^d)}(P_0, \bar{P}).$$

□

Remark 2.15. *Some remarks are in order: (1) We recall that thanks to Proposition 2.11, the distance $d_2^{\mathcal{P}(\mathbb{R}^d)}(P_0, \bar{P})$ is finite because $P_0 \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$; (2) Common noise does not ruin the usual convergence results. It makes sense because the common noise mainly acts like a drift term in Equation (3.1) and should not make the usual convergence results go haywire.*

3.3 Example

In this section, we consider an Ornstein-Uhlenbeck process with common noise, that is taking $V : x \mapsto |x|^2/2$ and $W = 0$. Let X be a real valued stochastic process evolving in \mathbb{R}^d , driven by the following stochastic differential equation:

$$dX_t = -X_t dt + \sigma dB_t + \sigma_0 dB_t^0. \quad (2.17)$$

Then, the process associated process $(m_t)_{t \geq 0}$ is solution of

$$d_t m_t = \nabla \cdot \left[\frac{\sigma^2 + \sigma_0^2}{2} \nabla m_t + m_t x \right] dt - \sigma_0 \nabla m_t \cdot dB_t^0, \quad (2.18)$$

Moreover, we consider the following initial condition $\mathcal{L}(X_0) = P_0$ (in the sense of Definition 3.1), for some $P_0 \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$. In this particular setting, we are able to explicitly describe the invariant measure.

Proposition 2.16. *The unique invariant measure \bar{P} of the process on the space $\mathcal{P}(\mathbb{R}^d)$ is the image, by the function $x \in \mathbb{R}^d \mapsto \mathcal{N}_d(-x, \sigma^2 \text{Id}) \in \mathcal{P}(\mathbb{R}^d)$ of the measure γ_{σ_0} ; where $\gamma_{\sigma_0}(dx) = (2\sigma_0^2)^{-1/2} e^{-|x|^2/2\sigma_0^2} dx$, where for all $\mu \in \mathbb{R}^d$, $\Sigma \in \mathcal{M}_d(\mathbb{R})$, $\mathcal{N}_d(\mu, \Sigma)$ denotes a gaussian distribution in dimension d centered in μ and with variance-covariance matrix Σ .*

The fact that in this case we are able to exhibit the invariant measure comes from the linearity of the equation and the linearity of the confinement forces which derives from a quadratic potential, as shown in the proof. The convergence to the equilibria holds at an exponential thanks to Theorem 2.12.

Proof of Proposition 2.16. Let be the process X^0 defined dynamic given by:

$$\begin{cases} dX_t^0 = -X_t^0 dt - \sigma_0 dB_t^0 \\ \mathcal{L}(X_0^0) = \mathcal{N}(0, \sigma_0^2). \end{cases}$$

The initial condition is such that the process X^0 is stationary. We now define the process $(\bar{m}_t)_t$ with \bar{m}_0 , such that $\mathcal{L}(\bar{m}_0) = P$ as initial condition and staying

$$d_t \bar{m}_t = \nabla \cdot \left[\frac{\sigma^2 + \sigma_0^2}{2} \nabla \bar{m}_t + \bar{m}_t x \right] dt - \sigma_0 \nabla \bar{m}_t \cdot dB_t^0. \quad (2.19)$$

Let us define $\tilde{m}_t := (\text{Id} - \sigma^0 B_t^0) \# \bar{m}_t$, where $\#$ stands for the pushforward operator. Then, applying Itô-Wentzell formula, we get that \tilde{m}_t satisfies

$$\begin{cases} \partial_t \tilde{m}_t = \nabla \cdot \left[\frac{\sigma^2}{2} \nabla \tilde{m}_t + \tilde{m}_t (x + B_t^0) \right], \\ \tilde{m}_0 = \bar{m}_0. \end{cases} \quad (2.20)$$

Moreover, we get that $m_t := \mathcal{N}(-(X_t^0 + \sigma_0 B_t^0), \sigma^2 \text{Id})$ satisfies,

$$d_t m_t(x) = -\sigma^{-2} (x + X_t^0 + \sigma_0 B_t^0) m_t(x) d(X_t^0 + \sigma_0 B_t^0) = \sigma^{-2} (x + X_t^0 + \sigma_0 B_t^0) \cdot X_t^0 m_t(x) dt. \quad (2.21)$$

Then, as

$$\frac{1}{2} \sigma_0^2 \Delta m_t + \text{div} (m_t (x + B_t^0)) = -\sigma^{-2} X_t^0 \nabla m_t = \sigma^{-2} X_t^0 \cdot (x + X_t^0 + \sigma_0 B_t^0) m_t(x),$$

it shows that $(m_t)_t$ satisfies (2.20) with $m_0 = \bar{m}_0$ and then $((\text{Id} + \sigma^0 B_t^0) \# m_t)_t$ is solution of (2.19) with the same initial condition. Finally,

$$\bar{m}_t(dx) = (\text{Id} + \sigma^0 B_t^0) \# m_t(dx) = c \exp\{-|x + X_t^0|^2 / 2\sigma^2\} dx.$$

As $(X_t^0)_t$ is stationary in \mathbb{R}^d , $(\bar{m}_t)_t$ is also in $\mathcal{P}(\mathbb{R}^d)$, which shows the first part of the Proposition. The uniqueness of this invariant measure is then given by Corollary 2.13. \square

4 Non-convex potential without idiosyncratic noise

In this section, we consider the case of a non-convex potential V in the particular setting $\sigma = 0$. It turns out that in some specific situations, we can use the presence of interaction to achieve an exponential rate of convergence to the unique invariant measure. More precisely, let us consider m with dynamic given by

$$d_t m_t = \nabla \cdot \left(\frac{\sigma_0^2}{2} \nabla m_t + m_t (\nabla V + \nabla W * m_t) \right) dt - \sigma_0 \nabla m_t \cdot dB_t^0. \quad (2.22)$$

In the following of the section, let us consider that the potential V satisfies Assumption 1:

1. There exists a continuous function $\kappa : [0, +\infty) \rightarrow \mathbb{R}$, such that

$$\liminf_{r \rightarrow +\infty} \kappa(r) > 0,$$

and

$$(\nabla V(x) - \nabla V(y)) \cdot (x - y) \geq \kappa(|x - y|)|x - y|^2.$$

2. ∇V is L_V -Lipschitz continuous.

Moreover, in the sequel, we only consider the particular setting where W is quadratic in order to be able to control the long time dynamic of $(m_t)_t$ with respect to its first moment.

Assumption 4. *There exists $\alpha > 0$, such that $W(x) = \alpha|x|^2/2$.*

4.1 Existence of an invariant measure

This section begins with a discussion on the existence of an invariant measure, which in this case can be described explicitly. This comes from the absence of idiosyncratic noise within our system coupled with the strong convexity of the interaction, which allows the measure process driven by Equation (2.22), to get close to its first moment on a large time scale, a concept further detailed in Lemma 2.19. Then, we see that the variance within the stochastic flow of measure decays to zero. It is then natural that the invariant measure finds its support in Dirac masses, as detailed in the following Proposition.

Proposition 2.17. *Under Assumptions 1 and 4, for α large enough, there exists a unique invariant measure \bar{P} in the sense of Definition 3.1, and this measure is supported by Dirac masses. More precisely,*

$$\bar{P}(dm) = \int_{\mathbb{R}^d} \delta_{\delta_a} m_0(da)$$

where m_0 is the probability measure on \mathbb{R}^d solution of $-\frac{\sigma_0^2}{2}\Delta m_0 - \nabla \cdot (\nabla V m_0) = 0$.

Proof of Proposition 2.17. If $\bar{P}(dm) = \int_{\mathbb{R}^d} \delta_{\delta_a} m_0(da)$, then for all twice differentiable function in the sense of Lions derivatives $F \in \mathcal{C}^2(\mathcal{P}(\mathbb{R}^d), \mathbb{R})$

$$\begin{aligned} I(F) &:= \int_{\mathcal{P}_2(\mathbb{R}^d)} \left[\int_{\mathbb{R}^d} \left(D_m F(m, x) \cdot (-\nabla V(x) - \nabla W * m(x)) + \frac{\sigma_0^2}{2} \operatorname{div}_x D_m F(m, x) \right) m(dx) \right. \\ &\quad \left. + \frac{\sigma_0^2}{2} \int_{\mathbb{R}^{2d}} \operatorname{Tr} [D_{mm}^2 F(m, x, y)] m(dx)m(dy) \right] \bar{P}(dm) \\ &= \int_{\mathbb{R}^d} \left[D_m F(\delta_a, a) \cdot (-\nabla V(a)) + \frac{\sigma_0^2}{2} \operatorname{div}_x D_m F(\delta_a, a) + \frac{\sigma_0^2}{2} \operatorname{Tr} [D_{mm}^2 F(\delta_a, a, a)] \right] m_0(da), \end{aligned} \tag{2.23}$$

where, if we define $\varphi(a) = F(\delta_a)$, then

$$\nabla \varphi(a) = D_m F(\delta_a, a), \quad \Delta \varphi(a) = \operatorname{Tr} [D_{mm}^2 F(\delta_a, a, a) + D_{xm}^2 F(\delta_a, a)].$$

Then,

$$I(F) = \int_{\mathbb{R}^d} \left[-\nabla \varphi(a) \cdot \nabla V(a) + \frac{\sigma_0^2}{2} \Delta \varphi(a) \right] m_0(da) = 0, \tag{2.24}$$

as m_0 is solution of $\frac{\sigma_0^2}{2} \Delta m_0 + \nabla \cdot (\nabla V m_0) = 0$. \square

4.2 Uniqueness of the invariant measure and exponential decay

The previous section shows that for $\sigma = 0$, we can explicitly identify an invariant measure, though its uniqueness is uncertain due to the non-convex nature of the confinement potential. This section establishes that, in the absence of idiosyncratic noise, the invariant measure is indeed unique. More precisely, let us state the following result:

Theorem 2.18. *Whenever $\sigma = 0$ and under Assumptions 3, for any initial conditions $P_0, Q_0 \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$, there exists a constant C that does not depend on t such that*

$$d_1^{\mathcal{P}(\mathbb{R}^d)}(P_t, Q_t) \leq C \left(e^{-\ell \sigma_0^2 t} + e^{-(\alpha - 2L_V)t} \right),$$

for some positive constant ℓ . Moreover, \bar{P} defined in Proposition 2.17 is the unique invariant measure in the sense of Definition 3.4.

This is the main result of this paper, showing is that when $\sigma = 0$, there is only one invariant measure, which contrasts with scenarios lacking common noise where multiple invariant solutions exist as we are in a context with non-convex confinement potential. This theorem reveals that introducing finite dimensional noise to the system uniquely determines the invariant measure, a result that is stronger than previous one in the literature. Earlier studies (see e.g [ABKO23]) showed that adding cylindrical noise could achieve invariant measure uniqueness, but our results go further by showing that even finite dimensional noise is enough for this purpose.

Proposition 2.17 shows that when $\sigma = 0$, we are able to identify the unique invariant measure. Moreover, in Theorem 2.18, we can see that the string interaction structure helps in order to get exponentially fast convergence to this invariant measure. This may seem surprising initially, but the explanation lies in how the interaction term affects convergence rates. Specifically, in contexts of strong interaction, the process (m_t) is significantly drawn towards its first moment on a large time scale. This observation plays a crucial role in proving Theorem 2.18, showing that the law of the infinite-dimensional random measure process m can be equated with the law of its first moment, as detailed in the following lemma.

Lemma 2.19. *Let us consider a probability measure $P_0 \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$. If $(m_t)_t$ defines the stochastic flow of measure solution of (2.22) with initial condition P_0 , we define for all $t > 0$, $\bar{p}_t \in \mathcal{P}_2(\mathbb{R}^d)$ the law of the first moment of m_t , $\bar{m}_t = \int_{\mathbb{R}^d} x m_t(dx)$. Then, we have for some constant C that depends only on P_0*

$$d_1^{\mathcal{P}(\mathbb{R}^d)}(P_t, \delta_{\bar{p}_t}) \leq C e^{-(\alpha - 2C_V)t}.$$

Remark 2.20. *The specific form of the invariant measure, as detailed in Proposition 2.17, showcases a scenario unique to quadratic interactions. However, the insights from Lemma 2.19 holds true even when we tweak the interaction term to something like*

$$(x, \mu) \mapsto \nabla - W \left(x - \int y \mu(dy) \right),$$

provided W is a uniformly convex potential. The critical aspect we are looking at here is not the specific form of the interaction but how it allows us to identify the process with its mean on a large time scale. Then, uniqueness of the invariant measure holds for more general interaction potential.

Proof of Lemma 2.19. The proof of this result is quite straightforward and due to the strong interaction regime. Let us now consider the process X evolving in \mathbb{R}^d with dynamic

$$dX_t = -\nabla V(X_t)dt - \alpha(X_t - \mathbb{E}_{\mathbb{P}^1}[X_t])dt + \sigma_0 dB_t^0.$$

Now, by definition of the Wasserstein distance, we know that

$$d_1^{\mathcal{P}(\mathbb{R}^d)}(P_t, \delta_{\bar{p}_t}) \leq \mathbb{E}[d_1^{\mathbb{R}^d}(m_t, \bar{p}_t)].$$

It is also known that the conditional law of the process X at time $t > 0$ is given by m_t . Then

$$\mathbb{E}[d_1^{\mathbb{R}^d}(m_t, \bar{p}_t)] \leq \mathbb{E}[|X_t - \mathbb{E}_{\mathbb{P}^1}[X_t]|].$$

We are now left with the study of a stochastic process evolving into a finite dimensional space. Writing the dynamic of the process and using the fact that ∇V is Lipschitz continuous, we get

$$\frac{d}{dt} \mathbb{E}[|X_t - \mathbb{E}_{\mathbb{P}^1}[X_t]|^2] \leq -2(\alpha - 2C_V) \mathbb{E}[|X_t - \mathbb{E}_{\mathbb{P}^1}[X_t]|^2].$$

This gives the result applying Grönwall Lemma, we get

$$\mathbb{E}[|X_t - \mathbb{E}_{\mathbb{P}^1}[X_t]|^2] \leq \mathbb{E}[|X_0 - \mathbb{E}_{\mathbb{P}^1}[X_0]|^2] \exp(-2(\alpha - 2C_V)t)$$

Moreover, we can write

$$\mathbb{E}[|X_0 - \mathbb{E}_{\mathbb{P}^1}[X_0]|^2] \leq 2\mathbb{E}[|X_0|^2].$$

This term is finite as P_0 is assumed to admit a finite moment of order 2, which concludes the proof. \square

The aforementioned result shows that over large time scales, the conditional law of the process X can be identified with its first moment. Consequently, to understand the long-term behavior of the solution (m_t) to Equation (2.22), it is enough to study the long-term behavior of

$$\left(\int_{\mathbb{R}^d} x m_t(dx) \right)_{t \geq 0}.$$

In particular, we would like to show that it admits a unique equilibrium. We are then left with the study of the following dynamic:

$$d\mathbb{E}_{\mathbb{P}^1}[X_t] = -\mathbb{E}_{\mathbb{P}^1}[\nabla V(X_t)]dt + \sigma_0 dB_t^0.$$

The dynamic here looks like the one of a classical diffusion process. This similarity is explicit in

the following expression:

$$d\mathbb{E}_{\mathbb{P}^1}[X_t] = -\nabla V(\mathbb{E}_{\mathbb{P}^1}[X_t])dt + \varepsilon_t^V dt + \sigma_0 dB_t^0,$$

where $\varepsilon_t^V := -\mathbb{E}_{\mathbb{P}^1}[\nabla V(X_t)] + \nabla V(\mathbb{E}_{\mathbb{P}^1}[X_t])$. Further, using the Lipschitz continuity of V and with a similar approach as in the proof of Lemma 2.19, we can show that ε_t^V decays exponentially fast to 0 over time. Then, the dynamics of $(\mathbb{E}_{\mathbb{P}^1}[X_t])_{t \geq 0}$ closely look like the one of a diffusion process with confining potential. Following the methodology of [Ebe16], we show that such process exhibits a contraction property in the Wasserstein distance. More precisely, we have the following result:

Proposition 2.21. *Let us consider two probability measures $P_0, Q_0 \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$. Moreover, if $(m_t^P)_t$ and $(m_t^Q)_t$ define the stochastic flows of measure solution of (2.22) with initial condition P_0 and Q_0 respectively, we define for all $t > 0$, \bar{p}_t , and \bar{q}_t the law of the first moment of m_t^P , and m_t^Q respectively. Then, we have for some constant C that depends only on P_0 and Q_0*

$$d_1^{\mathbb{R}^d}(\bar{p}_t, \bar{q}_t) \leq C(e^{-(\alpha-2C_V)t} + e^{-\ell\sigma_0^2 t}).$$

In order to prove Proposition 2.21, let us proceed as in [DEGZ20], and introduce the following quantity:

$$\begin{aligned} R_0 &= \inf \{s \geq 0, \kappa(r) \geq 0, \forall r \geq s\}, \\ R_1 &= \inf \{s \geq R_0, s(s - R_0)\kappa(r) \geq 4\sigma_0^2, \forall r \geq s\}. \end{aligned}$$

Moreover, we consider $\varphi, \Phi, g : [0, +\infty) \rightarrow [0, +\infty)$ defined by

$$\begin{aligned} \varphi(r) &= \exp\left(-\frac{1}{2\sigma_0^2} \int_0^r s\kappa_-(s)ds\right), \\ \Phi(r) &= \int_0^r \varphi(s)ds, \\ g(r) &= 1 - \frac{\ell}{2} \int_0^{r \wedge R_1} \Phi(s)/\varphi(s)ds, \end{aligned}$$

where $\kappa_- = \max(0, -\kappa)$ and $\ell = \left(\int_0^{R_1} \Phi(s)\varphi(s)^{-1}ds\right)^{-1}$. We now define an increasing function $f : [0, +\infty) \rightarrow [0, +\infty)$ by:

$$f(r) = \int_0^r \varphi(s)g(s)ds.$$

The function f that has been constructed is clearly positive, non-decreasing and concave. Moreover, it satisfies

$$\varphi(R_0)r/2 \leq f(r) \leq r. \quad (2.25)$$

This ensures that $(x, y) \mapsto f(|x - y|)$ defines a distance which is equivalent to the Euclidean one. Below, we will use contraction properties in Wasserstein-1 distance based on the underlying distance $f(|x - y|)$. These contraction property is a consequence of Proposition 2.22.

Proposition 2.22. *The following inequality holds for all $r > 0$:*

$$f''(r) - \frac{1}{2\sigma_0^2} r \kappa(r) f'(r) \leq -\ell f(r)/2.$$

The proof of this Proposition is found in Appendix 5.4 for sake of clarity but is exactly the same as in [Ebe16]. However, it is essential to emphasize that the function f was constructed to achieve a contraction inequality, ensuring uniform propagation of chaos over time. These techniques were greatly inspired by [LR86] and further developed in [Ebe16]. Let us now consider Let us also consider $\delta > 0$, and define X^δ and Y^δ with common dynamic given by:

$$dX_t^\delta = -\nabla V(X_t^\delta)dt - \alpha(X_t^\delta - \mathbb{E}_{\mathbb{P}_1}[X_t^\delta])dt + \sigma_0\{\pi_\delta(E_t^\delta)dB_t^0 + \lambda_\delta(E_t^\delta)d\tilde{B}_t^0\}$$

and

$$dY_t^\delta = -\nabla V(Y_t^\delta)dt - \alpha(Y_t^\delta - \mathbb{E}_{\mathbb{P}_1}Y_t^\delta)dt + \sigma_0\{(\text{Id} - 2e_t^\delta e_t^{\delta T})\pi_\delta(E_t^\delta)dB_t^0 + \lambda_\delta(E_t^\delta)d\tilde{B}_t^0\}.$$

where

- B^0 and \tilde{B}^0 are independent Brownian motions adapted to \mathbb{F}^0
- For all $t \geq 0$ and $\delta > 0$, we denote

$$E_t^\delta = \mathbb{E}_{\mathbb{P}_1}[X_t^\delta - Y_t^\delta];$$

- For all $t \geq 0$ and $\delta > 0$,

$$e_t^\delta = \begin{cases} E_t^\delta/|E_t^\delta| & \text{if } |E_t^\delta| \neq 0, \\ 0 & \text{else.} \end{cases}$$

- We define a non-decreasing and continuous function π , such that for $x \in \mathbb{R}^N$,

$$\pi(x) = \begin{cases} 1 & \text{if } |x| \geq 1 \\ 0 & \text{if } |x| \leq 1/2, \end{cases}$$

and, consider a non negative function λ such that

$$\pi(x)^2 + \lambda(x)^2 = 1, \quad \forall x \in \mathbb{R}^N.$$

Moreover, we extend π on the whole space, with the constraint that this remains a non decreasing and Lipschitz continuous function. Finally, we define $\pi_\delta : x \mapsto \pi(x/\delta)$.

The coupling just introduced is known as *coupling by reflection*. However, its application in this context is not immediately intuitive. Particularly intriguing is the reflection performed on the common noise component. However, the forthcoming Lemma proves that such a reflection on the common noise does not affect the properties of the conditional law.

Lemma 2.23. *There exist two d -dimensional Brownian motions (β_t^0) and $(\tilde{\beta}_t^0)$ adapted to the filtration \mathbb{F}^0 such that if we define for all $t \geq 0$, $m_t^{X,\delta} = \mathcal{L}^1(X_t^\delta)$ and $m_t^{Y,\delta} = \mathcal{L}^1(Y_t^\delta)$, we have*

$$d_t m_t^{X,\delta} = \nabla \cdot \left(\frac{\sigma_0^2}{2} \nabla m_t^{X,\delta} + m_t^{X,\delta} (\nabla V + \alpha(m_t^{X,\delta} - \int_{\mathbb{R}^d} x m_t^{X,\delta}(dx))) \right) dt - \sigma_0 \nabla m_t \cdot d\beta_t^0,$$

and

$$d_t m_t^{Y,\delta} = \nabla \cdot \left(\frac{\sigma_0^2}{2} \nabla m_t^{Y,\delta} + m_t^{Y,\delta} (\nabla V + \alpha(m_t^{Y,\delta} - \int_{\mathbb{R}^d} x m_t^{Y,\delta}(dx))) \right) dt - \sigma_0 \nabla m_t \cdot d\tilde{\beta}_t^0.$$

The proof of this lemma is quite straightforward and then postponed to the Appendix 5.5. This is quite natural because we only reflect with respect to something that is measurable with respect to the common noise, see [CD18b, Thm 4.14] for further details.

Now that we defined the coupling and that Lemma 2.23 ensures that this coupling gives the expected Equation, for any $\delta > 0$, we can show the following proposition:

Proposition 2.24. *For any $\delta > 0$ and $t > 0$,*

$$\begin{aligned} d|\mathbb{E}_{\mathbb{P}^1}[X_t^\delta] - \mathbb{E}_{\mathbb{P}^1}[Y_t^\delta]| &= -e_t^\delta \cdot \left(\mathbb{E}_{\mathbb{P}^1}[\nabla V(X_t^\delta)] - \mathbb{E}_{\mathbb{P}^1}[\nabla V(Y_t^\delta)] \right) dt \\ &\quad + 2\sigma_0 \pi_\delta(E_t^\delta) (e_t^\delta)^T dB_t^0. \end{aligned} \quad (2.26)$$

Proof. Using Itô's formula, we get that for any $\delta > 0$,

$$\begin{aligned} d|\mathbb{E}_{\mathbb{P}^1}[X_t^\delta] - \mathbb{E}_{\mathbb{P}^1}[Y_t^\delta]|^2 &= -2(\mathbb{E}_{\mathbb{P}^1}[X_t^\delta] - \mathbb{E}_{\mathbb{P}^1}[Y_t^\delta]) \cdot (\mathbb{E}_{\mathbb{P}^1}[\nabla V(X_t^\delta)] - \mathbb{E}_{\mathbb{P}^1}[\nabla V(Y_t^\delta)]) dt \\ &\quad + 4\sigma_0(\mathbb{E}_{\mathbb{P}^1}[X_t^\delta] - \mathbb{E}_{\mathbb{P}^1}[Y_t^\delta]) \pi_\delta(E_t^\delta) e_t^\delta (e_t^\delta)^T dB_t^0 \\ &\quad + 4\sigma_0^2 \pi_\delta(E_t^\delta)^2 (e_t^\delta)^T dt. \end{aligned}$$

For any $\varepsilon > 0$, let us introduce $\psi_\varepsilon : [0, +\infty) \ni r \mapsto (r + \varepsilon)^{1/2}$. This function is continuously twice differentiable, then we can write

$$\begin{aligned} d\psi_\varepsilon(|\mathbb{E}_{\mathbb{P}^1}[X_t^\delta] - \mathbb{E}_{\mathbb{P}^1}[Y_t^\delta]|^2) &= -2\psi'_\varepsilon(|E_t^\delta|^2) E_t^\delta \cdot (\mathbb{E}_{\mathbb{P}^1}[\nabla V(X_t^\delta)] - \mathbb{E}_{\mathbb{P}^1}[\nabla V(Y_t^\delta)]) dt \\ &\quad + 4\sigma_0 \psi'_\varepsilon(|E_t^\delta|^2) E_t^\delta \pi_\delta(E_t^\delta) e_t^\delta (e_t^\delta)^T dB_t^0 \\ &\quad + 4\sigma_0^2 \psi'_\varepsilon(|E_t^\delta|^2) \pi_\delta(E_t^\delta)^2 (e_t^\delta)^T dt \\ &\quad + 8\sigma_0^2 \psi''_\varepsilon(|E_t^\delta|^2) |E_t^\delta|^2 \psi_\delta(E_t^\delta) (e_t^\delta)^T dt. \end{aligned}$$

We now want to take the limit $\varepsilon \rightarrow 0$. Using dominated convergence theorem and stochastic dominated convergence theorem as stated in [RY99], combined to the fact that $4r\psi'_\varepsilon(r^2) \leq 1$, we can deal with the first two lines and get that for all $t \geq 0$

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \int_0^t 2\psi'_\varepsilon(|E_s^\delta|^2) E_s^\delta \cdot (\mathbb{E}_{\mathbb{P}^1}[\nabla V(X_s^\delta)] - \mathbb{E}_{\mathbb{P}^1}[\nabla V(Y_s^\delta)]) ds \\ = \int_0^t e_s^\delta \cdot (\mathbb{E}_{\mathbb{P}^1}[\nabla V(X_s^\delta)] - \mathbb{E}_{\mathbb{P}^1}[\nabla V(Y_s^\delta)]) ds \end{aligned} \quad (2.27)$$

and

$$\lim_{\varepsilon \rightarrow 0} \int_0^t 4\sigma_0 \psi'_\varepsilon(|E_s^\delta|^2) E_s^\delta \pi_\delta(E_s^\delta) e_s^\delta (e_s^\delta)^T dB_s^0 = \int_0^t 4\sigma_0 \pi_\delta(E_s^\delta) (e_s^\delta)^T dB_s^0 \quad (2.28)$$

For the two last, we need to take advantage of the presence of the function π_δ for $\delta > 0$. In fact, we have

$$\begin{aligned} & \left| 4\sigma_0^2 \psi'_\varepsilon(|E_t^\delta|^2) \pi_\delta(E_t^\delta)^2 (e_t^\delta)^T + 8\sigma_0^2 \psi''_\varepsilon(|E_t^\delta|^2) |E_t^\delta|^2 \psi_\delta(E_t^\delta) (e_t^\delta)^T \right| \\ & \leq \left| \pi_\delta(E_t^\delta)^2 \sigma_0^2 \left(4\psi'_\varepsilon(|E_t^\delta|^2) + 8\psi''_\varepsilon(|E_t^\delta|^2) |E_t^\delta|^2 \right) \right|. \end{aligned}$$

Moreover, we know that $\psi'_\varepsilon(r^2) + 2\psi''_\varepsilon(r^2)r^2 \leq r^{-3}$, for all $r > 0$ and $\varepsilon \leq 1$. Using the presence of π_δ , we have that the integrand is null near 0. Then, we can once again apply dominated convergence theorem and obtain

$$\lim_{\varepsilon \rightarrow 0} \int_0^t 4 \left\{ \sigma_0^2 \psi'_\varepsilon(|E_s^\delta|^2) \pi_\delta(E_s^\delta)^2 (e_s^\delta)^T + 8\sigma_0^2 \psi''_\varepsilon(|E_s^\delta|^2) |E_s^\delta|^2 \psi_\delta(E_s^\delta) (e_s^\delta)^T \right\} ds = 0. \quad (2.29)$$

Finally, combining Equations (3.61), (3.63) and (3.65), we get the expected result. \square

Using the results of Proposition 2.24, we are ready for the proof of Proposition 2.21:

Proof of Proposition 2.21. From Proposition 2.21 and using Itô formula to the function f , we obtain

$$\begin{aligned} df(|E_t^\delta|) &= -f'(|E_t^\delta|) e_t^\delta \cdot \left(\mathbb{E}_{\mathbb{P}_1}[\nabla V(X_t^\delta)] - \mathbb{E}_{\mathbb{P}_1}[\nabla V(Y_t^\delta)] \right) dt \\ &\quad + 2\sigma_0 f'(|E_t^\delta|) \pi_\delta(E_t^\delta) (e_t^\delta)^t dB_t^0 \\ &\quad + 2\sigma_0^2 f''(|E_t^\delta|) \pi_\delta(E_t^\delta)^2 dt. \end{aligned}$$

To control the first term, we carry out the following decomposition

$$\begin{aligned} -f'(|E_t^\delta|) e_t^\delta \cdot \left(\mathbb{E}_{\mathbb{P}_1}[\nabla V(X_t^\delta)] - \mathbb{E}_{\mathbb{P}_1}[\nabla V(Y_t^\delta)] \right) &= -f'(|E_t^\delta|) e_t^\delta \cdot \left(\mathbb{E}_{\mathbb{P}_1}[\nabla V(X_t^\delta)] - \nabla V(\mathbb{E}_{\mathbb{P}_1}[X_t^\delta]) \right) \\ &\quad - f'(|E_t^\delta|) e_t^\delta \cdot \left(\nabla V(\mathbb{E}_{\mathbb{P}_1}[X_t^\delta]) - \nabla V(\mathbb{E}_{\mathbb{P}_1}[Y_t^\delta]) \right) \\ &\quad - f'(|E_t^\delta|) e_t^\delta \cdot \left(\mathbb{E}_{\mathbb{P}_1}[\nabla V(X_t^\delta)] - \nabla V(\mathbb{E}_{\mathbb{P}_1}[X_t^\delta]) \right). \end{aligned}$$

Taking the expectation and leveraging on the fact that f' is globally bounded, the existence of $C_1 > 0$, independent of t and δ , such that

$$\begin{cases} -f'(|E_t^\delta|) e_t^\delta \cdot \left(\mathbb{E}_{\mathbb{P}_1}[\nabla V(X_t^\delta)] - \nabla V(\mathbb{E}_{\mathbb{P}_1}[X_t^\delta]) \right) \leq C_1 \mathbb{E} [|X_t^\delta - \mathbb{E}_{\mathbb{P}_1}[X_t^\delta]|] \\ -f'(|E_t^\delta|) e_t^\delta \cdot \left(\mathbb{E}_{\mathbb{P}_1}[\nabla V(X_t^\delta)] - \nabla V(\mathbb{E}_{\mathbb{P}_1}[X_t^\delta]) \right) \leq C_1 \mathbb{E} [|X_t^\delta - \mathbb{E}_{\mathbb{P}_1}[X_t^\delta]|] \end{cases}$$

Moreover, using Proposition 2.19, we obtain

$$\begin{aligned} & -\mathbb{E} \left[f'(|E_t^\delta|) e_t^\delta \cdot \left(\mathbb{E}_{\mathbb{P}_1}[\nabla V(X_t^\delta)] - \mathbb{E}_{\mathbb{P}_1}[\nabla V(Y_t^\delta)] \right) \right] \\ & \leq C_2 e^{-(\alpha-2L_V)t} - \mathbb{E} \left[f'(|E_t^\delta|) e_t^\delta \cdot \left(\nabla V(\mathbb{E}_{\mathbb{P}_1}[X_t^\delta]) - \nabla V(\mathbb{E}_{\mathbb{P}_1}[Y_t^\delta]) \right) \right]. \end{aligned}$$

Now, using the contraction property of f stated in Proposition 2.22, we get

$$\frac{d}{dt} \mathbb{E}[f(|E_t^\delta|)] \leq C_1 e^{-(\alpha-2L_V)t} - \ell\sigma_0^2 f(|E_t^\delta|) + \ell\sigma_0^2 \delta + |\kappa_-|_\infty \delta. \quad (2.30)$$

Now, letting $\delta \rightarrow 0$, we obtain:

$$\frac{d}{dt} \mathbb{E}[f(|E_t|)] \leq C_2 e^{-(\alpha-2L_V)t} - \ell\sigma_0^2 f(|E_t|).$$

Using Grönwall Lemma in Equation (2.30), we get

$$\mathbb{E}[f(|\mathbb{E}_{\mathbb{P}^1}[X_t] - \mathbb{E}_{\mathbb{P}^1}[Y_t|)])] \leq C_3(e^{-(\alpha-2L_V)t} + e^{-\ell\sigma_0^2 t}),$$

for some C_3 that only depends on P_0 and Q_0 . Finally,

$$d_f^{\mathbb{R}^d}(\bar{p}_t, \bar{q}_t) \leq \mathbb{E}[|\mathbb{E}_{\mathbb{P}^1}[X_t] - \mathbb{E}_{\mathbb{P}^1}[Y_t|)]|] \leq C_4(e^{-(\alpha-2L_V)t} + e^{-\ell\sigma_0^2 t}).$$

Thanks to Equation (2.25), we know that $d_f^{\mathbb{R}^d}$ and $d_1^{\mathbb{R}^d}$ are equivalent, and we have the expected result:

$$d_1^{\mathbb{R}^d}(\bar{p}_t, \bar{q}_t) \leq C(e^{-(\alpha-2L_V)t} + e^{-\ell\sigma_0^2 t}),$$

for some $C > 0$. Moreover, $C < +\infty$ as soon as

$$d_1^{\mathbb{R}^d}(\bar{p}_0, \bar{q}_0) < +\infty,$$

which is the case since P_0 and Q_0 admit a finite moment of order 2. \square

The proof of Theorem 2.18 is now quite straightforward. In fact, we have shown in Lemma 2.19 that we can identify on a large time scale the process of interest $(m_t)_t$ solution of Equation (2.22) with its first moment. Then, in Proposition 2.21, we proved that the dynamic of the latter admits a unique equilibrium, which shows the expected result.

Proof of Theorem 2.18. Taking advantage of the previous sections, we can perform the following decomposition:

$$d_1^{\mathcal{P}(\mathbb{R}^d)}(P_t, Q_t) \leq d_1^{\mathcal{P}(\mathbb{R}^d)}(P_t, \delta_{\bar{p}_t}) + d_1^{\mathcal{P}(\mathbb{R}^d)}(\delta_{\bar{p}_t}, \delta_{\bar{q}_t}) + d_1^{\mathcal{P}(\mathbb{R}^d)}(\delta_{\bar{q}_t}, Q_t).$$

From Lemma 2.19, we immediately get

$$d_1^{\mathcal{P}(\mathbb{R}^d)}(P_t, \delta_{\bar{p}_t}) + d_1^{\mathcal{P}(\mathbb{R}^d)}(\delta_{\bar{q}_t}, Q_t) \leq C e^{-(\alpha+2L_V)t}.$$

Moreover, for the remaining term, one can write that $d_1^{\mathcal{P}(\mathbb{R}^d)}(\delta_{\bar{p}_t}, \delta_{\bar{q}_t}) \leq d_1^{\mathbb{R}^d}(\bar{p}_t, \bar{q}_t)$. Then, one can conclude to the expected result using the result of Proposition 2.21, we can conclude the proof. \square

5 Appendix

5.1 Proof of Proposition 2.8

Equation (2.9) comes from Section 1.5 in [LSZ20] (Equation (1.15)). Hence, if (m_t) is an invariant measure in the sense of Definition 3.4, then \bar{P} straightforwardly is a solution of Equation (2.10). Conversely, let us consider a probability measure \bar{P} , solution of Equation (2.10). More precisely, for all $F \in \mathcal{C}_b^2(\mathcal{P}(\mathbb{R}^d))$

$$\langle \bar{P}, \mathcal{M}F \rangle = 0.$$

Let us consider $T > 0$, then

$$\begin{aligned} & \int_{\mathcal{P}(\mathbb{R}^d)} \|\nabla V + \nabla W * m\|_{L^2(m)}^2 \bar{P}(dm) \\ & \leq C \left(1 + \int_{\mathcal{P}(\mathbb{R}^d)} \int_{\mathbb{R}^d} |\nabla V(x)|^2 m(dx) \bar{P}(dm) + \int_{\mathcal{P}(\mathbb{R}^d)} \int_{\mathbb{R}^d} |\nabla W * m(x)|^2 m(dx) \bar{P}(dm) \right) \\ & \leq C \left(1 + \int_{\mathcal{P}(\mathbb{R}^d)} \int_{\mathbb{R}^d} |x|^2 m(dx) \bar{P}(dm) + \int_{\mathcal{P}(\mathbb{R}^d)} |\nabla W * m(0)|^2 \bar{P}(dm) \right) \\ & \leq C \left(1 + \int_{\mathcal{P}(\mathbb{R}^d)} \int_{\mathbb{R}^d} |x|^2 m(dx) \bar{P}(dm) \right), \end{aligned}$$

for some constant C that may change from line to line and depends only on the Lipschitz constant of ∇V and ∇W . Then using Proposition 2.10, we get that

$$\int_{\mathcal{P}(\mathbb{R}^d)} \|\nabla V + \nabla W * m\|_{L^2(m)}^2 \bar{P}(dm) < +\infty.$$

Applying Theorem 1.5 in [LSZ20], we get the existence of a process $(\mu_t) \in \mathcal{C}([0, T], \mathcal{P}(\mathbb{R}^d))$ such that μ has dynamic given by (3.1) and for all $t \in [0, T]$, $\mathcal{L}(\mu_t) = \bar{P}$. Then by weak uniqueness of the solutions of (3.1), we get that \bar{P} is an invariant measure for m in the sense of Definition 3.4.

5.2 Proof of Lemma 2.9

We consider the process (X_t) , driven by the dynamic (3.5) and with initial condition X_0 such that $\mathcal{L}(X_0) = P_0$, in the sense of Definition 3.1. Let us denote in the following $m_2(t) = \mathbb{E}[|X_t|^2]$. Then expanding using Ito formula and taking the time derivative, gives:

$$\begin{aligned} m_2'(t) &= -2\mathbb{E}[X_t \cdot (\nabla V(X_t) + \nabla W * m_t(X_t))] + (\sigma^2 + \sigma_0^2)d \\ &= -2\mathbb{E}[X_t \cdot (\nabla V(X_t) - \nabla V(0))] - 2\nabla V(0)\mathbb{E}[X_t] - 2\mathbb{E}[X_t \cdot \nabla W * m_t(X_t)] + (\sigma_0^2 + \sigma^2)d. \end{aligned}$$

Moreover, if \tilde{X}_t an independent copy of X_t , then:

$$\begin{aligned} \nabla W * m_t(X_t) &= \int_{\mathbb{R}^d} \nabla W(X_t - y) m_t(dy) \\ &= \mathbb{E}[\nabla W(X_t - \tilde{X}_t) | X_t, \mathcal{F}_0^0]. \end{aligned}$$

Now, the fact that W is even gives $2\mathbb{E}[X_t \nabla W * m_t(X_t)] = \mathbb{E}[(X_t - \tilde{X}_t) \cdot \nabla W(X_t - \tilde{X}_t)]$. This decomposition is the key, the end of the proof is straightforward using Assumption made on the potential W (see Assumption 2).

5.3 A measurable selection result

In this part, we will mainly prove Lemma 2.14:

Lemma 2.25. *Let m and ρ be two probability measure valued random variables which are \mathcal{F}_0^0 -measurable. Assuming that both random measures m and ρ admit a moment of order two almost surely, then there exists a random variable ξ , such that almost surely:*

$$\xi \in \arg \min_{\pi \in \Pi(m, \rho)} \int_{\mathbb{R}^d} |x - y|^2 \pi(dx, dy).$$

Proof. The purpose of the proof is to show that there exists a measurable function $\phi : \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathcal{P}(\mathbb{R}^d)$, such that for all $(m, \rho) \in \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathcal{P}(\mathbb{R}^d)$, $\phi(m, \rho) \in \Pi_{\text{opt}}(m, \rho)$, the set of minimizers for the transport problem. First of all, it is straightforward that Π_{opt} is never empty. Then, we need to show that the set valued function

$$\Phi : \begin{array}{l} \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d) \rightarrow 2^{\mathcal{P}(\mathbb{R}^d)} \\ (m, \rho) \mapsto \Pi(m, \rho), \end{array}$$

is measurable, where 2^A stands for the set of subsets of A . Considering the graph Γ_Φ , of Φ :

$$\Gamma_\Phi = \{((m, \rho), \xi), \xi \in \Phi(m, \rho)\},$$

it is clear that it is a closed set, and then measurable. In particular, the multi-application Φ is measurable. Moreover, Lemma 12.1.7 in [SV79], the application Ψ which associate to any compact set $K \in \mathcal{P}(\mathbb{R})^2$ the set

$$\Psi(K) = \arg \inf_{\pi \in K} \int_{\mathbb{R}^2} |x - y| \pi(dx, dy)$$

is measurable. Then the function

$$\Lambda : \begin{array}{l} \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d) \rightarrow \text{Ens}(\mathcal{P}(\mathbb{R}^d)) \\ (m, \rho) \mapsto \Pi_{\text{opt}}(m, \rho), \end{array}$$

is a compound of two mesurables functions. Finally by the measurable selection theorem, there exists a measurable function ϕ such that for all $(m, \rho) \in \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d)$, $\phi(m, \rho) \in \Pi_{\text{opt}}(m, \rho)$. \square

5.4 Proof of Proposition 2.22

This proof is given in [Ebe16], we repeat it here for the reader convenience. We begin with the easy case, which is whenever $r < R_1$. In fact in this case we have:

$$\begin{aligned} f''(r) &= \varphi'(r)g(r) + \varphi(r)g'(r) \\ &= -\frac{1}{2(\sigma^2 + \sigma_0^2)}r\kappa_-(r)f(r) - \frac{\ell}{2}\Phi(r) \\ &\leq -\frac{1}{2(\sigma^2 + \sigma_0^2)}r\kappa_-(r)f(r) - \frac{\ell}{2}f(r). \end{aligned}$$

This gives:

$$\begin{aligned} f''(r) - \frac{1}{2(\sigma^2 + \sigma_0^2)}r\kappa(r)f'(r) &\leq -\frac{1}{2(\sigma^2 + \sigma_0^2)}rf(r)(\kappa_-(r) + \kappa(r)) - \frac{\ell}{2}f(r) \\ &\leq -\frac{\ell}{2}f(r), \end{aligned}$$

because $\kappa_-(r) - \kappa(r) \geq 0$. The case where $r \geq R_1$ is more intricate. We can easily see that $f''(r) = 0$, and

$$\begin{aligned} -\frac{1}{2(\sigma^2 + \sigma_0^2)}r\kappa(r)f'(r) &\leq -\frac{1}{4(\sigma^2 + \sigma_0^2)}r\kappa(r)\varphi(R_0) \\ &\leq -r\varphi(R_0)(R_1(R_1 - R_0))^{-1}. \end{aligned} \tag{2.31}$$

We easily show that the function $r \mapsto r/\Phi(r)$ is non-decreasing on $[R_1, +\infty[$. Then, it comes that $rR_1^{-1} \leq \Phi(r)\Phi(R_1)^{-1}$, which gives:

$$-\frac{1}{2(\sigma^2 + \sigma_0^2)}r\kappa(r)f'(r) \leq -\varphi(R_0)\Phi(r)(\Phi(R_1)(R_1 - R_0))^{-1}.$$

Moreover,

$$\Phi(r) = \int_0^r \varphi(s)ds = (r - R_0)\varphi(R_0) + \Phi(R_0),$$

and

$$\begin{aligned} \int_{R_0}^{R_1} \Phi(s)\varphi(s)^{-1}ds &= \Phi(R_0)\varphi^{-1}(R_0)(R_1 - R_0) + \frac{1}{2}(R_1 - R_0)^2 \\ &= (R_1 - R_0)\varphi(R_0)^{-1} \left(\Phi(R_0) + \frac{1}{2}(R_1 - R_0)\varphi(R_0) \right) \\ &\geq (R_1 - R_0)\Phi(R_1)\varphi(R_0)^{-1}/2. \end{aligned}$$

Finally,

$$\begin{aligned} -\varphi(R_0)\Phi(r)(\Phi(R_1)(R_1 - R_0))^{-1} &\leq -\frac{1}{2}\Phi(r) \left(\int_{R_0}^{R_1} \Phi(s)\varphi(s)^{-1}ds \right)^{-1} \\ &\leq -\frac{\ell}{2}f(r), \end{aligned}$$

because $\Phi(r) \geq f(r)$ for all $r \geq 0$;

5.5 Proof of Lemma 2.23

Let us consider a bounded and twice differentiable function $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$, using Itô's Lemma

$$\begin{aligned} d\phi(X_t) &= \nabla\phi(X_t)dX_t + \frac{1}{2}\Delta\phi(X_t)d\langle X \rangle_t \\ &= -\nabla\phi(X_t)\nabla V(X_t)dt - \nabla\phi(X_t)\nabla W(X_t - \mathbb{E}_1[X_t])dt \\ &\quad + \sigma_0\nabla\phi(X_t)\{\pi_\delta(E_t)dB_t^0 + \lambda_\delta(E_t)d\tilde{B}_t^0\} \\ &\quad + \frac{\sigma_0^2}{2}\Delta\phi(X_t)dt. \end{aligned}$$

Now, taking the expectation under \mathbb{P}^1 , we have

$$\begin{aligned} d\mathbb{E}_{\mathbb{P}^1}[\phi(X_t)] &= -\mathbb{E}_{\mathbb{P}^1}[\nabla\phi(X_t)\nabla V(X_t)]dt - \mathbb{E}_{\mathbb{P}^1}[\nabla\phi(X_t)\nabla W(X_t - \mathbb{E}_1[X_t])]dt \\ &\quad + \sigma_0\mathbb{E}_{\mathbb{P}^1}[\nabla\phi(X_t)]\{\pi_\delta(E_t)dB_t^0 + \lambda_\delta(E_t)d\tilde{B}_t^0\} \\ &\quad + \frac{\sigma_0^2}{2}\mathbb{E}_{\mathbb{P}^1}[\Delta\phi(X_t)]dt. \end{aligned}$$

This is quite important here to notice that everything comes from the fact that we perform the reflection with a term that is measurable with respect to the *common noise* part. Then, performing integration by part we get the expected result.

Chapter 3

Ergodicity of some stochastic Fokker-Planck equations with additive common noise

Abstract

In this paper we consider stochastic Fokker-Planck Partial Differential Equations (PDEs), obtained as the mean-field limit (i.e., as the number of particles tends to ∞) of weakly interacting particle systems subjected to both independent (or idiosyncratic) and common Brownian noises. We provide sufficient conditions under which the deterministic counterpart of the Fokker-Planck equation, which corresponds to particle systems that are just subjected to independent noises, has several invariant measures, but for which the stochastic version admits a unique invariant measure under the presence of the additive common noise. The very difficulty comes from the fact that the common noise is just of finite dimension while the state variable, which should be seen as the conditional marginal law of the system given the common noise, lives in a space of infinite dimension. In this context, our result holds true if, in addition to standard confining properties, the mean field interaction term forces the system to be attracted by its conditional mean given the common noise and the intensity of the idiosyncratic noise is small.

Contents

1	Introduction	77
1.1	Deterministic Fokker-Planck equations	77
1.2	Common noise	78
1.3	Our contribution	79
1.4	Further connections with the literature on mean field models	80
1.5	Organisation of the paper	80
1.6	Notation	80
2	Main result: uniqueness recovery thanks to common noise	82
2.1	Stochastic Fokker-Planck PDE and Conditional McKean-Vlasov SDE	82
2.2	Assumptions, existence and uniqueness of solutions	83
2.3	Restoration of uniqueness	85
2.4	Discussions and examples	91
3	Proof of Theorem 3.8	95

3.1	First preliminary estimate: conditional variance	95
3.2	Second preliminary estimate: conditional exponential moments	97
3.3	Completion of the proof of Theorem 3.8	99
4	Proof of Proposition 3.7	100
4.1	Ansatz	100
4.2	Contraction for classical diffusion processes	100
4.3	Proof of Proposition 3.7	102
5	Appendix	112
5.1	Proof of Equation (3.38)	112
5.2	Proof of Equation (3.55)	113

1 Introduction

This paper addresses the long-term behavior of solutions to a class of non-linear Stochastic Partial Differential Equations (SPDEs) of the form

$$d_t m_t = \nabla \cdot \left(\frac{\sigma^2 + \sigma_0^2}{2} \nabla m_t - m_t b(\cdot, m_t) \right) dt - \sigma_0 \nabla m_t \cdot dB_t^0, \quad \text{on } [0, +\infty) \times \mathbb{R}^d, \quad (3.1)$$

where $B^0 = (B_t^0)_{t \geq 0}$ is a d -dimensional Brownian motion, called *common noise*. Above, the unknown $(m_t)_{t \geq 0}$ is regarded as a stochastic process with values in the space $\mathcal{P}(\mathbb{R}^d)$ of probability measures on \mathbb{R}^d . Accordingly, the coefficient b is a function from $\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$ to \mathbb{R}^d depending on both space and measure arguments. The two parameters σ and σ_0 are non-negative scalars: the former is the intensity of the so-called idiosyncratic noise and the latter is the intensity of the common noise.

Equation (3.1) is in fact intended to describe the flow of conditional marginal laws of the solution to the conditional McKean-Vlasov equation:

$$dX_t = b(X_t, \mathcal{L}(X_t|B^0))dt + \sigma dB_t + \sigma_0 dB_t^0, \quad t \geq 0, \quad (3.2)$$

where B is another Brownian motion, independent of B^0 , and $\mathcal{L}(\cdot|B^0)$ stands for the conditional law given the realization of the common noise. Formally, m_t in (3.1) is expected to coincide with $\mathcal{L}(X_t|B^0)$ in (3.2). The rigorous connection between (3.1) and (3.2) is addressed more carefully in the core of the paper.

1.1 Deterministic Fokker-Planck equations

When $\sigma_0 = 0$, i.e., in absence of common noise, Equation (3.1) boils down to a standard Fokker-Planck equation describing the evolution of the marginal laws of the solutions to the standard McKean-Vlasov equation (3.2) obtained by replacing $\mathcal{L}(X_t|B^0)$ by the law of X_t . In this setting, the long-time analysis of the solutions has been a notoriously challenging problem for over twenty years, giving rise to numerous contributions, many of them still recent. While the problem is of mathematical interest in its own right, it also finds some application to calculus of variation since certain equations like (3.1) can be interpreted as gradient flows on the space of probability measures, see the seminal works [AGS08, CMV03, JKO98, Ott99, Ott01]. For instance, such a gradient structure occurs when $b(x, \mu) = -\nabla V(x) - \nabla W * \mu(x)$, for two differentiable real-valued functions V and W on \mathbb{R}^d , W being symmetric, where $*$ stands for the convolution operator. In this case, the potential lying above the dynamics writes

$$\mu \in \mathcal{P}(\mathbb{R}^d) \mapsto \int_{\mathbb{R}^d} V(x) \mu(dx) + \frac{1}{2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} W(x-y) \mu(dx) \mu(dy) + \frac{\sigma^2}{2} \int_{\mathbb{R}^d} \ln\left(\frac{d\mu}{dx}(x)\right) \mu(dx). \quad (3.3)$$

In fact, due to the mean field interaction (whether it derives from a potential or not), the sole presence of the noise B in the dynamics (3.2) (with $\sigma_0 = 0$) does not suffice to guarantee the uniqueness of an invariant measure (or equivalently of a stationary solution to (3.1)) under quite general conditions on b . This is in contrast with the long run behavior of Stochastic Differential Equations (SDEs) (or equivalently linear Fokker-Planck equations), for which mere confining properties of the drift suffice to

produce ergodicity (provided $\sigma > 0$). In the mean field setting, some further structural conditions are necessary. In the potential example (3.3), a standard condition is to demand V and W to be (strongly) convex, which forces in fact a form of convexity of the potential (3.3) on the space of probability measures. We refer for instance to the earlier works [BRV98, CMV03, Mal01, Mal03], which also include additional convergence results to the stationary regime. For a tiny list of variants, in which *perturbations* of the convex potential case are addressed, we refer to [Bas20, BGG13, But14, DMT19, DEGZ20]. We refrain from detailing all the possible extensions of the previous cases under which the invariant measure remains unique. Let us just say, for our purposes, that the above list of references includes cases where the lack of convexity is compensated for by the presence of a strong enough diffusion coefficient. This *large noise* regime is compared with our own setting in Remark 3.9 below. At this point, we would especially like to stress that there are known explicit simple cases in which uniqueness does not hold, which observation supports our previous claim: the presence of a nonlinear mean field term may *easily* lead to the existence of multiple invariant measures, even in situations where the mass stays confined under the dynamics (3.1). Specifically, it has been proven in [HT10] (see also [Daw83]) that, when V is only uniformly convex outside of a ball but admits a double-well, W is quadratic (but non-zero) and the diffusion coefficient σ is sufficiently small, there exist several stationary solutions to (3.1) when it derives from the potential (3.3) (and $\sigma_0 = 0$).

1.2 Common noise

Very basically, our primary objective in this article is to revisit the class of examples addressed in [HT10], but in presence of a common noise, i.e., $\sigma_0 > 0$. In this regard, it is worth noting that, when $b(x, \mu) = -\nabla V(x) - \nabla W * \mu(x)$, with V being uniformly convex and W being convex, the common noise does not change the picture recalled in the previous paragraph. Indeed, the second author has shown in [Mai23] that (3.1) has a unique invariant measure, which is consistent with the results obtained earlier without common noise with the slight subtlety that the invariant measure is then understood as a probability measure on $\mathcal{P}(\mathbb{R}^d)$.

Of course, our goal is to go one step beyond and prove that the common noise can change the picture positively, meaning that it can force uniqueness of the invariant measure even though the deterministic analogue of (3.1) (i.e., with $\sigma_0 = 0$) has several stationary solutions. Actually, this type of result, if it holds true, must be part of the very broad theory of ergodic Markov processes, with the specific feature that the state space here is $\mathcal{P}(\mathbb{R}^d)$. It is worth observing that, from the same point of view, the deterministic and linear analogue of (3.2) (i.e., b only depends on x and $\sigma_0 = 0$) should be regarded as a Fokker-Planck equation obtained by forcing a standard ordinary differential equation by a Brownian motion of intensity σ . This is just to say that, in the Euclidean setting, the same program is very well-understood and just consists in addressing the ergodic properties of a non-degenerate SDE. For sure, the very difficulty in our setting comes from the fact that the unknown in (3.1) lives in a space of infinite dimension (once again, the space of probability measures) while the noise (i.e., B^0) is just finite-dimensional and thus *completely* degenerate. This says that, at best, noise can be expected to restore uniqueness of the invariant measure only in specific cases. And, precisely, this is our objective to identify a class of cases that would become ergodic under the action of the common noise and that would include some of the examples addressed in [HT10].

In fact, this question was already addressed by the second author in [Mai23] when $\sigma = 0$, leaving open setting with $\sigma > 0$. Apart from this, the idea of using the averaging properties of the common noise has been used in several contexts, for instance in the analysis of existence and uniqueness of equilibria to mean field games (see for instance [DFT20, Tch18b] for games with a finite dimensional common noise and [Del19] for games with an infinite dimensional common noise). More recently, the first author has just released an article, [DH24], in which the ergodic properties of (3.1) (or, more precisely, (3.2)) are studied when b is general but the common noise is infinite dimensional and the dimension d is 1. Part of the challenge then precisely lies in the construction of the noise, which is a question different from the one addressed here. A similar problem was addressed in the prior work [ABKO23], for another type of infinite dimensional common noise that may not preserve the space of probability measures.

1.3 Our contribution

We here focus on drifts of the type

$$b(x, \mu) = G(x) + F \left(x - \int_{\mathbb{R}^d} x \mu(dx) \right), \quad x \in \mathbb{R}^d, \mu \in \mathcal{P}(\mathbb{R}^d),$$

where F and G are functions from \mathbb{R}^d into itself. The function F is assumed to be strictly decreasing, which occurs for instance if $F = -\nabla W$ for a strictly convex function. More interestingly, G is just assumed to be confining in a mild sense that is detailed in Assumption 5 below. In particular, G is not required to be strictly decreasing and, when it derives from a potential V , V may not be convex. Therefore, this framework allows us to choose $G = -\nabla V$ for a non-convex potential V and $F = -\nabla W$, for $W(x) = \alpha x^2$ for some $\alpha > 0$, and thus contains the case treated by [HT10] (up to an additional common noise), since $b(x, \mu)$ rewrites $-\nabla V(x) - \nabla W * \mu(x)$.

Our main contribution is to establish that, for a small diffusion coefficient σ and for an interaction force F that is sufficiently decreasing, the system has a unique invariant measure, which attracts exponentially fast any other initial condition (with appropriate integrability properties). Our strategy of proof is based on the key idea that the long run behavior of (3.2) should be dictated by the long time dynamics of the conditional mean $\mathbb{E}[X_t|B^0]$ given the common noise. However, this intuition is not completely correct, as the distance between X_t and its conditional mean $\mathbb{E}[X_t|B^0]$ does not tend to 0 in long time, at least when $\sigma > 0$. In fact, some residual fluctuation persists due to the presence of the idiosyncratic noise and this is one of our main achievement here to explain how to handle this residual fluctuation in long time. This is precisely the point where we need σ to be small and the interaction force F to be sufficiently decreasing. As for the function G , it mainly plays a role in the long time analysis of the conditional mean $\mathbb{E}[X_t|B^0]$ itself, which is shown to share many similarities with the long time analysis of the (standard) SDE driven by the drift G and the noise $\sigma_0 B^0$. In particular, we draw heavily on previous works on coupling methods for long time analysis of SDEs, especially the approach developed in [Ebe16] on couplings by reflection (inspired from [LR86]), which plays a key role in our study. Notice that our result complements the earlier one [Mai23] dedicated to the case $\sigma = 0$.

1.4 Further connections with the literature on mean field models

Stochastic PDEs like (3.1) and related conditional McKean-Vlasov equations of type (3.2) were introduced in [DV95, KX99]. Further, they were studied in a series of works [LPS13, LPS14, GS15, FG16, GS17, FG19] in connection with stochastic scalar conservation laws, in which case σ_0 is typically required to depend on the local value at point x of dm_t/dx . The more recent contribution [CG19] addresses uniqueness to (3.1) under weaker conditions than in [KX99]. Equation (3.1) has also become very popular in mean field game theory, see for instance the book [CDLL19]. A variant, including a reflection term, has been studied recently in [BCCdRH20]. As for (3.2), we refer to [HvS21] for a general existence and uniqueness result of weak solutions.

Last but not least, it is worth recalling that the connection between nonlinear Fokker-Planck equations and McKean-Vlasov equations (without common noise) goes back to the pioneering works of [Kac56, McK66, Fun84]. In this context, a key question concerns the particular approximation of the solutions, usually referred to as *propagation of chaos*, see for instance [Szn91, M el96]. In connection with the existence of stationary solutions to the Fokker-Planck equation (3.1), it is in general a difficult question to wonder whether propagation of chaos holds uniformly in time or not. We feel better not to give a list of references in this direction but we quote that the question appears for instance in the earlier contribution [CGM08]. Obviously, this would be very interesting to address the same problem but in presence of a common noise for the model studied here.

1.5 Organisation of the paper

In Section 2, we list the assumptions and give the main statements of the paper. We also address some examples and compare in particular our results with those from [HT10] (when $\sigma_0 = 0$ and uniqueness does not hold). The proof of the main theorem (Theorem 3.8) is split in two parts. A first step is to derive Theorem 3.8 from a key auxiliary estimate (Proposition 3.7) on the contraction properties of the semigroup induced by the solution of (3.1). This is done in Section 3. A second step, which is in fact the core of the paper, is to prove this key estimate. We do so in Section 4, using coupling arguments. Auxiliary results are proven in Appendix.

1.6 Notation

Throughout the paper, for a Polish space E , $\mathcal{P}(E)$ stands for the space of Borel probability measures on E equipped with the topology of weak convergence and the corresponding Borel σ -algebra (which is induced by the mappings $m \in \mathcal{P}(E) \mapsto m(A)$ for any Borel subset A of E). Whenever there exists a distance d so that (E, d) is a metric space, we call $\mathcal{P}_p(E)$, for any $p > 0$, the collection of elements $\mu \in \mathcal{P}_p(E)$ such that

$$\exists x_0 \in E : \int_E d(x_0, x)^p \mu(dx) < +\infty.$$

In fact, the integral above is finite or not, whatever the choice of x_0 . We then define for any $p \geq 1$, the p -Wasserstein distance on $\mathcal{P}_p(E)$ as

$$d_p^E(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \left(\int_{E \times E} d(x, y)^p \pi(dx, dy) \right)^{1/p}, \quad \mu, \nu \in \mathcal{P}_1(E),$$

where $\Pi(\mu, \nu)$ stands for the set of all couplings of μ and ν (i.e., all the joint probability measures on $E \times E$ with μ and ν as first and second marginals).

Also, for any measurable $\varphi : E \rightarrow \mathbb{R}$ and any probability measure $m \in \mathcal{P}(E)$, we define the duality product $\langle \cdot ; \cdot \rangle_E$ as

$$\langle m; \varphi \rangle_E = \int_E \varphi(x) m(dx),$$

whenever the integral in the right-hand side does exist (for instance so is the case if φ is bounded or φ is at most of linear growth and $m \in \mathcal{P}_1(E)$ and so forth...). When $E = \mathbb{R}^d$, we define

$$\begin{aligned} \mu_1(m) &:= \int_{\mathbb{R}^d} x m(dx), & m \in \mathcal{P}_1(\mathbb{R}^d); \\ \mu_2(m) &:= \int_{\mathbb{R}^d} |x|^2 m(dx), & m \in \mathcal{P}_2(\mathbb{R}^d); \\ v(m) &:= \int_{\mathbb{R}^d} |x - \mu_1(m)|^2 m(dx), & m \in \mathcal{P}_2(\mathbb{R}^d); \end{aligned}$$

which stand respectively for the expectation, the moment of order two and the variance of m . It is important to pay attention to the fact that despite the similarities in the notation, the objects $\mu_1(m) \in \mathbb{R}^d$ and $\mu_2(m) \in \mathbb{R}$ are of different dimensions.

When m is random, say is a measurable mapping from $(\Omega_0, \mathcal{F}^0, \mathbb{P}_0)$ into $\mathcal{P}(E)$ (or $\mathcal{P}_1(E)$), its law P is an element of $\mathcal{P}(\mathcal{P}(E))$ (or $\mathcal{P}(\mathcal{P}_1(E))$), in the sense that, for any bounded measurable function $\phi : \mathcal{P}(E) \rightarrow \mathbb{R}$, $\mathbb{E}_0[\phi(m)] = \langle P; \phi \rangle_{\mathcal{P}(E)}$ (and similarly, when working on $\mathcal{P}_1(E)$), where \mathbb{E}_0 is the expectation associated with \mathbb{P}_0 . Repeatedly in this paper, we also consider continuous stochastic processes $(m_t)_{t \geq 0}$, typically constructed on $(\Omega_0, \mathcal{F}^0, \mathbb{F}^0, \mathbb{P}_0)$, with values in the space $\mathcal{P}_1(\mathbb{R}^d)$ equipped with $d_1^{\mathbb{R}^d}$ (and thus also in $\mathcal{P}(\mathbb{R}^d)$, equipped with the weak convergence topology). When $(m_t)_{t \geq 0}$ is Markovian, we define its semigroup $(\mathcal{P}_t)_{t \geq 0}$ by letting, for any function $\phi : \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}$ and any $t \geq 0$,

$$\mathcal{P}_t \phi : \begin{array}{l} \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R} \\ m \mapsto \mathbb{E}_0[\phi(m_t) | m_0 = m]. \end{array}$$

Moreover, for any twice differentiable function $f : \mathbb{R}^d \rightarrow \mathbb{R}$, ∇f , Δf and $\nabla^2 f$ respectively stand for the gradient, Laplacian and Hessian matrix of f . For any differentiable function $F : \mathbb{R}^d \rightarrow \mathbb{R}^d$, we denote by $DF : \mathbb{R}^d \rightarrow \mathcal{M}_d(\mathbb{R})$ its Jacobian matrix.

For any real squared matrix $A \in \mathcal{M}_d(\mathbb{R})$, we denote the Euclidian norm of A by

$$\|A\| = \sqrt{\text{Tr}(A^\top A)},$$

with A^\top and Tr respectively standing for the transpose matrix of A and the Trace operator on $\mathcal{M}_d(\mathbb{R})$. Lastly, the Euclidean norm on \mathbb{R}^d is denoted $|\cdot|$ and the inner product \cdot .

2 Main result: uniqueness recovery thanks to common noise

2.1 Stochastic Fokker-Planck PDE and Conditional McKean-Vlasov SDE

In this paper, we focus on a mean field model with common noise in which the nonlinearity in the dynamics occurs through the mean (i.e., the expectation). To make it clear, on a filtered probability space $(\Omega_0, \mathcal{F}^0, \mathbb{F}^0, \mathbb{P}_0)$ (satisfying the usual conditions) equipped with a d -dimensional \mathbb{F}^0 -Brownian motion B^0 and for $F, G : \mathbb{R}^d \rightarrow \mathbb{R}^d$ two interaction and confinement forces, we are interested in the solutions of the following Stochastic Partial Differential Equation (SPDE):

$$d_t m_t = \nabla \cdot \left(\frac{\sigma^2 + \sigma_0^2}{2} \nabla m_t - m_t (G + F(\cdot - \mu_1(m_t))) \right) dt - \sigma_0 \nabla m_t \cdot dB_t^0, \quad t \geq 0, \quad (3.4)$$

which is understood the weak sense, i.e., the process $(m_t)_{t \geq 0}$ satisfies for all $t \geq 0$ and $\varphi \in C_c^\infty(\mathbb{R}^d)$,

$$d \langle m_t, \varphi \rangle = \langle m_t, \mathcal{L}_m \varphi \rangle dt + \sigma_0 \langle m_t, (\nabla \varphi)^\top \rangle dB_t^0,$$

where for any probability measure $m \in \mathcal{P}_1(\mathbb{R}^d)$, the operator \mathcal{L}_m acts on a smooth function φ with compact support in the following manner:

$$\mathcal{L}_m \varphi = - (G + F(\cdot - \mu_1(m))) \cdot \nabla \varphi + \frac{\sigma_0^2 + \sigma^2}{2} \Delta \varphi.$$

Equation (3.4) admits a Lagrangian representation in the form of conditional McKean-Vlasov equation. Throughout the article, this McKean-Vlasov equation is constructed on a product probability space, obtained by tensorizing the filtered probability space $(\Omega_0, \mathcal{F}^0, \mathbb{F}^0, \mathbb{P}_0)$ (on which Equation (3.4) is defined) with another filtered probability space $(\Omega_1, \mathcal{F}^1, \mathbb{F}^1, \mathbb{P}_1)$. The product structure is denoted

$$(\Omega := \Omega_0 \times \Omega_1, \mathcal{F}, \mathbb{F}, \mathbb{P}),$$

where $(\mathcal{F}, \mathbb{P})$ is the completion of $(\mathcal{F}^0 \otimes \mathcal{F}^1, \mathbb{P}_0 \otimes \mathbb{P}_1)$ and \mathbb{F} is the right continuous augmentation of $(\mathcal{F}_t^0 \otimes \mathcal{F}_t^1)_{t \geq 0}$. In this paper we denote by $\mathbb{E}_0, \mathbb{E}_1$ and \mathbb{E} the expectations on $(\Omega_0, \mathcal{F}^0, \mathbb{F}^0, \mathbb{P}_0)$, $(\Omega_1, \mathcal{F}^1, \mathbb{F}^1, \mathbb{P}_1)$ and $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ respectively. Analogously, for any random variable X defined on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, we denote by $\mathcal{L}_0(X)$ and $\mathcal{L}_1(X)$ the conditional laws of the random variable X given \mathcal{F}^1 and \mathcal{F}^0 respectively. Of course, $\mathcal{L}(X)$ stands for the law of X on the whole product space.

Next, the initial law of (3.4) may be random and then distributed according to a probability measure $P_0 \in \mathcal{P}(\mathcal{P}_1(\mathbb{R}^d))$. The initial condition m_0 is then defined as an \mathcal{F}_0^0 -measurable random variable with values in $\mathcal{P}_1(\mathbb{R}^d)$ such that $\mathcal{L}(m_0) = P_0$. We can now define on the whole probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ a random variable X_0 such that $\mathcal{L}_1(X_0) = m_0$ almost surely. Precisely, by Lemma 2.4 in [CD18b], for \mathbb{P}^0 -a.e. $\omega_0 \in \Omega_0$, $X_0(\omega_0, \cdot)$ is a random variable on $(\Omega_1, \mathcal{F}^1, \mathbb{P}_1)$ and the conditional law of X_0 given \mathcal{F}^0 , $\mathcal{L}_1 : \Omega_0 \ni \omega_0 \mapsto \mathcal{L}(X_0(\omega_0, \cdot))$, defines a random variable from $(\Omega_0, \mathcal{F}^0, \mathbb{P}_0)$ into $\mathcal{P}(\mathbb{R}^d)$, which can be taken equal to m_0 almost surely.

In addition to B^0 , we consider a d -dimensional \mathbb{F}^1 -Brownian motion B supported by the space $(\Omega_1, \mathcal{F}^1, \mathbb{P}_1)$. The Lagrangian formulation of (3.4) then reads in the form of the following conditional

McKean-Vlasov equation, set on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$:

$$dX_t = G(X_t)dt + F(X_t - \mathbb{E}_1[X_t])dt + \sigma dB_t + \sigma_0 dB_t^0, \quad t \geq 0, \quad (3.5)$$

with X_0 as initial condition.

2.2 Assumptions, existence and uniqueness of solutions

In the sequel, we make the following assumptions on the coefficients F and G . We start with G :

Assumption 5. *The drift $G : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is differentiable in \mathbb{R}^d . Moreover,*

- G is confining in the sense that there exists a function $\kappa : [0; +\infty) \rightarrow \mathbb{R}$ such that

$$\forall x, y \in \mathbb{R}^d, \quad (G(x) - G(y)) \cdot (x - y) \leq -\frac{\sigma_0^2}{2} \kappa(|x - y|) |x - y|^2,$$

with

$$\limsup_{r \rightarrow +\infty} \kappa(r) > 0 \quad \text{and} \quad \int_0^1 r \kappa(r)^- dr < \infty,$$

where $\kappa^- = \max(0, -\kappa)$.

- G is Lipschitz-continuous on \mathbb{R}^d , meaning that there exists a constant $L_G > 0$ such that

$$\forall x, y \in \mathbb{R}^d, \quad |G(x) - G(y)| \leq L_G |x - y|.$$

- G is differentiable and its Jacobian matrix DG is Lipschitz-continuous on \mathbb{R}^d , meaning that there exists a constant $C_G > 0$ such that for any $x, y \in \mathbb{R}^d$,

$$\|DG(x) - DG(y)\| \leq C_G |x - y|.$$

In order to state properly the assumption on F , we introduce the following definition:

Definition 3.1. *For a pair $(\alpha, C) \in \mathbb{R} \times [0, +\infty)$, we call $\mathcal{S}(\alpha, C, L)$ the collection of L -Lipschitz-continuous and differentiable functions H from \mathbb{R}^d into itself that are α -decreasing and whose derivative is C Lipschitz-continuous, i.e., for all $(x, y) \in \mathbb{R}^{2d}$,*

$$(i) \quad (x - y) \cdot (H(x) - H(y)) \leq -\alpha |x - y|^2;$$

$$(ii) \quad \|DH(x) - DH(y)\| \leq C |x - y|.$$

Given Definition 3.1, we make the following assumption on F :

Assumption 6. *$F : \mathbb{R}^d \rightarrow \mathbb{R}^d$ satisfies $F(0) = 0$. Moreover, there exist $\alpha_F > 0$ and $C_F \geq 0$ such that $F \in \mathcal{S}(\alpha_F, C_F, C_F)$.*

About Assumption 5

The following remarks about Assumption 5 are in order:

- Assumption 5 is quite intuitive and consistent with the ideas presented in [HT10]. Essentially, this assumption guarantees that G is confining and behaves linearly at infinity. Also, it is worth noting that our approach can accommodate a force that derives from a non-convex potential, which, around the origin, behaves for example like $x \mapsto |x|^4 - |x|^2$.
- When $\sigma_0 > 0$, there exists a canonical choice of κ , given by

$$\kappa(r) := \inf \left\{ -\frac{2}{\sigma_0^2} \frac{(x-y) \cdot (G(x) - G(y))}{|x-y|^2}; \quad x, y \in \mathbb{R}^d \text{ s.t. } |x-y| = r \right\}, \quad r > 0.$$

Assumption 5 says that κ is necessarily positive outside of a ball. This implies in particular that $\sigma_0^2 \kappa^- / 2$ is bounded from below by some constant $m_G \in \mathbb{R}$, which gives

$$(x-y) \cdot (G(x) - G(y)) \leq -m_G |x-y|^2, \quad x, y \in \mathbb{R}^d. \quad (3.6)$$

This implies in particular that $G \in \mathcal{S}(m_G, C_G, L_G)$.

When $m_G > 0$ we say that the drift is (strictly) decreasing. This case is easier to study under similar conditions on F because the monotone structure of G ensures that two solutions of the SDE (3.5) (with different initial conditions) get closer in large time. In this context, uniqueness of the invariant measure of the solutions of Equation (3.4) is covered in [Mai23, Section 2], but it is fair to say that the common noise then plays little role because the invariant measure is also unique when $\sigma_0 = 0$. Notice that in the specific case when G derives from a potential, this potential is strictly convex under the condition $m_G > 0$.

About Assumption 6

Assumption 6 is somewhat surprising, especially as we are asking below the coefficient α_F to be large (see the statement of Theorem 3.8). Intuitively, one might indeed expect that uniqueness of the invariant measure becomes less likely if the interaction is strong. At least, this is exactly what happens in absence of common noise, the extreme case being $F \equiv 0$ in which the dynamics become a mere diffusion equation (the long time analysis of which is much easier to study). However, the situation is different here, thanks to the peculiar structure of the dynamics and to the presence of the common noise ($\sigma_0 > 0$). In brief, F has a contracting effect, which forces the process to be attracted by its conditional expectation (given the common noise), with the latter being a nearly solution of an ergodic SDE. Even though this picture is not exactly correct due to some minor errors caused by the idiosyncratic noise, this is indeed a key step in our analysis to quantify the accuracy of this approximation and to derive from it uniqueness of the invariant measure. We refer to Section 3.1.

Throughout, we work under Assumptions 5 and 6. In this context, we claim

Proposition 3.2. *Under Assumptions 5 and 6 and for an initial condition X_0 satisfying $\mathbb{E}[|X_0|] < \infty$ (or equivalently $\mathbb{E}_0 \int_{\mathbb{R}^d} |x| m_0(dx) < \infty$ for $m_0 := \mathcal{L}_0(X_0)$), the conditional McKean-Vlasov equation (3.5)*

has a unique \mathbb{F} -progressively measurable solution $(X_t)_{t \geq 0}$, with continuous trajectories, such that, for all $T > 0$, $\mathbb{E}[\sup_{0 \leq t \leq T} |X_t|] < \infty$.

Proof. The proof consists in a straightforward fixed point argument, using the Lipschitz properties of the two coefficients F and G . We refer to [CD18b, Chap. 2]. \square

By the superposition principle for conditional McKean-Vlasov equations (see [LSZ23]), we deduce that existence and uniqueness also hold true for the stochastic Fokker-Planck equation:

Proposition 3.3. *Under Assumptions 5 and 6 and for an initial condition m_0 satisfying $\mathbb{E}_0 \int_{\mathbb{R}^d} |x| m_0(dx) < \infty$, the stochastic Fokker-Planck equation (3.4) has a unique solution $(m_t)_{t \geq 0}$ in the space of \mathbb{F}^0 -progressively measurable processes with values in $\mathcal{P}_1(\mathbb{R}^d)$ satisfying $\mathbb{E}_0[\sup_{0 \leq t \leq T} \int_{\mathbb{R}^d} |x| m_t(dx)] < \infty$.*

Proof. Existence of a solution is a straightforward consequence of the existence part in the statement of Proposition 3.2. This is Proposition 1.2 in [LSZ23]. Uniqueness is more difficult to obtain. We invoke Theorem 1.3 in [LSZ23]. In brief, any solution $(\tilde{m}_t)_{t \geq 0}$ (in the weak sense) to (3.4) induces a weak solution to the McKean-Vlasov equation $\tilde{X} = (\tilde{X}_t)_{t \geq 0}$ to (3.5), weak in the sense that the private (or idiosyncratic) noise, say \tilde{B} , becomes part of the solution. It holds $\tilde{m}_t = \mathcal{L}(\tilde{X}_t | \mathcal{F}_T^0)$. In fact, by strong uniqueness to (3.5), a relevant form of Yamada-Watanabe theorem applies to the current setting and implies that $\mathcal{L}(\tilde{X}_t | \mathcal{F}_T^0)$ is necessarily equal to $\mathcal{L}_0(X_t)$, with $X = (X_t)_{t \geq 0}$ being the solution given by Proposition 3.2. \square

2.3 Restoration of uniqueness

We now collect our main results about the long time behaviour of the process $(m_t)_{t \geq 0}$, solution of Equation (3.4). Before we do so, we first state the precise definition of an invariant measure for such a stochastic process, when seen as a process with values in the space of probability measures:

Definition 3.4 (Invariant measure). *We say that $\bar{P} \in \mathcal{P}_2(\mathcal{P}_1(\mathbb{R}^d))$ is an invariant measure for the process $(m_t)_{t \geq 0}$ when the latter is regarded as taking values in $\mathcal{P}_1(\mathbb{R}^d)$, if the law of m_t is independent of t , when m_0 is distributed according to \bar{P} , i.e., for all continuous and bounded function $\phi \in \mathcal{C}_b(\mathcal{P}_1(\mathbb{R}^d))$*

$$\mathbb{E}_0[\phi(m_t)] = \mathbb{E}_0[\phi(m_0)] = \int_{\mathcal{P}_1(\mathbb{R}^d)} \phi(m) \bar{P}(dm), \quad \forall t > 0.$$

Of course, if \bar{P} is an invariant probability distribution, then the process $(m_t)_{t \geq 0}$ with initial condition \bar{P} , has the same law as the process $(m_{t+T})_{t \geq 0}$ for any $T > 0$. This follows directly from the fact that m_0 and m_T have the same law combined with the weak Markov property. Moreover, it is worth observing that, under the condition $\bar{P} \in \mathcal{P}_2(\mathcal{P}_1(\mathbb{R}^d))$ imposed in the statement, the function $m \mapsto d_1^{\mathbb{R}^d}(\delta_0, m) = \int_{\mathbb{R}^d} |x| m(dx)$ is square integrable. This constraint may seem rather artificial. In fact, given the form of the dynamics in Equation (3.4), it is natural to require the invariant measures to be supported by $\mathcal{P}_1(\mathbb{R}^d)$. Then, the aforementioned integrability condition provides an additional growth property that makes easier the identification of those invariant measures, see the proof of Proposition 3.5 right below.

Let us begin with the existence of an invariant measure \bar{P} :

Proposition 3.5. *Under Assumptions 5 and 6, there exists at least one invariant measure \bar{P} for the process defined by Equation (3.4). Moreover, under the condition $\alpha > L_G$, any invariant measure belongs to $\mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$.*

The proof of Proposition 3.5 makes use of the notion of derivatives on the space of measures. Following [Car10] (see also [CD18a, Chap. 5] and [CD18b, Chap. 4]), we recall the following definition:

Definition 3.6. *Let us define $\mathcal{C}_b^2(\mathcal{P}_2(\mathbb{R}^d))$ as the collection of continuous and bounded functions $\Phi : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ with the following properties (throughout the definition, $\mathcal{P}_2(\mathbb{R}^d)$ is equipped with $d_2^{\mathbb{R}^d}$):*

- *There exists a unique jointly continuous function $\partial_m \Phi : (m, x) \in \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \rightarrow \partial_m \Phi(m, x) \in \mathbb{R}^d$, at most of quadratic growth in x for any m , such that*

$$\lim_{h \rightarrow 0} \frac{\Phi(m + h(m' - m)) - \Phi(m)}{h} = \int_{\mathbb{R}^d} \partial_m \Phi(m, v)(m' - m)(dv),$$

for all $m, m' \in \mathcal{P}(\mathbb{R}^d)$ and

$$\int_{\mathbb{R}^d} \partial_m \Phi(m, v)m(dv) = 0, \quad m \in \mathcal{P}(\mathbb{R}^d);$$

- *For any $m \in \mathcal{P}_2(\mathbb{R}^d)$, the mapping $x \mapsto \partial_m \Phi(m, x)$ is differentiable, with the gradient being denoted $D_m \Phi(m, x)$; the mapping $(m, x) \mapsto D_m \Phi(m, x)$ is jointly continuous in (m, x) and at most of linear growth in x uniformly in m in bounded subsets of $\mathcal{P}_2(\mathbb{R}^d)$;*
- *For any $x \in \mathbb{R}^d$, every component of the \mathbb{R}^d -valued function $m \mapsto D_m \Phi(m, x)$ satisfies the same conditions as in the first bullet point, resulting in a mapping $(m, x, y) \in \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \times \mathbb{R}^d \mapsto D_m^2 \Phi(m, x, y) \in \mathbb{R}^{d \times d}$, which is jointly continuous (in the three arguments) and at most of quadratic growth in (x, y) , uniformly in m in bounded subsets of $\mathcal{P}_2(\mathbb{R}^d)$;*
- *For any $m \in \mathcal{P}_2(\mathbb{R}^d)$, the function $x \in \mathbb{R}^d \mapsto D_m \Phi(m, x)$ is differentiable; the Jacobian, denoted $(m, x) \mapsto D_{xm}^2 \Phi(m, x)$, is jointly continuous in (m, x) and at most of linear growth in x , uniformly in m in bounded subsets of $\mathcal{P}_2(\mathbb{R}^d)$.*

Proof of Proposition 3.5. Step 1. The proof of the existence of an invariant measure closely follows the approach taken in Proposition 2 of [Mai23], using the results of Section 1.2 of [LSZ23]. The key step therein is to control, uniformly in time, the second-order moments of the solutions to Equation (3.4), which can be achieved with relative ease in our context. Details are left to the reader.

Step 2. We now address the second part of the statement (any invariant measure in $\mathcal{P}_2(\mathcal{P}_1(\mathbb{R}^d))$ is in fact in $\mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$). Let us thus consider an invariant measure $\bar{P} \in \mathcal{P}_2(\mathcal{P}_1(\mathbb{R}^d))$ and call m_0 an \mathcal{F}_0 -valued random variable with values in $\mathcal{P}_1(\mathbb{R}^d)$ such that $\mathcal{L}_0(m_0) = \bar{P}$. We denote by $(m_t)_{t \geq 0}$ the solution to Equation (3.4) with m_0 as initial condition. We also introduce for any $R > 0$, a smooth function $\chi_R^d : \mathbb{R}^d \rightarrow \mathbb{R}^d$ that coincides with the identity on the ball $B_{\mathbb{R}^d}(0, R)$ and that is equal to 0 outside $B_{\mathbb{R}^d}(0, 2R)$. We can assume the Lipschitz constant of χ_R^d to be bounded by 1. For any $R > 0$, we denote by P_0^R the law of $\chi_R^d \# m_0$ (where $\#$ is the pushforward operator) and construct $(m_t^R)_{t \geq 0}$, the solution to Equation (3.4) with $\chi_R^d \# m_0$ as initial condition. Letting $(P_t^R)_{t \geq 0} := (\mathcal{L}_0(m_t^R))_{t \geq 0}$,

we deduce from the truncation performed on the initial condition and from Equation (3.4) that, for any $R > 0$ and any $t \geq 0$, $P_t^R(\mathcal{P}_2(\mathbb{R}^d)) = 1$. We obtain from the chain rule proven in [CD18a, Chap. 5] and [CD18b, Chap. 4] (see [LSZ23] for the form that is retained below) that for any function $\Phi \in \mathcal{C}_b^2(\mathcal{P}_2(\mathbb{R}^d))$, for any $t \geq 0$,

$$\langle P_t^R - P_0^R; \Phi \rangle = \int_0^t \langle P_s^R; \mathcal{M}\Phi \rangle ds, \quad (3.7)$$

where for any $m \in \mathcal{P}_2(\mathbb{R}^d)$,

$$\begin{aligned} \mathcal{M}\Phi(m) := & \int_{\mathbb{R}^d} \left[D_m \Phi(m, x) \cdot (G(x) + F(x - \mu_1(m))) + \frac{\sigma^2 + \sigma_0^2}{2} \text{Tr} [D_{xm}^2 \Phi(m, x)] \right] m(dx) \\ & + \frac{\sigma_0^2}{2} \int_{\mathbb{R}^{2d}} \text{Tr} [D_{mm}^2 \Phi(m, x, y)] m(dx) m(dy). \end{aligned}$$

Let us now consider $\Phi(m) := \int_{\mathbb{R}^d} \phi(x - \mu_1(m)) m(dx)$, for a function $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ which we assume to be bounded and smooth, with bounded derivatives of order 1 and 2. In particular, Φ is bounded. Now, $m \mapsto \Phi(m) \in \mathcal{C}_b^2(\mathcal{P}_2(\mathbb{R}^d))$ and we can write, for any $m \in \mathcal{P}_2(\mathbb{R}^d)$ and $x, y \in \mathbb{R}^d$,

$$\begin{aligned} D_m \Phi(m, x) &= \nabla \phi(x - \mu_1(m)) - \int_{\mathbb{R}^d} \nabla \phi(y - \mu_1(m)) m(dy), \\ D_{xm}^2 \Phi(m, x) &= \nabla^2 \phi(x - \mu_1(m)), \\ D_{mm}^2 \Phi(m, x, y) &= -\nabla^2 \phi(x - \mu_1(m)) - \nabla^2 \phi(y - \mu_1(m)) + \int_{\mathbb{R}^d} \nabla^2 \phi(z - \mu_1(m)) m(dz). \end{aligned}$$

Plugging this into Equation (3.7), we obtain

$$\begin{aligned} & \int_{\mathcal{P}_2(\mathbb{R}^d)} \Phi(m) (P_t^R - P_0^R)(dm) \\ &= \int_0^t \left[\int_{\mathcal{P}_2(\mathbb{R}^d)} \int_{\mathbb{R}^d} \nabla \phi(x - \mu_1(m)) \cdot (G(x) + F(x - \mu_1(m))) m(dx) P_s^R(dm) \right. \\ & \quad \left. - \int_{\mathcal{P}_2(\mathbb{R}^d)} \left(\int_{\mathbb{R}^d} \nabla \phi(y - \mu_1(m)) m(dy) \right) \cdot \left(\int_{\mathbb{R}^d} (G(x) + F(x - \mu_1(m))) m(dx) \right) P_s^R(dm) \right. \\ & \quad \left. + \frac{1}{2} (\sigma^2 + \sigma_0^2) \int_{\mathbb{R}^d} \text{Tr} [\nabla^2 \phi(z - \mu_1(m))] m(dz) - \frac{1}{2} \sigma_0^2 \int_{\mathbb{R}^d} \text{Tr} [\nabla^2 \phi(z - \mu_1(m))] m(dz) \right] ds, \end{aligned}$$

which gives (by simplifying the last line)

$$\begin{aligned}
 & \int_{\mathcal{P}_2(\mathbb{R}^d)} \Phi(m)(P_t^R - P_0^R)(dm) \\
 &= \int_0^t \left[\int_{\mathcal{P}_2(\mathbb{R}^d)} \int_{\mathbb{R}^d} \nabla \phi(x - \mu_1(m)) \cdot (G(x) + F(x - \mu_1(m)))m(dx) P_s^R(dm) \right. \\
 & \quad \left. - \int_{\mathcal{P}_2(\mathbb{R}^d)} \left(\int_{\mathbb{R}^d} \nabla \phi(y - \mu_1(m))m(dy) \right) \cdot \left(\int_{\mathbb{R}^d} (G(x) + F(x - \mu_1(m)))m(dx) \right) P_s^R(dm) \right. \\
 & \quad \left. + \frac{1}{2} \sigma^2 \int_{\mathbb{R}^d} \text{Tr}[\nabla^2 \phi(z - \mu_1(m))]m(dz) \right] ds.
 \end{aligned} \tag{3.8}$$

Step 3. We now want to let $R \rightarrow \infty$. Let us begin with the term on the left-hand side

$$\int_{\mathcal{P}(\mathbb{R}^d)} \Phi(m)(P_t^R - P_0^R)(dm) = \int_{\mathcal{P}(\mathbb{R}^d)} \Phi(m)(P_t^R - \bar{P})(dm) + \int_{\mathcal{P}(\mathbb{R}^d)} \Phi(m)(\bar{P} - P_0^R)(dm).$$

Using the Lipschitz-continuity of F and G , we can show that, for any $T \geq 0$, there exists a constant $C_T \geq 0$, such that, for any $t \in [0, T]$,

$$d_2^{\mathcal{P}_1(\mathbb{R}^d)}(P_t^R, \bar{P}) \leq C_T d_2^{\mathcal{P}_1(\mathbb{R}^d)}(P_0^R, \bar{P}). \tag{3.9}$$

Moreover, we can write

$$d_2^{\mathcal{P}_1(\mathbb{R}^d)}(P_0^R, \bar{P})^2 \leq \int_{\mathcal{P}(\mathbb{R}^d)} \left(\int_{\mathbb{R}^d} |\chi_R^d(x) - x|m(dx) \right)^2 \bar{P}(dm).$$

Since χ_R^d is equal to 0 in 0 and is 1-Lipschitz continuous and since $\bar{P} \in \mathcal{P}_2(\mathcal{P}_1(\mathbb{R}^d))$, the above right-hand side tends to 0 as R tends to ∞ , that is

$$\lim_{R \rightarrow +\infty} d_2^{\mathcal{P}_1(\mathbb{R}^d)}(P_0^R, \bar{P}) = 0. \tag{3.10}$$

Equations (3.9) and (3.10) say that for all $T \geq 0$,

$$\lim_{R \rightarrow +\infty} \sup_{0 \leq t \leq T} d_2^{\mathcal{P}_1(\mathbb{R}^d)}(P_t^R, \bar{P}) = 0. \tag{3.11}$$

Back to (3.8), we observe that all the functions of m that are integrated with respect to \bar{P} therein are continuous in m with respect to $d_1^{\mathbb{R}^d}$. Moreover, they are all at most of linear growth with respect to $\int_{\mathbb{R}^d} |x|m(dx)$. This makes it possible to let R tend to $+\infty$. We get

$$\begin{aligned}
 & \int_{\mathcal{P}_1(\mathbb{R}^d)} \int_{\mathbb{R}^d} \nabla \phi(x - \mu_1(m)) \cdot (G(x) + F(x - \mu_1(m)))m(dx) \bar{P}(dm) \\
 & \quad - \int_{\mathcal{P}_1(\mathbb{R}^d)} \left(\int_{\mathbb{R}^d} \nabla \phi(y - \mu_1(m))m(dy) \right) \cdot \left(\int_{\mathbb{R}^d} (G(x) + F(x - \mu_1(m)))m(dx) \right) \bar{P}(dm) \\
 & \quad + \frac{1}{2} \sigma^2 \int_{\mathcal{P}_1(\mathbb{R}^d)} \int_{\mathbb{R}^d} \text{Tr}[\nabla^2 \phi(z - \mu_1(m))]m(dz) \bar{P}(dm) = 0.
 \end{aligned}$$

Step 4. Now, we take $\phi(x)$ in the form $\vartheta(|x|)$ where ϑ is a smooth function from $[0, +\infty)$ into itself, that is equal to the identity on $[0, A]$ and that is equal to $3A/2$ on $[2A, +\infty)$, for some $A > 0$. We assume ϑ' to be non-negative and bounded by the identity and ϑ'' to take values in $[0, 2]$. Under this choice, $\nabla\phi(x)$ rewrites $[\vartheta'(|x|)/|x|]x$ and $\text{Tr}[\nabla^2\phi(x)]$ is bounded by a constant c_d only depending on the dimension d . Then, we can find a constant C , only depending on G (and whose value is allowed to change from line to line) such that, for any $m \in \mathcal{P}_2(\mathbb{R}^d)$,

$$\begin{aligned} & \left| \int_{\mathbb{R}^d} \nabla\phi(x - \mu_1(m)) \cdot G(x)m(dx) \right| \\ & \leq \left| \int_{\mathbb{R}^d} \nabla\phi(x - \mu_1(m)) \cdot [G(x) - G(\mu_1(m))]m(dx) \right| + \left| \int_{\mathbb{R}^d} \nabla\phi(x - \mu_1(m)) \cdot G(\mu_1(m))m(dx) \right| \\ & \leq L_G \int_{\mathbb{R}^d} \vartheta'(|x - \mu_1(m)|)|x - \mu_1(m)|m(dx) + C \int_{\mathbb{R}^d} \vartheta'(|x - \mu_1(m)|)(1 + |\mu_1(m)|)m(dx) \\ & \leq L_G \int_{\mathbb{R}^d} \vartheta'(|x - \mu_1(m)|)|x - \mu_1(m)|m(dx) + C(1 + d_1^{\mathbb{R}^d}(\delta_0, m)^2). \end{aligned}$$

We insist on the fact that C is independent of ϑ . Now using the assumption on F (see Assumption 6), we obtain

$$\int_{\mathbb{R}^d} \nabla\phi(x - \mu_1(m)) \cdot F(x - \mu_1(m))m(dx) \leq -\alpha \int_{\mathbb{R}^d} \vartheta'(|x - \mu_1(m)|)|x - \mu_1(m)|m(dx).$$

Rewriting the conclusion of the third step gives

$$\begin{aligned} & (\alpha - L_G) \int_{\mathcal{P}_1(\mathbb{R}^d)} \int_{\mathbb{R}^d} \vartheta'(|x - \mu_1(m)|)|x - \mu_1(m)|m(dx)\bar{P}(dm) \\ & \leq c_d\sigma^2 + C \left(1 + \int_{\mathcal{P}_1(\mathbb{R}^d)} d_1^{\mathbb{R}^d}(\delta_0, m)^2\bar{P}(dm) \right). \end{aligned} \tag{3.12}$$

Now, we can choose ϑ' along a non-decreasing sequence converging pointwise to the identity function on $[0, +\infty)$. By monotone convergence, we obtain

$$(\alpha - L_G) \int_{\mathcal{P}_1(\mathbb{R}^d)} \int_{\mathbb{R}^d} |x - \mu_1(m)|^2 m(dx)\bar{P}(dm) \leq c_d\sigma^2 + C \left(1 + \int_{\mathcal{P}_1(\mathbb{R}^d)} d_1^{\mathbb{R}^d}(\delta_0, m)^2\bar{P}(dm) \right),$$

which completes the proof. \square

Uniqueness is much more challenging to establish. When the intensity of the idiosyncratic noise σ is positive, this question has not been addressed yet. The following proposition is one key step in this regard and demonstrates that, under certain assumptions, the initial condition is forgotten in long time.

Proposition 3.7. *For any $\alpha_F > \max(L_G, m_G)$ in Assumption 6, for any $\sigma_0 > 0$, there exists $\bar{\sigma} > 0$, depending on F, G, σ_0 and d , such that for all $\sigma \leq \bar{\sigma}$, for any 1-Lipschitz continuous function $\phi : \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}$ (with $\mathcal{P}_1(\mathbb{R}^d)$ being equipped with $d_1^{\mathbb{R}^d}$) and any $m, \tilde{m} \in \mathcal{P}_2(\mathbb{R}^d)$,*

$$|\mathcal{P}_t\phi(m) - \mathcal{P}_t\phi(\tilde{m})| \leq \mathcal{K}(m, \tilde{m})e^{-ct}, \quad t \geq 0,$$

for some $\mathfrak{c} := c(G, F, \sigma, \sigma_0) > 0$ and some $\mathcal{K}(m, \tilde{m})$, which is described next.

Indeed, given the choice of α_F , there exist a constant $c_1^{\alpha_F} > 0$, depending on G and F , and a constant C_d , depending on d (and thus independent of ϕ , m and \tilde{m}), such that

$$\mathcal{K}(m, \tilde{m}) := C_d(\mu_2(m)^{1/2} + \mu_2(\tilde{m})^{1/2}) \exp\left(\frac{c_1^{\alpha_F}}{c_{\alpha_F}} \left(v(m)^{1/2} + v(\tilde{m})^{1/2}\right)\right),$$

where $c_{\alpha_F} := \alpha_F - L_G$. We recall that μ_1 , μ_2 and v are defined in Section 1.6.

Proposition 3.7 shows that we forget the initial condition at an exponential rate for any sufficiently integrable initial conditions m and \tilde{m} . This is a key result of the paper, the proof of which is postponed to Section 4.

As a consequence of Proposition 3.7, we get uniqueness of the invariant measure and deduce that the system converges exponentially fast toward the stationary regime, both facts being stated in the main statement below.

Theorem 3.8. *For any $\alpha_F > \max(L_G, m_G)$ in Assumption 6, for any $\sigma_0 > 0$, there exists $\bar{\sigma} > 0$, depending on F, G, σ_0 and d , such that for all $\sigma \leq \bar{\sigma}$, there is a unique invariant measure $\bar{P} \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$ for the process $(m_t)_{t \geq 0}$, solution of Equation (3.4). More precisely, there exists a constant $C > 0$, such that for any probability measure $m \in \mathcal{P}(\mathbb{R}^d)$ and any 1-Lipschitz continuous function $\phi : \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}$, (with $\mathcal{P}_1(\mathbb{R}^d)$ being equipped with $d_1^{\mathbb{R}^d}$), we have*

$$\left| \mathcal{P}_t \phi(m) - \int_{\mathcal{P}(\mathbb{R}^d)} \mathcal{P}_t \phi(\tilde{m}) \bar{P}(d\tilde{m}) \right| \leq C(1 + \mu_2(m)^{1/2}) \exp(\lambda v(m)^{1/2}) e^{-ct}, \quad t \geq 0,$$

with $\lambda := c_1^{\alpha_F} / c_{\alpha_F} > 0$, and for \mathfrak{c} , c_{α_F} and $c_1^{\alpha_F}$ as in the statement of Proposition 3.7.

Remark 3.9. *Let us comment here on the choice of parameters in the above statement. First, it is important to note that the type and strength of the interaction play a central role in our framework. The interaction, in both its form and intensity ($\alpha_F \geq \bar{\alpha}$), forces the long-time behaviour of the entire process $(m_t)_{t \geq 0}$ to be dictated by its mean $(\mu_1(m_t))_{t \geq 0}$, which is the cornerstone of the proof. Asking the interaction to be strong, as we require in the condition $\alpha_F \geq \bar{\alpha}$, may seem surprising at first sight. Indeed, this is fundamentally different from the standard picture that exists for models without common noise, where strong interactions are typically expected to give rise to several invariant measures. In our case, the presence of the common noise, even of a possibly small intensity, induces a phase transition. For any positive value of σ_0 and for a sufficiently large interaction (independently of the value of σ_0), we can indeed find $\sigma > 0$ such that uniqueness of the invariant measure is ensured in presence of a common noise of intensity σ_0 but is lost when the common noise is disabled. This makes a conceptual difference with the previous results obtained on the subject in [JJ24], where uniqueness is restored when the global contribution of the two noises is large enough with respect to the strength of the interaction. In comparison, σ and σ_0 have opposite roles to each other in our approach.*

2.4 Discussions and examples

In this section, we present two explicit examples in which common noise forces uniqueness of the invariant measure. The first example fits explicitly the framework of this work and the second one may be seen as a variant of (3.4). At last, we also provide a counter-example where uniqueness does not hold despite the presence of the common noise.

A prototype: dynamics driven by a confining potential and a linear interaction.

Let us consider the case $G(x) = -\nabla V(x)$, $x \in \mathbb{R}^d$, for a differentiable function $V : \mathbb{R}^d \rightarrow \mathbb{R}$, and $F(x) = -\alpha x$, $x \in \mathbb{R}^d$, for a certain $\alpha > 0$. Then, (3.4) becomes

$$d_t m_t = \frac{\sigma^2 + \sigma_0^2}{2} \Delta m_t dt + \nabla \cdot \left[m_t \left(\nabla V + \alpha (\cdot - \mu_1(m_t)) \right) \right] dt - \sigma_0 \nabla m_t \cdot dB_t^0, \quad t \geq 0. \quad (3.13)$$

When $\sigma_0 = 0$, this equation reduces to

$$\partial_t m_t = \frac{\sigma^2}{2} \Delta m_t + \nabla \cdot \left[m_t \left(\nabla V + \alpha (\cdot - \mu_1(m_t)) \right) \right]. \quad (3.14)$$

When V is a double well potential, the latter equation may have several stationary solutions. For instance, if $\sigma = 0$ (in addition to the assumption $\sigma_0 = 0$), any critical point x_0 of V induces a stationary solution, concentrated at x_0 . In the standard example where $d = 1$ and $V(x) = |x|^4/4 - |x|^2/2$ (see for instance [Daw83, HT10]), those stationary solutions are the Dirac masses δ_1 , δ_0 and δ_{-1} . Non-uniqueness persists when σ is strictly positive but small: with the same potential V and for σ less than a certain threshold $\bar{\sigma}$, there are three invariant measures, respectively centered around 1, 0 and -1 , to the McKean-Vlasov SDE

$$dX_t = -V'(X_t)dt - \alpha(X_t - \mathbb{E}_1[X_t])dt + \sigma dB_t, \quad t \geq 0.$$

Lack of uniqueness can be explained as follows. There is a competition between the noise B (weighted by the intensity factor σ) and the interaction term function $(x, m) \mapsto -\alpha(x - \mu_1(m))$. On the one hand, the process X is attracted by the minimizers of V and by its expectation. On the other hand, it is subjected to the diffusive effect of the noise. Whenever the coefficient σ is too small, the interaction term dominates. Any invariant measure has a small variance and must be concentrated around a minimizer of V , hence forcing the multiplicity of the stationary solutions to (3.14).

When $\sigma_0 > 0$ but $\sigma = 0$ (and under conditions similar to Assumption 5), whenever α is assumed large enough, the second author has shown in [Mai23] that the process $(m_t)_{t \geq 0}$ admits a unique invariant measure $P_0 \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$, and thus established restoration of uniqueness in this situation. In this particular case, the invariant measure is supported by Dirac masses:

$$\bar{P}_0(dm) = \int_{\mathbb{R}^d} \delta_{\delta_a}(dm) m^*(da),$$

where m^* is solution of the stationary Fokker-Planck equation $(\sigma_0^2/2)\Delta m^* + \nabla \cdot (m^* \nabla V) = 0$. Moreover, as consequence of the contracting properties of F , it can be shown that

$$v(m_t) \rightarrow 0,$$

when $t \rightarrow +\infty$, which implies that, asymptotically (in time), the solution X to (3.5) coincides with its conditional expectation given the common noise, hence justifying the shape of the measure \bar{P}_0 .

In the case $\sigma_0 > 0$ and σ small but (strictly) positive, our result extends the analysis carried out in [Mai23]. However, the long-time behaviour of $(v(m_t))_{t \geq 0}$ is different and the latter does not tend to 0 as t tends to ∞ . This difference is substantial and makes the proof significantly more challenging in our analysis. The argument is explained in the next two sections. In short, one shows that the long-time behavior of the conditional mean $(\mu_1(m_t))_{t \geq 0}$ coincides with that one of the standard SDE $dY_t = -\nabla V(Y_t)dt + \sigma_0 dB_t^0, t \geq 0$ up to an error that is small with σ and $1/\alpha$. Although the fluctuation $(X_t - \mathbb{E}_1(X_t) = X_t - \mu_1(m_t))_{t \geq 0}$ does not vanish in long time, which is the main mathematical difficulty here, the regime $0 < \sigma \ll 1$ shares many conceptual similarities with the case $\sigma = 0$. In particular, the long-time behavior of $(m_t)_{t \geq 0}$ is mostly dictated by the ergodic properties of the standard SDE obtained by putting $F \equiv 0$ and $\sigma = 0$ in (3.5). This example is typical of our study. When the value of σ is significantly large and α is relatively small, it is straightforward to establish the uniqueness of the invariant measure using coupling arguments on the idiosyncratic noise, as in [DEGZ20]. Still, it is fair to say that we do not have, at this stage, a good understanding of the behaviour of the model between these two regimes.

A variant: dynamics driven by a quadratic potential and a nonlinear first order mean field term

In this paragraph, we cook up a variant of the model covered by Theorem 3.8 that does not exactly fit the form postulated in (3.4) but that can be studied in a similar manner. This example is built up as a perturbation of an Ornstein-Uhlenbeck process. Precisely, we consider the stochastic Fokker-Planck equation

$$d_t m_t = \frac{\sigma^2 + \sigma_0^2}{2} \Delta m_t dt + \nabla \cdot [m_t (ax - f(\mu_1(m_t)))] dt - \sigma_0 \nabla m_t \cdot dB_t^0, \quad t \geq 0, \quad (3.15)$$

whose probabilistic counterpart writes

$$dX_t = -aX_t dt + f(\mathbb{E}_1[X_t]) dt + \sigma dB_t + \sigma_0 dB_t^0, \quad t \geq 0,$$

for some $a > 0$ and $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$. Equivalently, the conditional law $(\mathcal{L}_1(X_t))_{t \geq 0}$ of X given the common noise $(\mathcal{L}_1(X_t))_t$ is a solution to (3.15).

Obviously, the mean field term $f(\mathbb{E}_1[X_t])$ cannot be put in the form $F(X_t - \mathbb{E}_1[X_t])$, from which we see that this example is outside the scope of Theorem 3.8. That said, it is in fact not that far from the framework addressed in this article and, in particular, it obeys phenomena similar to those underpinning the proof of Theorem 3.8.

Actually, the long-time analysis of equation (3.15) is rather straightforward. When $\sigma_0 = 0$, any invariant measure m is solution to

$$\frac{\sigma^2}{2}\Delta m - \nabla \cdot [m(f(\mu_1(m)) - ax)] = 0, \quad x \in \mathbb{R}^d.$$

By integrating with respect to x , the previous equation leads to the condition

$$f(\mu_1(m)) = a\mu_1(m), \tag{3.16}$$

which is in fact sufficient in the following sense: once $\mu_1(m)$ has been found, the entire measure m can be defined as the invariant measure of the Ornstein-Uhlenbeck dynamics $dX_t = -a(X_t - \mu_1(m))dt + \sigma dB_t$, $t \geq 0$. Therefore, Equation (3.16) says that there exist several invariant measures as soon as f/a admits several fixed points in \mathbb{R}^d .

In presence of a common noise (i.e., $\sigma_0 > 0$), the conditional mean $(\mu_1(m_t))_{t \geq 0}$ solves the SDE

$$d\mu_1(m_t) = -a\mu_1(m_t)dt + f(\mu_1(m_t))dt + \sigma_0 dB_t^0, \quad t \geq 0, \tag{3.17}$$

and under appropriate confining conditions on the *effective* drift $x \mapsto -ax + f(x)$ (which may be compatible with the fact that (3.16) has several fixed points), (3.17) admits a unique invariant measure, say m^* (on \mathbb{R}^d), which implies that any two invariant measures on $\mathcal{P}_1(\mathbb{R}^d)$ (in the sense of Definition 3.4) must have the same marginal law by the projection $m \mapsto \mu_1(m)$. As above, the entire invariant measure (on $\mathcal{P}_1(\mathbb{R}^d)$) can be recovered by observing that $d(X_t - \mathbb{E}_1(X_t)) = -a(X_t - \mathbb{E}_1(X_t))dt + \sigma dB_t$, $t \geq 0$. This shows that the invariant measure is

$$\bar{P}(dm) = \int_{\mathbb{R}^d} \delta_{\mathcal{N}_d(\theta, [\sigma^2/(2a)]I_d)}(dm) m^*(d\theta), \tag{3.18}$$

where $\mathcal{N}_d(\theta, [\sigma^2/(2a)]I_d)$ is the d -dimensional Gaussian law with θ as mean and $\sigma^2/(2a)I_d$ as covariance (I_d standing for the d -dimensional identity matrix). In particular, there is a unique invariant measure even though (3.16) has several fixed points.

The spirit of this new example is clear: similar to the prototype addressed in the previous paragraph, its long-time analysis is in fact governed by the simpler Fokker-Planck equation (3.17) for the sole conditional mean. Here the situation is even simpler because the residual fluctuation $(X_t - \mathbb{E}_1(X_t))_{t \geq 0}$ has a trivial behavior (whilst the analysis of the residual is non-trivial in the prototype example).

Combining the two examples

For sure, the reader may wonder about a global framework that would cover both the prototype and the variant examples. While this would be indeed possible to extend in such a way the current setting, we have decided not to go up to this level of generality in order to keep the presentation and the notation at a reasonable level. In order to guess what the more general form of (3.4) should be, one first needs to understand which of the two coefficients F and G (in (3.4)) correspond respectively to $x \mapsto -ax$ and $m \mapsto f(\mu_1(m))$ in (3.15). At first sight, one may be tempted to regard $x \mapsto -ax$ as a specific example of G , but this is the wrong choice. In fact, the correct answer is to write $-ax$ as $-a(x - \mu_1(m))$, which

prompts us to associate the coefficient a in this model with the coefficient α in the prototype example. Next, we should see the remaining coefficient $m \mapsto -a\mu_1(m) + f(\mu_1(m))$ as a *new* G , with the subtlety that, in our global framework, G must be allowed to depend on both x and $\mu_1(m)$. In clear, the new version of (3.4) should be

$$d_t m_t = \nabla \cdot \left(\frac{\sigma^2 + \sigma_0^2}{2} \nabla m_t - m_t [G(\mu_1(m_t), \cdot) + F(\cdot - \mu_1(m_t))] \right) dt - \sigma_0 \nabla m_t \cdot dB_t^0, \quad (3.19)$$

for a function $G : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$. Back to the prototype example addressed in the previous paragraph, this says that the long-time behavior of $(m_t)_{t \geq 0}$ should now be compared with those of the standard SDE $dY_t = G(Y_t, Y_t)dt + \sigma_0 dB_t^0$, up to some fluctuation terms that one may expect to control properly if F is sufficiently decreasing. In this context, what truly matters are the confining properties of the *doubled* mapping $(x, x) \mapsto G(x, x)$.

We strongly believe that Theorem 3.8 could be extended to this setting, with a similar analysis. From a technical point of view, the gain would be limited: the difficulty would not come from the variable $\mu_1(m)$ in G but from the variable x (precisely because the strategy is to replace in the end the argument X_t by its conditional expectation $\mathbb{E}_1(X_t)$ in (3.5)). This difficulty is already present in the prototype example (3.13) and, in contrast, it is clear from the example (3.15) that the presence of the variable $\mu_1(m)$ in G does not raise any substantial difficulty. This explains how choice to work on (3.4) (and not on (3.19)).

An example without restoration of uniqueness

As highlighted earlier, the finite-dimensional nature of the common noise constitutes a significant limitation to obtain ergodic properties on the process $(m_t)_{t \geq 0}$ (see in contrast the recent work [DH24] by the first author for a 1d case with an infinite dimensional noise). In particular, this is the thrust of our work to identify one class of mean field dynamics for which the common noise really helps in this matter. However, it is clear that there is no chance to get uniqueness of the invariant measure for a generic class of stochastic Fokker-Planck SPDEs forced by a *simple* noise like B^0 .

A counter-example in dimension $d = 1$ is

$$d_t m_t = \frac{\sigma(v(m_t))^2 + \sigma_0^2}{2} \partial_{xx}^2 m_t dt + \partial_x(xm_t) dt - \sigma_0 \partial_x m_t dB_t^0,$$

for some function $\sigma : [0, +\infty) \rightarrow (0, +\infty)$. Whenever $\sigma_0 = 0$ and the function $\sigma^2(\cdot)/4$ admits several fixed points, the stationary solutions to the above equation cannot be unique. In this case, $(m_t)_{t \geq 0}$ is indeed solution of

$$\partial_t m_t = \frac{\sigma(v(m_t))^2}{2} \partial_{xx}^2 (m_t) + \partial_x(xm_t), \quad t \geq 0. \quad (3.20)$$

Then, similar to the stationary solutions found in the previous example, the stationary solutions (equivalently, the invariant measures of (3.5)) are here solutions of

$$\frac{dm}{dx}(x) = \exp\left(-2 \frac{x^2}{\sigma^2(v(m))}\right), \quad x \in \mathbb{R}.$$

Then, one can show that, if the function $x \mapsto \sigma^2(x)/4$ admits several fixed points, then there are several stationary solutions to (3.20).

When adding common noise, the situation does not get better. For instance, assuming that the initial condition $P_0 = \delta_{\delta_0}$, one can show that for all $t > 0$, \mathbb{P}_0 -almost surely, m_t is a (random) Gaussian probability measure with parameters

$$\begin{cases} \mu_1(m_t) = \sigma_0 \int_0^t e^{s-t} dB_s^0, \\ v(m_t) = \int_0^t e^{2(s-t)} \sigma^2(v(m_s)) ds. \end{cases}$$

Then, it is not difficult to see that $(\mu_1(m_t))_{t \geq 0}$ is an ergodic process (it is in fact an Ornstein-Uhlenbeck process with a negative mean-reverting parameter). As for the dynamics of the variance $(v(m_t))_{t \geq 0}$, we get

$$\partial_t v(m_t) = -2v(m_t) + \frac{1}{2} \sigma^2(v(m_t)), \quad t \geq 0.$$

Obviously, the above equation has several stationary solutions (say σ^*) if the function $x \mapsto \sigma^2(x)/4$ admits several fixed points. In the latter case, $(m_t)_{t \geq 0}$ has several invariant measures, whose form is similar to (3.18) except that σ therein is now replaced by any σ^* . In this specific example, it is noteworthy that the common noise has no influence on the mean field term $(v(m_t))_{t \geq 0}$ (which is completely deterministic). Clearly, this is a consequence of the additive structure of the noise, the effect of which is just to shift (or to translate) the measures $(m_t)_{t \geq 0}$.

Conclusion

All these examples illustrate that the form of the mean field interaction in (3.4), based on the sole (conditional) mean state is key in the derivation of Theorem 3.8. In particular, this structure makes it possible to transmit the noise from the dynamics of the state variable $(X_t)_{t \geq 0}$ in (3.5) to its conditional mean $(\mathbb{E}_1(X_t))_{t \geq 0}$. As suggested in the discussion on the prototype example, the next step is to prove that the long run behavior of $(\mathbb{E}_1(X_t))_{t \geq 0}$ is dictated by the ergodic properties of the SDE $dY_t = G(Y_t)dt + \sigma_0 dB_t^0$, at least if α is large and σ is small. This is the main line of the proof that is presented in the next two sections.

3 Proof of Theorem 3.8

In this section, we explain how to deduce Theorem 3.8 from Proposition 3.7, but the proof of the latter is postponed to Section 4.

Before moving on to the proof of Theorem 3.8, we need two preliminary estimates: the first one addresses the conditional variance of the solution to (3.4), and the second one the conditional exponential moments of the solution.

3.1 First preliminary estimate: conditional variance

Given a solution X to Equation (3.5), we first focus on the conditional expectation given the common

noise $(\mathbb{E}_1[X_t])_{t \geq 0}$:

Proposition 3.10. *Let us consider a d -dimensional Brownian motion β^0 defined on $(\Omega_0, \mathcal{F}^0, \mathbb{P}_0)$ and adapted to the filtration \mathbb{F}^0 . Let Assumptions 5 and 6 be in force with the additional constraint $c_{\alpha_F} := \alpha_F - L_G > 0$. Let us consider $P_0 \in \mathcal{P}(\mathcal{P}_1(\mathbb{R}^d))$ and X_0 such that $\mathcal{L}_1(X_0) = m_0$ a.s. for some random measure m_0 satisfying $\mathcal{L}_0(m_0) = P_0$. Then, for X with dynamic given by*

$$\begin{cases} dX_t = G(X_t)dt + F(X_t - \mathbb{E}_1[X_t])dt + \sigma dB_t + \sigma_0 d\beta_t^0 \\ X_{t=0} = X_0, \end{cases}$$

we have for all $t \geq 0$,

$$\mathbb{E}_1 [|X_t - \mathbb{E}_1[X_t]|^2] \leq \mathbb{E}_1 [|X_0 - \mathbb{E}_1[X_0]|^2] e^{-2c_{\alpha_F} t} + \frac{d\sigma^2}{c_{\alpha_F}}, \quad (3.21)$$

where the previous inequality holds $\mathbb{P}_0 - a.s.$

Remark 3.11. (1) *Inequality (3.21) does not depend on the choice of the Brownian motion $(\beta_t^0)_{t \geq 0}$ in the dynamics of X . Accordingly, it can be recast in terms of the solution (3.4) in the form*

$$v(m_t) \leq v(m_0) e^{-2c_{\alpha_F} t} + \frac{d\sigma^2}{c_{\alpha_F}}, \quad t \geq 0. \quad (3.22)$$

(2) *The result can be interpreted as follows. When $\sigma = 0$, the result is consistent with [Mai23] and says that the process $(X_t)_{t \geq 0}$ is attracted by its conditional expectation in the long run. When $\sigma > 0$, the result provides a sharp estimate of the residual conditional variance.*

(3) *Back to Theorem 3.8, the challenge is precisely to prove that the long-time behaviour of X is dictated by the long-time behaviour of $\mathbb{E}_1[X]$ even though the residual term is not zero.*

Proof of Proposition 3.10. Using the dynamics of the process X , one can write

$$d(X_t - \mathbb{E}_1[X_t]) = (G(X_t) - \mathbb{E}_1[G(X_t)])dt + (F(X_t - \mathbb{E}_1[X_t]) - \mathbb{E}_1[F(X_t - \mathbb{E}_1[X_t])])dt + \sigma dB_t$$

Then, applying Itô's formula, it comes

$$\begin{aligned} d|X_t - \mathbb{E}_1[X_t]|^2 &= 2(X_t - \mathbb{E}_1[X_t]) \cdot (G(X_t) - \mathbb{E}_1[G(X_t)])dt \\ &\quad + 2(X_t - \mathbb{E}_1[X_t]) \cdot (F(X_t - \mathbb{E}_1[X_t]) - \mathbb{E}_1[F(X_t - \mathbb{E}_1[X_t])])dt \\ &\quad + 2\sigma(X_t - \mathbb{E}_1[X_t]) \cdot dB_t \\ &\quad + d\sigma^2 dt \end{aligned}$$

Using the fact that G is Lipschitz continuous and taking expectation, we get that

$$\frac{d}{dt} \mathbb{E}_1 [|X_t - \mathbb{E}_1[X_t]|^2] \leq 2L_G \mathbb{E}_1 [|X_t - \mathbb{E}_1[X_t]|^2] + 2\mathbb{E}_1 [(X_t - \mathbb{E}_1[X_t]) \cdot F(X_t - \mathbb{E}_1[X_t])] + d\sigma^2.$$

Moreover, we can write

$$\begin{aligned} \mathbb{E}_1 [(X_t - \mathbb{E}_1[X_t]) \cdot F(X_t - \mathbb{E}_1[X_t])] &= \mathbb{E}_1 [(X_t - \mathbb{E}_1[X_t]) \cdot (F(X_t - \mathbb{E}_1[X_t]) - F(0))] \\ &\leq -\alpha_F \mathbb{E}_1 [|X_t - \mathbb{E}_1[X_t]|^2], \end{aligned}$$

where we recall that $\alpha_F > 0$ is defined in Assumption 6. Then,

$$\frac{d}{dt} \mathbb{E}_1 [|X_t - \mathbb{E}_1[X_t]|^2] \leq -2(\alpha_F - L_G) \mathbb{E}_1 [|X_t - \mathbb{E}_1[X_t]|^2] + d\sigma^2.$$

Finally, we get the result using Grönwall's Lemma, as soon as $\alpha_F > L_G$. \square

3.2 Second preliminary estimate: conditional exponential moments

Let us state the following lemma which allows us to control the exponential moments of the process $(m_t)_{t \geq 0}$ under one invariant measure \bar{P} (which will be in the end 'the' invariant measure):

Lemma 3.12. *Under Assumptions 5 and 6, assuming that $\alpha_F > L_G$, there exists an invariant measure $\bar{P} \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$ for the process $(m_t)_{t \geq 0}$, such that*

$$\int_{\mathcal{P}(\mathbb{R}^d)} \exp\left(2c_2 v(\tilde{m})^{1/2}\right) \bar{P}(d\tilde{m}) < +\infty, \quad (3.23)$$

where $c_2 := c_1^{\alpha_F} / c_{\alpha_F}$.

Remark 3.13. *The result presented in Lemma 3.12 aligns with the global picture we gave for the behavior of the model. In particular, it is natural (and in fact well-expected) in the regime $\sigma \ll 1$ (which corresponds to the framework of Proposition 3.7). Indeed, we already know from [Mai23] (see also Section 2.4) that, if $\sigma = 0$, the invariant measure \bar{P}^0 is supported by Dirac masses; therefore, the conclusion of Lemma 3.12 becomes straightforward when replacing \bar{P} by \bar{P}^0 in the statement. The thrust of Lemma 3.12 is thus to extend the result to positive values of σ . Although our proof relies on a direct computation, it is worth noticing that any invariant measure \bar{P} (for $\sigma > 0$) should satisfy*

$$d_1^{\mathcal{P}(\mathbb{R}^d)}(\bar{P}, \bar{P}^0) \leq C\sigma^2,$$

for some constant C independent of σ . In particular, any invariant \bar{P} is close to \bar{P}^0 when σ is small, hence justifying our intuition that Lemma 3.12 is a perturbation of the regime $\sigma = 0$.

Proof of Lemma 3.12. Consider a probability measure $P_0 \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ with $P_0(\mathcal{P}_2(\mathbb{R}^d)) = 1$ and the measure valued stochastic process $(m_t)_{t \geq 0}$ solving (3.4) with $\mathcal{L}_0(m_0) = P_0$ as initial condition. Then, letting $P_t = \mathcal{L}_0(m_t)$, $\forall t \geq 0$, we can easily show with arguments similar to those used in the proof of Proposition 2 in [Mai23] that the sequence $(Q_T)_{T \geq 0}$ defined by

$$Q_T := \frac{1}{T} \int_0^T P_t dt$$

admits a converging subsequence, still denoted by $(Q_T)_{T \geq 0}$, that converges to an invariant measure \bar{P} .

For any $T > 0$,

$$\begin{aligned} \int_{\mathcal{P}(\mathbb{R}^d)} \exp\left(2c_2v(\tilde{m})^{1/2}\right) \mathcal{Q}_T(d\tilde{m}) &= \frac{1}{T} \int_0^T \int_{\mathcal{P}(\mathbb{R}^d)} \exp\left(2c_2v(\tilde{m})^{1/2}\right) P_t(d\tilde{m}) \\ &= \frac{1}{T} \int_0^T \mathbb{E}_0 \left[\exp\left(2c_2v(m_t)^{1/2}\right) \right] dt. \end{aligned}$$

Moreover, (3.22) gives

$$\begin{aligned} \int_{\mathcal{P}(\mathbb{R}^d)} \exp\left(2c_2v(\tilde{m})^{1/2}\right) \mathcal{Q}_T(d\tilde{m}) &\leq \frac{1}{T} \int_0^T \mathbb{E}_0 \left[\exp\left(2c_2v(m_0)^{1/2} + \frac{2c_2\sigma}{\sqrt{c_{\alpha_F}}}\right) \right] dt \\ &\leq \mathbb{E}_0 \left[\exp\left(2c_2v(m_0)^{1/2} + \frac{2c_2\sigma}{\sqrt{c_{\alpha_F}}}\right) \right], \end{aligned}$$

which can be assumed to be finite by choosing an appropriate initial condition. Recall indeed that the choice of P_0 in the construction of the sequence $(\mathcal{Q}_T)_{T \geq 0}$ is free. Therefore, assuming that

$$\int_{\mathcal{P}(\mathbb{R}^d)} \exp(2c_2v(\tilde{m})^{1/2}) P_0(d\tilde{m}) < +\infty, \quad (3.24)$$

and using lower-semi-continuity of the function $P \mapsto \langle \exp(2c_2v(\cdot)^{1/2}); P \rangle_{\mathcal{P}(\mathbb{R}^d)}$, we get that

$$\int_{\mathcal{P}(\mathbb{R}^d)} \exp\left(2c_2v(\tilde{m})^{1/2}\right) \bar{P}(d\tilde{m}) \leq \liminf_{T \rightarrow +\infty} \int_{\mathcal{P}(\mathbb{R}^d)} \exp\left(2c_2v(\tilde{m})^{1/2}\right) \mathcal{Q}_T(d\tilde{m}) < +\infty, \quad (3.25)$$

which is (3.23). In order to prove that $\bar{P} \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$, we use the same approach and get for any $T > 0$,

$$\begin{aligned} \int_{\mathcal{P}(\mathbb{R}^d)} \mu_2(\tilde{m}) \mathcal{Q}_T(d\tilde{m}) &= \frac{1}{T} \int_0^T \int_{\mathcal{P}(\mathbb{R}^d)} \mu_2(\tilde{m}) P_t(d\tilde{m}) dt \\ &= \frac{1}{T} \int_0^T \mathbb{E}_0[\mu_2(m_t)] dt. \end{aligned} \quad (3.26)$$

Now, we claim that, under (3.24) and Assumptions 5 and 6,

$$\sup_{t \geq 0} \mathbb{E}_0[\mu_2(m_t)] < +\infty.$$

Indeed, there exist two positive constants $b_1, b_2 > 0$, such that for all $(x, y) \in \mathbb{R}^d$,

$$(x - y) \cdot (G(x) - G(y)) \leq -b_1|x - y|^2 + b_2.$$

Then, we get

$$\begin{aligned} \frac{d}{dt} \mathbb{E}_0[\mu_2(m_t)] &\leq -2b_1 \mathbb{E}_0[\mu_2(m_t)] + 2b_2 + (\sigma^2 + \sigma_0^2) \\ &\quad + 2|G(0)| \mathbb{E}_0[\mu_2(m_t)]^{1/2} + 2C_F \mathbb{E}_0 \left[\mu_2(m_t)^{1/2} v(m_t)^{1/2} \right]. \end{aligned}$$

Using once again (3.22) and Jensen's inequality to control the last term, we obtain

$$\begin{aligned} \frac{d}{dt} \mathbb{E}_0[\mu_2(m_t)] &\leq -2b_1 \mathbb{E}_0[\mu_2(m_t)] + 2 \left(|G(0)| + C_F v(m_0)^{1/2} + \frac{C_F \sigma}{\sqrt{c_{\alpha_F}}} \right) \mathbb{E}_0[\mu_2(m_t)]^{1/2} \\ &\quad + 2b_2 + d(\sigma^2 + \sigma_0^2) \\ &\leq -b_1 \mathbb{E}_0[\mu_2(m_t)] + b_1^{-1} \left(|G(0)| + C_F v(m_0)^{1/2} + \frac{C_F \sigma}{\sqrt{c_{\alpha_F}}} \right)^2 + 2b_2 + d(\sigma^2 + \sigma_0^2), \end{aligned}$$

and conclude using Grönwall's Lemma that $\sup_{t \geq 0} \mathbb{E}_0[\mu_2(m_t)] < +\infty$. Finally, using the same lower-semi-continuity argument as in (3.25) and returning to (3.26), we get

$$\int_{\mathcal{P}(\mathbb{R}^d)} \mu_2(\tilde{m}) \bar{P}(d\tilde{m}) \leq \lim_{T \rightarrow +\infty} \int_{\mathcal{P}(\mathbb{R}^d)} \mu_2(\tilde{m}) \mathcal{Q}_T(d\tilde{m}) < +\infty. \quad (3.27)$$

The latter implies that $\bar{P} \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$. \square

3.3 Completion of the proof of Theorem 3.8

We are now ready for the proof of Theorem 3.8 which is a straightforward combination of Proposition 3.7 and Lemma 3.12.

Proof of Theorem 3.8. We know from Proposition 3.7 that for any $\alpha_F \geq \max(L_G, m_G)$ and any $\sigma_0 > 0$, there exist $\bar{\sigma} > 0$ (depending on F, G, σ_0 and d) such that for all $\sigma \leq \bar{\sigma}$ and for any 1-Lipschitz continuous function $\phi : \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}$ and all $m, \tilde{m} \in \mathcal{P}_2(\mathbb{R}^d)$,

$$|\mathcal{P}_t \phi(m) - \mathcal{P}_t \phi(\tilde{m})| \leq \mathcal{K}(m, \tilde{m}) e^{-ct}, \quad t \geq 0,$$

for some $c > 0$ independent of m and \tilde{m} . Let us now consider an invariant measure $\bar{P} \in \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$, with a finite conditional exponential moments, whose existence is ensured by Lemma 3.12. Then,

$$\begin{aligned} \left| \mathcal{P}_t \phi(m) - \int_{\mathcal{P}(\mathbb{R}^d)} \mathcal{P}_t \phi(\tilde{m}) \bar{P}(d\tilde{m}) \right| &\leq \int_{\mathcal{P}(\mathbb{R}^d)} |\mathcal{P}_t \phi(m) - \mathcal{P}_t \phi(\tilde{m})| \bar{P}(d\tilde{m}) \\ &\leq e^{-ct} \int_{\mathcal{P}(\mathbb{R}^d)} \mathcal{K}(m, \tilde{m}) \bar{P}(d\tilde{m}). \end{aligned} \quad (3.28)$$

Now, coming back to Proposition 3.7 for the shape of $\mathcal{K}(m, \tilde{m})$, we obtain that

$$\begin{aligned} \int_{\mathcal{P}(\mathbb{R}^d)} \mathcal{K}(m, \tilde{m}) \bar{P}(d\tilde{m}) &\leq C_d \mu_2(m)^{1/2} \exp(c_2 v(m)^{1/2}) \int_{\mathcal{P}(\mathbb{R}^d)} \exp(c_2 v(m)^{1/2}) \bar{P}(d\tilde{m}) \\ &\quad + C_d \exp(c_2 v(\tilde{m})^{1/2}) \int_{\mathcal{P}(\mathbb{R}^d)} \mu_2(\tilde{m})^{1/2} \exp(c_2 v(\tilde{m})^{1/2}) \bar{P}(d\tilde{m}), \end{aligned}$$

with $c_2 := c_1^{\alpha_F} / c_{\alpha_F}$.

Using Hölder inequality and leveraging on Lemma 3.12 we have

$$\int_{\mathcal{P}(\mathbb{R}^d)} \mathcal{K}(m, \tilde{m}) \bar{P}(d\tilde{m}) \leq C_d (1 + \mu_2(m)^{1/2}) \exp(c_2 v(m)^{1/2}).$$

Substituting in (3.28), we complete the proof. \square

4 Proof of Proposition 3.7

This section is dedicated to the proof of Proposition 3.7.

4.1 Ansatz

The strategy of proof of Proposition 3.7 relies on the auxiliary Proposition 3.10. This latter result says that the residual conditional variance of X (solution to Equation (3.5)) is small with σ . In turn, it prompts us to focus on the process $\mathbb{E}_1[X] = (\mathbb{E}_1[X_t])_{t \geq 0}$. The latter has the following dynamics:

$$d\mathbb{E}_1[X_t] = \mathbb{E}_1[G(X_t)]dt + \mathbb{E}_1[F(X_t - \mathbb{E}_1[X_t])]dt + \sigma_0 dB_t^0, \quad t \geq 0.$$

Whilst this is not a closed equation (in $\mathbb{E}_1[X_t]$), our approach is to write it as a perturbation of a finite-dimensional stochastic differential equation. Heuristically, we have (thanks to Proposition 3.10) that on a large time scale

$$\mathbb{E}_1[F(X_t - \mathbb{E}_1[X_t])] \approx \sigma^2/c_{\alpha_F}, \quad (3.29)$$

$$\mathbb{E}_1[G(X_t)] = G(\mathbb{E}_1[X_t]) + [\mathbb{E}_1[G(X_t)] - G(\mathbb{E}_1[X_t])] \approx G(\mathbb{E}_1[X_t]) + \sigma^2/c_{\alpha_F}. \quad (3.30)$$

Equations (3.29) and (3.30) are rather informal, and we refrain from giving here a rigorous meaning to the symbol \approx in both of them. However, these two expansions suggest that the dynamics of $(\mathbb{E}_1[X_t])_{t \geq 0}$ indeed resembles to the one of the diffusion process Y , constructed on $(\Omega_0, \mathbb{F}^0, \mathbb{P}_0)$ as the solution of

$$dY_t = G(Y_t) dt + \sigma_0 dB_t^0, \quad t \geq 0; \quad Y_0 = \mathbb{E}_1(X_0). \quad (3.31)$$

Here, it is worth emphasizing that, even obvious, (3.31) is not a McKean-Vlasov but a mere diffusion equation. In particular, the long time analysis of (3.31) falls within a much wider literature, since the study of the long time behaviour of diffusion processes has been an important topic of interest, see [ABC⁺00, BBCG08, BE85] to name just a few. Below, we make use of coupling arguments, in the spirit of Eberle [Ebe16], who stated contraction properties for the law of the solution of (3.31) by building on ideas developed earlier in [LR86]. The next section gives some details on this approach and the contraction results available in this framework.

4.2 Contraction for classical diffusion processes

Following [Ebe16], we introduce the following quantities :

$$\begin{aligned} R_0 &:= \inf\{R \geq 0 : \kappa(r) \geq 0, \quad \forall r \geq R\}, \\ R_1 &:= \inf\{R \geq R_0 : \kappa(r)R(R - R_0) \geq 8, \quad \forall r \geq R\}, \end{aligned}$$

so that $R_0 \leq R_1$. Thanks to Assumption 5, R_0 and R_1 are finite. Moreover, let

$$\begin{aligned}\varphi(r) &:= \exp\left(-\frac{1}{4} \int_0^r s \kappa(s)^- ds\right), \quad \Phi(r) := \int_0^r \varphi(s) ds, \\ g(r) &:= 1 - \left(\int_0^{r \wedge R_1} \Phi(s) \varphi(s)^{-1} ds\right) \left(2 \int_0^{R_1} \Phi(s) \varphi(s)^{-1} ds\right)^{-1},\end{aligned}$$

and

$$f(r) := \int_0^r \varphi(s) g(s) ds. \quad (3.32)$$

Then, we have that

- φ is non-increasing, $\varphi(0) = 1$, and $\varphi(r) = \varphi(R_0)$ for any $r \geq R_0$;
- g is non-increasing, $g(0) = 1$, and $g(r) = 1/2$ for any $r \geq R_1$;
- f' is bounded on $[0, +\infty)$;
- f is concave, $f(0) = 0$, $f'(0) = 1$, and

$$\Phi(r)/2 \leq f(r) \leq \Phi(r) \quad \text{for any } r \geq 0.$$

The last display combined with the fact that for all $r \geq 0$, $\Phi(r)/2 \geq \varphi(R_0)r/2$ and $\Phi(r) \leq r$, implies

$$\kappa_1 r \leq f(r) \leq r, \quad (3.33)$$

with $\kappa_1 := \varphi(R_0)/2$. This allows us to define the equivalent deformation of 1-Wasserstein distance

$$d_f^{\mathbb{R}^d}(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \int f(|x - y|) \pi(dx, dy),$$

for any two probability measures μ, ν on \mathbb{R}^d , where the infimum is taken over all couplings π of μ and ν . Last but not least, the function f we constructed satisfies

$$f''(r) - \frac{1}{4} r \kappa(r) f'(r) \leq -\frac{c}{2\sigma_0^2} f(r) \quad \text{for all } r > 0, \quad (3.34)$$

and for some $c > 0$. Such concentration property for the function f is classical, and the proof of inequality (3.34) can be found in [Con23, DEGZ20, Ebe16]. Leveraging on the contraction property of f given by Equation (3.34), we get the following result (as a direct consequence of [Ebe16]):

Theorem 3.14 (Corollary 2 in [Ebe16]). *Under Assumption 5 and with $\sigma_0 > 0$, there exists a constant $c := c_{\sigma_0} > 0$, such that for any pair (Y, \tilde{Y}) of solutions to Equation (3.31) with different initial conditions, we have*

$$d_f^{\mathbb{R}^d}(\mathcal{L}_1(Y_t), \mathcal{L}_1(\tilde{Y}_t)) \leq d_f^{\mathbb{R}^d}(\mathcal{L}_1(Y_0), \mathcal{L}_1(\tilde{Y}_0)) e^{-ct}, \quad t \geq 0. \quad (3.35)$$

Remark 3.15. Thanks to (3.33), we get that inequality (3.35) holds (up to some constant) when replacing $d_f^{\mathbb{R}^d}$ by the classical $d_1^{\mathbb{R}^d}$ Wasserstein distance.

The idea of the proof of Theorem 3.14 is to use the reflection coupling introduced in [LR86] and widely used in the literature [Con23, DEGZ20, Ebe16, EZ19]. The classical reflection coupling is constructed by multiplying the noise by a symmetry matrix mapping instantaneously $Y_t - \tilde{Y}_t$ on its opposite, see [LR86, Section 3] for details. In this framework, the distance between the two processes Y and \tilde{Y} evolves according to an Itô process with constant noise intensity. The hope is thus to benefit from the presence of the noise to obtain further recurrent properties that prevent the processes from staying away from one another. For sure, one cannot expect this intuition to hold true for any type of drift; in order to guarantee a relevant form of recurrence, some further confining properties are necessary.

When dealing with the stochastic differential equation (3.31), the condition $\limsup_{r \rightarrow +\infty} \kappa(r) > 0$ required in Assumption 5 is key in the verification of the latter confining properties. In itself, this is not an obvious result: one must indeed keep in mind that G is not strictly decreasing, meaning that κ may be negative on some part of the space. Clearly, things become even more subtle in presence of the conditional McKean-Vlasov interaction. Still, one would like to settle down a similar reflection argument for Equation (3.5), which raises preliminary questions of measurability. Briefly, in order to preserve the form of the common noise, one must consider a reflection matrix that is measurable with respect to the latter (and not to the idiosyncratic noise). The idea developed in the next subsection is to introduce a coupling (X, \tilde{X}) of solutions to (3.5) in which the reflection is achieved by reflecting with respect to $\mathbb{E}_1[X_t] - \mathbb{E}_1[\tilde{X}_t]$. However, while reflecting with respect to the conditional expectations may seem appropriate to our objective, it cannot suffice on its own to get a proof of Proposition 3.7 that would be a mere copy and paste of the argument of [Ebe16]. The difficulty that one meets when adapting [Ebe16] to (3.5) is clear: the processes that solve (3.5) are not those that are used in the reflection. This is where Proposition 3.10 comes in and this is one of the main innovation of our work: we use the fact that the distance between the solutions to (3.5) and their conditional expectations (which are also the processes entering the reflection) is small with σ . Mathematically, the difficulty is to revisit the whole machinery of [Ebe16] by using the fact that, although it is not zero, the latter distance is small.

4.3 Proof of Proposition 3.7

To ease the reading of this subsection, we briefly outline the strategy of proof:

Proof Outline.

- *Step 1.* We construct a coupling (X, \tilde{X}) inspired by the reflection coupling introduced in [LR86] (as explained in subsection 4.2) and adapted to the presence of a common noise.
- *Steps 2, 3 & 4.* We prove that this coupling allows the conditional expectations $\mathbb{E}_1[X]$ and $\mathbb{E}_1[\tilde{X}]$ to get closer in time, up to some residual error that depends on the distance between the centred processes $X - \mathbb{E}_1[X]$ and $\tilde{X} - \mathbb{E}_1[\tilde{X}]$. This is established thanks to the confining property of G guaranteed by Assumption 5.

- *Step 5.* In parallel (the key point is to run the two arguments at the same time), we prove the mirror result, namely we show that the centred processes $X - \mathbb{E}_1[X]$ and $\tilde{X} - \mathbb{E}_1[\tilde{X}]$ get closer in time, up to some residual error that depends on the distance between the conditional expectations $\mathbb{E}_1[X]$ and $\mathbb{E}_1[\tilde{X}]$. This is established thanks to the strictly contracting properties of the interaction term F guaranteed by Assumption 6.
- *Step 6.* We combine together the results of Step 4 and Step 5 to conclude.

Proof. Step 1. Let us consider a bounded and $d_1^{\mathbb{R}^d}$ -Lipschitz continuous function $\phi : \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}$. In particular there exists $\|\phi\|_{\text{lip}} > 0$ such that for any $m, \tilde{m} \in \mathcal{P}_1(\mathbb{R}^d)$,

$$|\phi(m) - \phi(\tilde{m})| \leq \|\phi\|_{\text{lip}} d_1^{\mathbb{R}^d}(m, \tilde{m}).$$

Let us then fix $m, \tilde{m} \in \mathcal{P}_2(\mathbb{R}^d)$ together with two independent random variables X_0, \tilde{X}_0 , defined on $(\Omega_1, \mathbb{F}^1, \mathbb{P}_1)$ such that $\mathcal{L}_1(X_0) = m$ and $\mathcal{L}_1(\tilde{X}_0) = \tilde{m}$. Then, we extend X_0 and \tilde{X}_0 to the product space $(\Omega, \mathbb{F}, \mathbb{P})$ in a natural manner.

Remark 3.16. *It is important to notice here that, even if the initial conditions X_0, \tilde{X}_0 are defined on the product space $(\Omega, \mathbb{F}, \mathbb{P})$, they are independent of \mathcal{F}^0 .*

For a given $\delta > 0$, let us now introduce the following coupling of two solutions of the SDE (3.5) with initial conditions X_0 and \tilde{X}_0 :

$$\begin{cases} dX_t^\delta = G(X_t^\delta)dt + F\left(X_t^\delta - \mathbb{E}_1[X_t^\delta]\right) dt + \sigma dB_t + \sigma_0 \left\{ \pi_\delta(E_t^\delta) dB_t^0 + \lambda_\delta(E_t^\delta) d\tilde{B}_t^0 \right\}, \\ d\tilde{X}_t^\delta = G(\tilde{X}_t^\delta)dt + F\left(\tilde{X}_t^\delta - \mathbb{E}_1[\tilde{X}_t^\delta]\right) dt + \sigma dB_t \\ \quad + \sigma_0 \left\{ \pi_\delta(E_t^\delta) \left(\text{Id} - 2e_t^\delta (e_t^\delta)^\top \right) dB_t^0 + \lambda_\delta(E_t^\delta) d\tilde{B}_t^0 \right\}, \end{cases} \quad (3.36)$$

where

- the initial conditions are given by $X_0^\delta = X_0$ and $\tilde{X}_0^\delta = \tilde{X}_0$;
- \tilde{B}^0 is a d -dimensional Brownian motion defined on $(\Omega_0, \mathcal{F}^0, \mathbb{P}_0)$ adapted to the filtration \mathbb{F}^0 and independent of B^0 ;
- for all $t \geq 0$, $E_t^\delta := \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]$, and

$$e_t^\delta := \begin{cases} E_t^\delta / |E_t^\delta| & \text{if } |E_t^\delta| \neq 0 \\ 0 & \text{otherwise;} \end{cases}$$

- the function π_δ is Lipschitz continuous on \mathbb{R}^d with value in $[0, 1]$, such that for any $x \in \mathbb{R}^d$,

$$\pi_\delta(x) = \begin{cases} 0 & \text{if } |x| \leq \delta/2 \\ 1 & \text{if } |x| \geq \delta. \end{cases}$$

- the function λ_δ is Lipschitz-continuous and satisfies $\lambda_\delta^2(x) = 1 - \pi_\delta^2(x)$, for all $x \in \mathbb{R}^d$.

The first thing to notice is that under the Lipschitz continuity assumptions on G and F , and the integrability properties of the initial conditions, we have existence and uniqueness of a weak solution of (3.36). Moreover, thanks to the choice of π_δ and λ_δ , the pair $(X^\delta, \tilde{X}^\delta)$ is a coupling of (Z, \tilde{Z}) where for Z and \tilde{Z} admits the same dynamic,

$$dZ_t = G(Z_t)dt + F(Z_t - \mathbb{E}_1[Z_t])dt + \sigma dB_t + \sigma_0 dB_t^0, \quad (3.37)$$

with initial condition $Z_0 = X_0$ and $\tilde{Z}_0 = \tilde{X}_0$. This is mainly due to Levy's characterisation of the Brownian motion which is applicable here as for all $x \in \mathbb{R}^d$, $\lambda_\delta^2(x) + \pi_\delta^2(x) = 1$. Details on the coupling method are given in subsection 4.2. Below, we focus on the long-time behavior of

$$\mathbb{E}_0[\mathbb{E}_1[|X_t^\delta - \tilde{X}_t^\delta|]],$$

for a given $\delta > 0$.

Step 2. Letting $A_t^\delta := X_t^\delta - \mathbb{E}_1[X_t^\delta]$ and $\tilde{A}_t^\delta := \tilde{X}_t^\delta - \mathbb{E}_1[\tilde{X}_t^\delta]$, we get that for all $t \geq 0$,

$$|X_t^\delta - \tilde{X}_t^\delta| \leq |\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]| + |A_t^\delta - \tilde{A}_t^\delta|.$$

The idea of the proof is to control both $|\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]|$ and $|A_t^\delta - \tilde{A}_t^\delta|$ simultaneously. This will allow us to prove that both terms converge in fact to 0 (in some sense). Let us begin with the difference of the conditional expectations:

$$\begin{aligned} d\left(\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]\right) &= \mathbb{E}_1[G(X_t^\delta) - G(\tilde{X}_t^\delta)]dt + \mathbb{E}_1[F(A_t^\delta) - F(\tilde{A}_t^\delta)]dt \\ &\quad + 2\sigma_0\pi_\delta(E_t^\delta)e_t^\delta \cdot dB_t^0. \end{aligned}$$

Using Itô's formula and recalling that $E_t^\delta = \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]$ and $e_t^\delta \cdot E_t^\delta = |E_t^\delta|$, we obtain

$$\begin{aligned} d\left|\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]\right|^2 &= 2\left(\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]\right) \cdot \left(\mathbb{E}_1[G(X_t^\delta)] - \mathbb{E}_1[G(\tilde{X}_t^\delta)]\right) dt \\ &\quad + 2\left(\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]\right) \cdot \left(\mathbb{E}_1[F(A_t^\delta)] - \mathbb{E}_1[F(\tilde{A}_t^\delta)]\right) dt \\ &\quad + 4\sigma_0\left|\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]\right|\pi_\delta(E_t^\delta)e_t^\delta \cdot dB_t^0 \\ &\quad + 2\sigma_0^2\pi_\delta^2(E_t^\delta)dt. \end{aligned}$$

This combined with the particular shape of the reflection, and noting that $(\int_0^t e_s^\delta \cdot dB_s^0)_t$ is a 1-dimensional Brownian motion tank's to Levy's characterisation allows us to write

$$\begin{aligned} d\left|\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]\right| &= e_t^\delta \cdot \left(\mathbb{E}_1[G(X_t^\delta)] - \mathbb{E}_1[G(\tilde{X}_t^\delta)]\right) dt \\ &\quad + e_t^\delta \cdot \left(\mathbb{E}_1[F(A_t^\delta)] - \mathbb{E}_1[F(\tilde{A}_t^\delta)]\right) dt \\ &\quad + 2\sigma_0\pi_\delta(E_t^\delta)e_t^\delta \cdot dB_t^0, \quad t \geq 0. \end{aligned} \quad (3.38)$$

Deriving Equation (3.38) is not straightforward and can be done using similar techniques as in [DEGZ20]. The underlying concept is that the presence of π_δ keeps the process E^δ from lingering near zero. The complete proof of (3.38) can be found in Appendix 5. We now recall (see Assumption 5) that κ is a continuous function on $(0, \infty)$ satisfying

$$\liminf_{r \rightarrow \infty} \kappa(r) > 0 \quad \text{and} \quad \int_0^1 r \kappa(r)^- dr < \infty. \quad (3.39)$$

Then, using Itô's formula once again, we obtain

$$\begin{aligned} df \left(\left| \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] \right| \right) &= f' \left(\left| \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] \right| \right) e_t^\delta \cdot \left(\mathbb{E}_1[G(X_t^\delta)] - \mathbb{E}_1[G(\tilde{X}_t^\delta)] \right) dt \\ &\quad + f' \left(\left| \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] \right| \right) e_t^\delta \cdot \left(\mathbb{E}_1[F(A_t^\delta)] - \mathbb{E}_1[F(\tilde{A}_t^\delta)] \right) dt \\ &\quad + 2\sigma_0 f' \left(\left| \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] \right| \right) \pi_\delta(E_t^\delta) e_t^\delta \cdot dB_t^0 \\ &\quad + 2\sigma_0^2 f'' \left(\left| \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] \right| \right) \pi_\delta^2(E_t^\delta) dt. \end{aligned}$$

Then, we rewrite

$$\begin{aligned} df \left(\left| \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] \right| \right) &= f' \left(\left| \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] \right| \right) e_t^\delta \cdot \left(G(\mathbb{E}_1[X_t^\delta]) - G(\mathbb{E}_1[\tilde{X}_t^\delta]) \right) dt \\ &\quad + 2\sigma_0^2 f'' \left(\left| \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] \right| \right) \pi_\delta^2(E_t^\delta) dt \\ &\quad + 2\sigma_0 f' \left(\left| \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] \right| \right) \pi_\delta(E_t^\delta) e_t^\delta \cdot dB_t^0 + r_t^1 dt, \end{aligned} \quad (3.40)$$

where for all $t \geq 0$, $r_t^1 := \mathbf{I}_1(t) + \mathbf{II}_1(t) + \mathbf{III}_1(t)$, with

$$\begin{aligned} \mathbf{I}_1(t) &= f' \left(\left| E_t^\delta \right| \right) e_t^\delta \cdot \mathbb{E}_1 \left[\int_0^1 \{ DG(\rho X_t^\delta + (1-\rho)\tilde{X}_t^\delta) - DG(\rho \mathbb{E}_1[X_t^\delta] + (1-\rho)\mathbb{E}_1[\tilde{X}_t^\delta]) \} d\rho \right. \\ &\quad \left. (\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]) \right]; \\ \mathbf{II}_1(t) &= f' \left(\left| E_t^\delta \right| \right) e_t^\delta \cdot \mathbb{E}_1 \left[\int_0^1 \{ DG(\rho X_t^\delta + (1-\rho)\tilde{X}_t^\delta) - DG(\rho \mathbb{E}_1[X_t^\delta] + (1-\rho)\mathbb{E}_1[\tilde{X}_t^\delta]) \} d\rho \right. \\ &\quad \left. (A_t^\delta - \tilde{A}_t^\delta) \right]; \\ \mathbf{III}_1(t) &= f' \left(\left| E_t^\delta \right| \right) e_t^\delta \cdot \mathbb{E}_1 \left[\int_0^1 DF(\rho A_t^\delta + (1-\rho)\tilde{A}_t^\delta) d\rho (A_t^\delta - \tilde{A}_t^\delta) \right], \end{aligned} \quad (3.41)$$

where we used the following two identities: $\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] + A_t^\delta - \tilde{A}_t^\delta = X_t^\delta - \tilde{X}_t^\delta$ and $\mathbb{E}_1[A_t^\delta] - \mathbb{E}_1[\tilde{A}_t^\delta] = 0$.

Step 3. Recall (3.34):

$$f''(r) - \frac{1}{4} r \kappa(r) f'(r) \leq -\frac{c}{2\sigma_0^2} f(r) \quad \text{for all } r > 0.$$

Combining the above equation with the fact that, from Assumption 5,

$$(x - y) \cdot (G(x) - G(y)) \leq -\frac{2}{\sigma_0^2} \kappa(|x - y|)|x - y|^2, \quad x, y \in \mathbb{R}^d,$$

we obtain

$$\begin{aligned} & f' \left(\left| \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] \right| \right) e_t^\delta \cdot \left(G(\mathbb{E}_1[X_t^\delta]) - G(\mathbb{E}_1[\tilde{X}_t^\delta]) \right) \\ & \quad + 2\sigma_0^2 f'' \left(\left| \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] \right| \right) \pi_\delta^2(E_t^\delta) \\ & \leq -\frac{\sigma_0^2}{2} f' \left(\left| E_t^\delta \right| \right) |E_t^\delta| \kappa \left(\left| E_t^\delta \right| \right) + 2\sigma_0^2 f'' \left(\left| E_t^\delta \right| \right) \pi_\delta^2(E_t^\delta) \\ & \leq 2\sigma_0^2 \left(-\frac{1}{4} f' \left(\left| E_t^\delta \right| \right) |E_t^\delta| \kappa \left(\left| E_t^\delta \right| \right) + f'' \left(\left| E_t^\delta \right| \right) \right) \pi_\delta^2(E_t^\delta) - \frac{\sigma_0^2}{2} f' \left(\left| E_t^\delta \right| \right) |E_t^\delta| \kappa \left(\left| E_t^\delta \right| \right) \lambda_\delta^2(E_t^\delta) \\ & \leq -cf \left(\left| \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] \right| \right) + h(\delta), \end{aligned} \tag{3.42}$$

where

$$h(\delta) := \frac{\sigma_0^2}{2} f' \left(\left| E_t^\delta \right| \right) |E_t^\delta| \kappa \left(\left| E_t^\delta \right| \right) \lambda_\delta^2(E_t^\delta).$$

Now, recalling that $\lambda_\delta(x) = 0$ as soon as $|x| \geq \delta$, we obtain for $\delta \in [0, 1]$,

$$|h(\delta)| \leq \frac{\sigma_0^2}{2} \|f'\|_\infty \delta \kappa_2, \tag{3.43}$$

where $\kappa_2 = \sup_{0 \leq r \leq 1} \kappa(r)$, from which we deduce that $h(\delta) \rightarrow 0$ with δ . Then, back to (3.40), we can write

$$\begin{aligned} df \left(\left| \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] \right| \right) & \leq -cf \left(\left| \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] \right| \right) dt + |r_t^1| dt + h(\delta) dt \\ & \quad + 2\sigma_0 f' \left(\left| \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] \right| \right) \pi_\delta(E_t^\delta) e_t^\delta \cdot dB_t^0, \end{aligned}$$

and

$$\frac{d}{dt} \mathbb{E}_0 \left[f \left(\left| \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] \right| \right) \right] \leq -c \mathbb{E}_0 \left[f \left(\left| \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] \right| \right) \right] + \mathbb{E}_0[|r_t^1|] + \mathbb{E}_0[|h(\delta)|].$$

Using Grönwall's Lemma, we obtain

$$\begin{aligned} \mathbb{E}_0 \left[f \left(\left| \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] \right| \right) \right] & \leq \mathbb{E}_0 \left[f \left(\left| \mathbb{E}_1[X_0^\delta] - \mathbb{E}_1[\tilde{X}_0^\delta] \right| \right) \right] e^{-ct} + \int_0^t e^{-c(t-s)} \mathbb{E}_0[|r_s^1|] ds \\ & \quad + \frac{h(\delta)}{c}. \end{aligned} \tag{3.44}$$

Step 4. It remains to control the three parts in the expansion of the residual term r^1 , see (3.41). For

the two first ones, we use the Lipschitz continuity of DG with respect to the norm $\|\cdot\|$ and write

$$\begin{aligned} |\mathbf{I}_1(t)| &\leq C_G \mathbb{E}_1 \left[|X_t^\delta - \mathbb{E}_1[X_t^\delta]|^2 + |\tilde{X}_t^\delta - \mathbb{E}_1[\tilde{X}_t^\delta]|^2 \right]^{1/2} \left| \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] \right|; \\ |\mathbf{II}_1(t)| &\leq C_G \mathbb{E}_1 \left[|X_t^\delta - \mathbb{E}_1[X_t^\delta]|^2 + |\tilde{X}_t^\delta - \mathbb{E}_1[\tilde{X}_t^\delta]|^2 \right]^{1/2} \mathbb{E}_1 \left[|A_t^\delta - \tilde{A}_t^\delta|^2 \right]^{1/2}. \end{aligned}$$

Letting $\eta_t^\delta = \mathbb{E}_1 \left[|X_t^\delta - \mathbb{E}_1[X_t^\delta]|^2 + |\tilde{X}_t^\delta - \mathbb{E}_1[\tilde{X}_t^\delta]|^2 \right]^{1/2}$, for all $t \geq 0$, and recalling from Proposition 3.10 that

$$\eta_t^\delta \leq \eta_0^\delta e^{-c_{\alpha_F} t} + 2\sigma c_{\alpha_F}^{-1/2} \sqrt{d},$$

we obtain

$$\begin{aligned} &|\mathbf{I}_1(t)| + |\mathbf{II}_1(t)| \\ &\leq C_G \left(\eta_0^\delta e^{-c_{\alpha_F} t} + 2\sigma c_{\alpha_F}^{-1/2} \sqrt{d} \right) \left(\mathbb{E}_1 \left[|A_t^\delta - \tilde{A}_t^\delta|^2 \right]^{1/2} + \left| \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] \right| \right). \end{aligned} \quad (3.45)$$

For the term $\mathbf{III}_1(t)$, we use the fact that $\mathbb{E}_1[A_t^\delta] = \mathbb{E}_1[\tilde{A}_t^\delta] = 0$ and write

$$\begin{aligned} |\mathbf{III}_1(t)| &= \left| \mathbb{E}_1 \left[\int_0^1 \{DF(\rho A_t^\delta + (1-\rho)\tilde{A}_t^\delta) - DF(0)\} d\rho \cdot (A_t^\delta - \tilde{A}_t^\delta) \right] \right| \\ &\leq C_F \left(\eta_0^\delta e^{-c_{\alpha_F} t} + 2\sigma c_{\alpha_F}^{-1/2} \sqrt{d} \right) \mathbb{E}_1 \left[|A_t^\delta - \tilde{A}_t^\delta|^2 \right]^{1/2}, \end{aligned} \quad (3.46)$$

where we used once again Proposition 3.10 together with the Lipschitz continuity of F to get the last line. Then, combining (3.40), (3.45) and (3.46), we have

$$\begin{aligned} \mathbb{E}_0[|r_t^1|] &\leq \mathbb{E}_0 [|\mathbf{I}_1(t)| + |\mathbf{II}_1(t)| + |\mathbf{III}_1(t)|] \\ &\leq (C_G + C_F) \left(\eta_0^\delta e^{-c_{\alpha_F} t} + 2\sigma c_{\alpha_F}^{-1/2} \sqrt{d} \right) \mathbb{E}_0 \left[\mathbb{E}_1 \left[|A_t^\delta - \tilde{A}_t^\delta|^2 \right]^{1/2} + \left| \mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta] \right| \right]. \end{aligned}$$

Here, we recall that η_0^δ is deterministic because m and \tilde{m} are deterministic, see Remark 3.16. Finally, using inequalities (3.33) and (3.44), we obtain

$$\begin{aligned} \kappa_1 \mathbb{E}_0 [|\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]|] &\leq e^{-ct} \left| \mathbb{E}_1[X_0^\delta] - \mathbb{E}_1[\tilde{X}_0^\delta] \right| + \frac{h(\delta)}{c} \\ &+ C_b \int_0^t e^{-c(t-s)} \left(\eta_0^\delta e^{-c_{\alpha_F} s} + 2\sigma c_{\alpha_F}^{-1/2} \sqrt{d} \right) \mathbb{E}_0 \left[\mathbb{E}_1 \left[|A_s^\delta - \tilde{A}_s^\delta|^2 \right]^{1/2} + \left| \mathbb{E}_1[X_s^\delta] - \mathbb{E}_1[\tilde{X}_s^\delta] \right| \right] ds, \end{aligned} \quad (3.47)$$

where $C_b := C_G + C_F$.

Step 5. We now focus on the (normed) difference process $(|A_t^\delta - \tilde{A}_t^\delta|)_{t \geq 0}$. Thanks to the properties of F , as stated in Assumption 6, we are here able to show that it decreases exponentially fast to zero.

Recalling from Step 2 that $A_t^\delta = X_t^\delta - \mathbb{E}_1[X_t^\delta]$ for $t \geq 0$, we have

$$dA_t^\delta = \left(G(X_t^\delta) - \mathbb{E}_1[G(X_t^\delta)] \right) dt + \left(F(X_t^\delta - \mathbb{E}_1[X_t^\delta]) dt - \mathbb{E}_1[F(X_t^\delta - \mathbb{E}_1[X_t^\delta])] dt + \sigma dB_t, \right)$$

from which we obtain

$$\begin{aligned} \frac{d}{dt}(A_t^\delta - \tilde{A}_t^\delta) &= \left[G(X_t^\delta) - G(\tilde{X}_t^\delta) \right] + \left[F(A_t^\delta) - F(\tilde{A}_t^\delta) \right] \\ &\quad - \left[\mathbb{E}_1[G(X_t^\delta)] - \mathbb{E}_1[G(\tilde{X}_t^\delta)] \right] - \left[\mathbb{E}_1[F(A_t^\delta)] - \mathbb{E}_1[F(\tilde{A}_t^\delta)] \right]. \end{aligned}$$

Then,

$$\begin{aligned} \frac{d}{dt}|A_t^\delta - \tilde{A}_t^\delta|^2 &= 2(A_t^\delta - \tilde{A}_t^\delta) \cdot (G(X_t^\delta) - G(\tilde{X}_t^\delta)) \\ &\quad + 2(A_t^\delta - \tilde{A}_t^\delta) \cdot (F(A_t^\delta) - F(\tilde{A}_t^\delta)) \\ &\quad - 2(A_t^\delta - \tilde{A}_t^\delta) \cdot (\mathbb{E}_1[G(X_t^\delta)] - \mathbb{E}_1[G(\tilde{X}_t^\delta)]) \\ &\quad - 2(A_t^\delta - \tilde{A}_t^\delta) \cdot (\mathbb{E}_1[F(A_t^\delta)] - \mathbb{E}_1[F(\tilde{A}_t^\delta)]) \end{aligned} \quad (3.48)$$

We can immediately leverage on Assumption 6 to control the second term on the right hand side:

$$2(A_t^\delta - \tilde{A}_t^\delta) \cdot (F(A_t^\delta) - F(\tilde{A}_t^\delta)) \leq -2\alpha_F |A_t^\delta - \tilde{A}_t^\delta|^2. \quad (3.49)$$

Let us now focus on the terms involving G . We have

$$\begin{aligned} &(A_t^\delta - \tilde{A}_t^\delta) \cdot \left(G(X_t^\delta) - G(\tilde{X}_t^\delta) - \mathbb{E}_1[G(X_t^\delta)] + \mathbb{E}_1[G(\tilde{X}_t^\delta)] \right) \\ &= (A_t^\delta - \tilde{A}_t^\delta) \cdot \left[\left(\int_0^1 \left\{ DG(\rho X_t^\delta + (1-\rho)\tilde{X}_t^\delta) - DG(\rho \mathbb{E}_1[X_t^\delta] + (1-\rho)\mathbb{E}_1[\tilde{X}_t^\delta]) \right\} d\rho \right) (X_t^\delta - \tilde{X}_t^\delta) \right] \\ &\quad + (A_t^\delta - \tilde{A}_t^\delta) \cdot \left[\left(\int_0^1 DG(\rho \mathbb{E}_1[X_t^\delta] + (1-\rho)\mathbb{E}_1[\tilde{X}_t^\delta]) d\rho \right) (X_t^\delta - \tilde{X}_t^\delta) \right] \\ &\quad - (A_t^\delta - \tilde{A}_t^\delta) \cdot \mathbb{E}_1 \left[\left(\int_0^1 \left\{ DG(\rho X_t^\delta + (1-\rho)\tilde{X}_t^\delta) - DG(\rho \mathbb{E}_1[X_t^\delta] + (1-\rho)\mathbb{E}_1[\tilde{X}_t^\delta]) \right\} d\rho \right) (X_t^\delta - \tilde{X}_t^\delta) \right] \\ &\quad - (A_t^\delta - \tilde{A}_t^\delta) \cdot \left[\left(\int_0^1 DG(\rho \mathbb{E}_1[X_t^\delta] + (1-\rho)\mathbb{E}_1[\tilde{X}_t^\delta]) d\rho \right) (\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]) \right]. \end{aligned}$$

By combining the second and fourth term in the right hand side, we obtain

$$\begin{aligned} &(A_t^\delta - \tilde{A}_t^\delta) \cdot \left(G(X_t^\delta) - G(\tilde{X}_t^\delta) - \mathbb{E}_1[G(X_t^\delta)] + \mathbb{E}_1[G(\tilde{X}_t^\delta)] \right) \\ &= (A_t^\delta - \tilde{A}_t^\delta) \cdot \left[\left(\int_0^1 \left\{ DG(\rho X_t^\delta + (1-\rho)\tilde{X}_t^\delta) - DG(\rho \mathbb{E}_1[X_t^\delta] + (1-\rho)\mathbb{E}_1[\tilde{X}_t^\delta]) \right\} d\rho \right) (X_t^\delta - \tilde{X}_t^\delta) \right] \\ &\quad + (A_t^\delta - \tilde{A}_t^\delta) \cdot \left(\int_0^1 DG(\rho \mathbb{E}_1[X_t^\delta] + (1-\rho)\mathbb{E}_1[\tilde{X}_t^\delta]) d\rho \right) (A_t^\delta - \tilde{A}_t^\delta) \\ &\quad - (A_t^\delta - \tilde{A}_t^\delta) \cdot \mathbb{E}_1 \left[\left(\int_0^1 \left\{ DG(\rho X_t^\delta + (1-\rho)\tilde{X}_t^\delta) - DG(\rho \mathbb{E}_1[X_t^\delta] + (1-\rho)\mathbb{E}_1[\tilde{X}_t^\delta]) \right\} d\rho \right) (X_t^\delta - \tilde{X}_t^\delta) \right] \\ &=: \bar{\mathbf{A}}_1(t) + \bar{\mathbf{A}}_2(t) + \bar{\mathbf{A}}_3(t). \end{aligned}$$

Then, if we perform the following decomposition of the first term on the right-hand side, we obtain

$$\bar{\mathbf{A}}_1(t) = \bar{\mathbf{A}}\bar{\mathbf{A}}_1(t) + \bar{\mathbf{A}}\bar{\mathbf{A}}_2(t), \quad (3.50)$$

with

$$\begin{aligned}\overline{\Delta\Delta}_1(t) &:= (A_t^\delta - \tilde{A}_t^\delta) \cdot \left[\left(\int_0^1 \left\{ DG(\rho X_t^\delta + (1-\rho)\tilde{X}_t^\delta) - DG(\rho\mathbb{E}_1[X_t^\delta] + (1-\rho)\mathbb{E}_1[\tilde{X}_t^\delta]) \right\} d\rho \right) \right. \\ &\quad \left. (\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]) \right] \\ \overline{\Delta\Delta}_2(t) &:= (A_t^\delta - \tilde{A}_t^\delta) \cdot \left(\int_0^1 \left\{ DG(\rho X_t^\delta + (1-\rho)\tilde{X}_t^\delta) \right. \right. \\ &\quad \left. \left. - DG(\rho\mathbb{E}_1[X_t^\delta] + (1-\rho)\mathbb{E}_1[\tilde{X}_t^\delta]) \right\} d\rho \right) (A_t^\delta - \tilde{A}_t^\delta).\end{aligned}$$

Then, combining the terms $\overline{\Delta}_2(t)$ and $\overline{\Delta\Delta}_2(t)$, we get

$$\begin{aligned}& (A_t^\delta - \tilde{A}_t^\delta) \cdot \left(G(X_t^\delta) - G(\tilde{X}_t^\delta) - \mathbb{E}_1[G(X_t^\delta)] + \mathbb{E}_1[G(\tilde{X}_t^\delta)] \right) \\ &= \overline{\Delta\Delta}_1(t) + [\overline{\Delta}_2(t) + \overline{\Delta\Delta}_2(t)] + \overline{\Delta}_3(t) \\ &=: \overline{\Delta\Delta}_1(t) + \overline{\Pi}_2(t) + \overline{\Delta}_3(t),\end{aligned}\tag{3.51}$$

where $\overline{\Delta\Delta}_1(t)$ and $\overline{\Delta}_3(t)$ have been already defined and $\overline{\Pi}_2(t)$ is defined by

$$\overline{\Pi}_2(t) := (A_t^\delta - \tilde{A}_t^\delta) \cdot \left(\int_0^1 DG(\rho X_t^\delta + (1-\rho)\tilde{X}_t^\delta) d\rho \right) (A_t^\delta - \tilde{A}_t^\delta).$$

We now control $\overline{\Pi}_2(t)$ as we handled Equation (3.49). We get

$$\overline{\Pi}_2(t) \leq m_G |A_t^\delta - \tilde{A}_t^\delta|^2,\tag{3.52}$$

where we recall from (3.6) that $m_G \in \mathbb{R}$ stands for the upper bound of the function $r \mapsto [-\sigma_0^2 \kappa(r)/2]$ introduced in Assumption 5. Indeed, as a consequence of 5, one can show that for any $x, y \in \mathbb{R}^d$,

$$y \cdot (DG(x)y) \leq m_G |y|^2.$$

Then, inserting (3.49), (3.51) and (3.52) into (3.48), we obtain

$$\begin{aligned}& \frac{d}{dt} |A_t^\delta - \tilde{A}_t^\delta|^2 \\ & \leq -2(\alpha_F - m_G) |A_t^\delta - \tilde{A}_t^\delta|^2 + \overline{\Delta\Delta}_1(t) + \overline{\Delta}_3(t) + 2(A_t^\delta - \tilde{A}_t^\delta) \cdot (\mathbb{E}_1[F(A_t^\delta)] - \mathbb{E}_1[F(\tilde{A}_t^\delta)]).\end{aligned}$$

By taking expectation under \mathbb{P}_1 and by using the fact that $\mathbb{E}_1[A_t^\delta] = \mathbb{E}_1[\tilde{A}_t^\delta] = 0$, we get rid the last two terms in the above right-hand side. We obtain

$$\frac{d}{dt} \mathbb{E}_1 \left[|A_t^\delta - \tilde{A}_t^\delta|^2 \right] \leq -2(\alpha_F - m_G) \mathbb{E}_1 \left[|A_t^\delta - \tilde{A}_t^\delta|^2 \right] + 2\mathbb{E}_1 [\overline{\Delta\Delta}_1(t)].\tag{3.53}$$

In order to control $\overline{\Delta\Delta}_1(t)$, we come back to (3.50) and use the fact that $\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]$ is measurable with respect to the common noise. By Cauchy-Schwarz' inequality, we obtain

$$\mathbb{E}_1 [\overline{\Delta\Delta}_1(t)] \leq C_G \mathbb{E}_1 [|A_t^\delta - \tilde{A}_t^\delta|^2]^{1/2} \eta_t^\delta |\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]|,$$

where we recall that $\eta_t^\delta = \mathbb{E}_1[|A_t^\delta|^2 + |\tilde{A}_t^\delta|^2]^{1/2}$. Then, from Proposition 3.10, we get

$$\mathbb{E}_1[\overline{\Delta\Delta}_1(t)] \leq \mathbb{E}_1[|A_t^\delta - \tilde{A}_t^\delta|^2]^{1/2} \left(\eta_0^\delta e^{-c\alpha_F t} + \sigma c_{\alpha_F}^{-1/2} \sqrt{d} \right) |\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]|,$$

and (3.53) rewrites

$$\begin{aligned} \frac{d}{dt} \mathbb{E}_1 \left[|A_t^\delta - \tilde{A}_t^\delta|^2 \right] &\leq -2(\alpha_F - m_G) \mathbb{E}_1 \left[|A_t^\delta - \tilde{A}_t^\delta|^2 \right] \\ &\quad + 2\mathbb{E}_1[|A_t^\delta - \tilde{A}_t^\delta|^2]^{1/2} \left(\eta_0^\delta e^{-c\alpha_F t} + \sigma c_{\alpha_F}^{-1/2} \sqrt{d} \right) |\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]|. \end{aligned} \quad (3.54)$$

Then, we get

$$\begin{aligned} \frac{d}{dt} \mathbb{E}_1 \left[|A_t^\delta - \tilde{A}_t^\delta|^2 \right]^{1/2} &\leq -(\alpha_F - m_G) \mathbb{E}_1 \left[|A_t^\delta - \tilde{A}_t^\delta|^2 \right]^{1/2} \\ &\quad + C_G (\eta_0^\delta e^{-c\alpha_F t} + \sigma c_{\alpha_F}^{-1/2} \sqrt{d}) |\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]|, \end{aligned} \quad (3.55)$$

see Appendix 5.2 for details. Using Grönwall's Lemma, we obtain

$$\begin{aligned} \mathbb{E}_1[|A_t^\delta - \tilde{A}_t^\delta|^2]^{1/2} &\leq \mathbb{E}_1[|A_0^\delta - \tilde{A}_0^\delta|^2]^{1/2} \exp(-(\alpha_F - m_G)t) \\ &\quad + C_G \int_0^t e^{-(\alpha_F - m_G)(t-s)} (\eta_0^\delta e^{-c\alpha_F s} + \sigma c_{\alpha_F}^{-1/2} \sqrt{d}) |\mathbb{E}_1[X_s^\delta] - \mathbb{E}_1[\tilde{X}_s^\delta]| ds. \end{aligned}$$

Finally, taking again expectation \mathbb{E}_0 ,

$$\begin{aligned} \mathbb{E}_0 \left[\mathbb{E}_1[|A_t^\delta - \tilde{A}_t^\delta|^2]^{1/2} \right] &\leq \exp(-(\alpha_F - m_G)t) \mathbb{E}_1[|A_0^\delta - \tilde{A}_0^\delta|^2]^{1/2} \\ &\quad + C_G \int_0^t e^{-(\alpha_F - m_G)(t-s)} (\eta_0^\delta e^{-c\alpha_F s} + \sigma c_{\alpha_F}^{-1/2} \sqrt{d}) \mathbb{E}_0 \left[|\mathbb{E}_1[X_s^\delta] - \mathbb{E}_1[\tilde{X}_s^\delta]| \right] ds. \end{aligned} \quad (3.56)$$

Step 6. Now, as (3.47) remains true replacing the rate c by $\tilde{c} \leq c$, we can assume without loss of generality that $\alpha_F > m_G + c$. We can then combine Equations (3.47) and (3.56) to obtain, for all $t \geq 0$,

$$\begin{aligned} &\mathbb{E}_0 \left[\mathbb{E}_1 \left[|A_t^\delta - \tilde{A}_t^\delta|^2 \right]^{1/2} + |\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]| \right] \\ &\leq \left(\mathbb{E}_1 \left[|A_0^\delta - \tilde{A}_0^\delta|^2 \right]^{1/2} + |\mathbb{E}_1[X_0^\delta] - \mathbb{E}_1[\tilde{X}_0^\delta]| \right) e^{-ct} + \frac{h(\delta)}{c} \\ &\quad + c_3^b \int_0^t e^{-c(t-s)} \left(\eta_0^\delta e^{-c\alpha_F s} + 2\sigma c_{\alpha_F}^{-1/2} \sqrt{d} \right) \mathbb{E}_0 \left[\mathbb{E}_1 \left[|A_s^\delta - \tilde{A}_s^\delta|^2 \right]^{1/2} + |\mathbb{E}_1[X_s^\delta] - \mathbb{E}_1[\tilde{X}_s^\delta]| \right] ds, \end{aligned} \quad (3.57)$$

for some constant $c_3^b > 0$ that depends on F and G .

With the notation $\Theta_t^\delta := \mathbb{E}_1 \left[|A_t^\delta - \tilde{A}_t^\delta|^2 \right]^{1/2} + |\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]|$, Equation (3.57) reads

$$\mathbb{E}_0[\Theta_t^\delta] \leq \Theta_0^\delta e^{-ct} + \frac{h(\delta)}{c} + c_3^b \int_0^t e^{-c(t-s)} \left(\eta_0^\delta e^{-c\alpha_F s} + 2\sigma c_{\alpha_F}^{-1/2} \sqrt{d} \right) \mathbb{E}_0 \left[\Theta_s^\delta \right] ds.$$

Then, using Grönwall Lemma, we obtain

$$\begin{aligned} e^{ct}\mathbb{E}_0\left[\Theta_t^\delta\right] &\leq\left(\Theta_0+\frac{h(\delta)}{c}\right)\exp\left(\frac{c_3^b}{c_{\alpha_F}}\eta_0+\frac{c_3^b\sigma\sqrt{d}}{\sqrt{c_{\alpha_F}}}t\right) \\ &+h(\delta)\int_0^te^{cs}\exp\left(\frac{c_3^b}{c_{\alpha_F}}\eta_0+\frac{c_3^b\sigma\sqrt{d}}{\sqrt{c_{\alpha_F}}}(t-s)\right)ds. \end{aligned} \quad (3.58)$$

We emphasize here that we write Θ_0 and η_0 without reference to δ because these quantities are predetermined and independent of δ . Now, let us notice that creating the couplage we defined two Brownian motions adapted to the filtration \mathbb{F}_0 :

$$\begin{aligned} \beta_t^\delta &= \int_0^t \pi_\delta(E_s^\delta)(\text{Id} - 2e_s^\delta(e_s^\delta)^\top)dB_s^0 + \int_0^t \lambda_\delta(E_s^\delta)d\tilde{B}_s^0 \\ \tilde{\beta}_t^\delta &= \int_0^t \pi_\delta(E_s^\delta)dB_s^0 + \int_0^t \lambda_\delta(E_s^\delta)d\tilde{B}_s^0. \end{aligned}$$

Setting $m_t^\delta = \mathcal{L}_1(X_t^\delta)$ et $\tilde{m}_t^\delta = \mathcal{L}_1(\tilde{X}_t^\delta)$, we get

$$\begin{aligned} d_t m_t^\delta &= \nabla \cdot \left(\frac{\sigma^2 + \sigma_0^2}{2} \nabla m_t^\delta - m_t^\delta \left(G + F(\cdot - \mu_1(m_t^\delta)) \right) \right) dt - \sigma_0 \nabla m_t^\delta \cdot d\beta_t^\delta, \\ d_t \tilde{m}_t^\delta &= \nabla \cdot \left(\frac{\sigma^2 + \sigma_0^2}{2} \nabla \tilde{m}_t^\delta - \tilde{m}_t^\delta \left(G + F(\cdot - \mu_1(\tilde{m}_t^\delta)) \right) \right) dt - \sigma_0 \nabla \tilde{m}_t^\delta \cdot d\tilde{\beta}_t^\delta. \end{aligned}$$

Then, one has $\mathcal{P}_t\phi(m) = \mathbb{E}_0[\phi(m_t^\delta)|m_0^\delta = m]$ and $\mathcal{P}_t\phi(\tilde{m}) = \mathbb{E}_0[\phi(\tilde{m}_t^\delta)|\tilde{m}_0^\delta = \tilde{m}]$. Then, we obtain that for any $\delta > 0$,

$$\begin{aligned} |\mathcal{P}_t\phi(m) - \mathcal{P}_t\phi(\tilde{m})| &\leq \|\phi\|_{\text{lip}}\mathbb{E}_0\left[\mathbb{E}_1\left[\left|X_t^\delta - \tilde{X}_t^\delta\right|\right]\right] \\ &\leq \|\phi\|_{\text{lip}}\mathbb{E}_0[\Theta_t^\delta] \end{aligned} \quad (3.59)$$

Plugging the estimate given by Equation (3.58) into Equation (3.59) gives

$$\begin{aligned} |\mathcal{P}_t\phi(m) - \mathcal{P}_t\phi(\tilde{m})| &\leq \|\phi\|_{\text{lip}}\left(\Theta_0+\frac{h(\delta)}{c}\right)\exp\left(\frac{c_3^b}{c_{\alpha_F}}\eta_0+\left(\frac{c_3^b\sigma\sqrt{d}}{\sqrt{c_{\alpha_F}}}-c\right)t\right) \\ &+ \|\phi\|_{\text{lip}}h(\delta)\int_0^te^{-c(t-s)}\exp\left(\frac{c_3^b}{c_{\alpha_F}}\eta_0+\frac{c_3^b\sigma\sqrt{d}}{\sqrt{c_{\alpha_F}}}(t-s)\right)ds. \end{aligned}$$

The latter being true for all $\delta > 0$, we can consider the limit $\delta \rightarrow 0$ to finally obtain

$$|\mathcal{P}_t\phi(m) - \mathcal{P}_t\phi(\tilde{m})| \leq \|\phi\|_{\text{lip}}\Theta_0\exp\left(\frac{c_3^b}{c_{\alpha_F}}\eta_0+\frac{c_3^b\sigma\sqrt{d}}{\sqrt{c_{\alpha_F}}}t-ct\right).$$

Then, for σ and α satisfying $\mathfrak{c} := \sigma c_3^b c_{\alpha_F}^{-1/2} - c < 0$, we get the expected exponential rate of convergence.

Moreover, writing

$$\begin{aligned} \Theta_0 \exp\left(\frac{c_3^b}{c_{\alpha_F}} \eta_0\right) &= \left(\mathbb{E}_1\left[|A_0 - \tilde{A}_0|^2\right]^{1/2} + |\mathbb{E}_1[X_0] - \mathbb{E}_1[\tilde{X}_0]|\right) \exp\left(\frac{c_3^b}{c_{\alpha_F}} \eta_0\right) \\ &\leq C_2 \left(\mu_2(m)^{1/2} + \mu_2(\tilde{m})^{1/2}\right) \exp\left(\frac{c_3^b}{c_{\alpha_F}} (v(m)^{1/2} + v(\tilde{m})^{1/2})\right), \end{aligned}$$

for a new constant positive C_2 , we complete the proof of Proposition 3.7. \square

5 Appendix

5.1 Proof of Equation (3.38)

Proof. Using Itô's formula, and recalling that $E^\delta = X^\delta - \tilde{X}^\delta$, we get that for any $\delta > 0$,

$$\begin{aligned} &d|\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]|^2 \\ &= 2(\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]) \cdot \left(\mathbb{E}_1[G(X_t^\delta)] - \mathbb{E}_1[G(\tilde{X}_t^\delta)]\right) dt \\ &+ 2(\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]) \cdot \left(\mathbb{E}_1[F(X_t^\delta - \mathbb{E}_1[X_t^\delta])] - \mathbb{E}_1[G(F(\tilde{X}_t^\delta - \mathbb{E}_1[\tilde{X}_t^\delta]))]\right) dt \\ &+ 4\sigma_0\pi_\delta(E_t^\delta)((\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]) \cdot e_t^\delta) e_t^\delta \cdot dB_t^0 \\ &+ 4\sigma_0^2\pi_\delta^2(E_t^\delta)dt. \end{aligned}$$

For any $\varepsilon > 0$, let us introduce $\psi_\varepsilon : [0, +\infty) \in r \mapsto (r + \varepsilon)^{1/2}$. Since this function is twice continuously differentiable, we can write

$$\begin{aligned} d\psi_\varepsilon\left(|\mathbb{E}_1[X_t^\delta] - \mathbb{E}_1[\tilde{X}_t^\delta]|^2\right) &= 2\psi'_\varepsilon(|E_t^\delta|^2)E_t^\delta \cdot \left(\mathbb{E}_1[G(X_t^\delta)] - \mathbb{E}_1[G(\tilde{X}_t^\delta)]\right) dt \\ &+ 2\psi'_\varepsilon(|E_t^\delta|^2)E_t^\delta \cdot \left(\mathbb{E}_1[F(X_t^\delta - \mathbb{E}_1[X_t^\delta])] - \mathbb{E}_1[F(\tilde{X}_t^\delta - \mathbb{E}_1[\tilde{X}_t^\delta])]\right) dt \\ &+ 4\sigma_0\psi'_\varepsilon(|E_t^\delta|^2)\pi_\delta(E_t^\delta)(E_t^\delta \cdot e_t^\delta)e_t^\delta \cdot dB_t^0 \tag{3.60} \\ &+ 4\sigma_0^2\psi'_\varepsilon(|E_t^\delta|^2)\pi_\delta^2(E_t^\delta)dt \\ &+ 8\sigma_0^2\psi''_\varepsilon(|E_t^\delta|^2)|E_t^\delta|^2\pi_\delta(E_t^\delta)dt. \end{aligned}$$

We now want to take the limit $\varepsilon \rightarrow 0$. Using dominated convergence theorem and stochastic dominated convergence theorem as stated in [RY99, Theorem 2.12], combined with the bound $4r\psi'_\varepsilon(r^2) \leq 1$, we can deal with the first two lines and get that for all $t \geq 0$,

$$\lim_{\varepsilon \rightarrow 0} \int_0^t 2\psi'_\varepsilon(|E_s^\delta|^2)E_s^\delta \cdot \left(\mathbb{E}_1[G(X_s^\delta)] - \mathbb{E}_1[G(\tilde{X}_s^\delta)]\right) ds = \int_0^t e_s^\delta \cdot \left(\mathbb{E}_1[G(X_s^\delta)] - \mathbb{E}_1[G(\tilde{X}_s^\delta)]\right) ds, \tag{3.61}$$

$$\begin{aligned} &\lim_{\varepsilon \rightarrow 0} \int_0^t 2\psi'_\varepsilon(|E_s^\delta|^2)E_s^\delta \cdot \left(\mathbb{E}_1[F(X_s^\delta - \mathbb{E}_1[X_s^\delta])] - \mathbb{E}_1[F(\tilde{X}_s^\delta - \mathbb{E}_1[\tilde{X}_s^\delta])]\right) ds \\ &= \int_0^t e_s^\delta \cdot \left(\mathbb{E}_1[F(X_s^\delta - \mathbb{E}_1[X_s^\delta])] - \mathbb{E}_1[F(\tilde{X}_s^\delta - \mathbb{E}_1[\tilde{X}_s^\delta])]\right) ds, \end{aligned} \tag{3.62}$$

and

$$\lim_{\varepsilon \rightarrow 0} \int_0^t 4\sigma_0 \psi'_\varepsilon(|E_s^\delta|^2) \pi_\delta(E_s^\delta) (E_s^\delta \cdot e_s^\delta) e_s^\delta \cdot dB_s^0 = \int_0^t 2\sigma_0 \pi_\delta(E_s^\delta) e_s^\delta \cdot dB_s^0, \quad (3.63)$$

almost surely. For the last two terms in Equation (3.60), let us remark that for all $r \geq 0$,

$$\psi'_\varepsilon(r) + 2\psi''(r) = \frac{1}{2(r+\varepsilon)^{1/2}} - \frac{r}{2(r+\varepsilon)^{3/2}} \rightarrow 0, \quad (3.64)$$

when ε tends to 0. In order to we need to properly take the limit $\varepsilon \rightarrow 0$ in Equation (3.60), we need to take advantage of the presence of the function π_δ for $\delta > 0$. In Indeed, we have

$$\begin{aligned} & \left| 4\sigma_0^2 \psi'_\varepsilon(|E_t^\delta|^2) \pi_\delta^2(E_t^\delta) + 8\sigma_0^2 \psi''_\varepsilon(|E_t^\delta|^2) |E_t^\delta|^2 \pi_\delta(E_t^\delta) \right| \\ & \leq \left| \pi_\delta^2(E_t^\delta) \sigma_0^2 \left(4\psi'_\varepsilon(|E_t^\delta|^2) + 8\psi''_\varepsilon(|E_t^\delta|^2) |E_t^\delta|^2 \right) \right|. \end{aligned}$$

Moreover, we know that $\psi'_\varepsilon(r^2) + 2\psi''_\varepsilon(r^2)r^2 \leq r^{-3}$, for all $r > 0$ and $\varepsilon \leq 1$. Using the presence of π_δ , we have that the integrand is null near 0. Then, we can once again apply dominated convergence theorem and obtain

$$\lim_{\varepsilon \rightarrow 0} \int_0^t 4 \left\{ \sigma_0^2 \psi'_\varepsilon(|E_s^\delta|^2) \pi_\delta(E_s^\delta)^2 + 8\sigma_0^2 \psi''_\varepsilon(|E_s^\delta|^2) |E_s^\delta|^2 \pi_\delta(E_s^\delta) \right\} ds = 0. \quad (3.65)$$

Finally, combining Equations (3.61), (3.62) (3.63) and (3.65), we get the expected result. \square

5.2 Proof of Equation (3.55)

Proof. To prove Equation, let us introduce the following technical Lemma:

Lemma 3.17. *For any positive and differentiable function $u : [0, +\infty) \rightarrow [0, +\infty)$ such that*

$$\frac{d}{dt} u(t) \leq 2k(t) \sqrt{u(t)},$$

for some function $k : [0, +\infty) \rightarrow \mathbb{R}$, we have

$$\frac{d}{dt} \sqrt{u(t)} \leq k(t).$$

Proof of Lemma 3.17. We consider for any $\varepsilon > 0$ the function $\psi_\varepsilon : [0, +\infty) \ni r \mapsto (r + \varepsilon)^{1/2}$. The function ψ_ε being differentiable with positive derivative, we get

$$\frac{d}{dt} \psi_\varepsilon(u(t)) \leq 2k(t) \sqrt{u(t)} \psi'_\varepsilon(u(t)).$$

Using the fact that $2r\psi'_\varepsilon(r^2) = r(r^2 + \varepsilon)^{-1/2} \leq 1$, we use dominated convergence theorem to conclude the proof. \square

Such a lemma is classical and applying it to Equation (3.54) concludes the proof of Equation (3.55). \square

Part II

Statistical analysis of diffusion models in presence of noise

Chapter 4

Invariant density estimation from noisy data

Abstract

We introduce a new approach for estimating the invariant density of a multidimensional diffusion when dealing with high-frequency observations blurred by independent noises. We consider the intermediate regime, where observations occur at discrete time instances $k\Delta_n$ for $k = 0, \dots, n$, under the conditions $\Delta_n \rightarrow 0$ and $n\Delta_n \rightarrow \infty$. Our methodology involves the construction of a kernel density estimator that uses a pre-averaging technique to proficiently remove noise from the data while preserving the analytical characteristics of the underlying signal and its asymptotic properties. The rate of convergence of our estimator depends on both the anisotropic regularity of the density and the intensity of the noise. We establish conditions on the intensity of the noise that ensure the recovery of convergence rates similar to those achievable without any noise. Furthermore, we prove a Bernstein concentration inequality for our estimator, from which we derive an adaptive procedure for the kernel bandwidth selection.

Contents

1	Introduction	119
1.1	Setting	119
1.2	Motivation	119
1.3	Organization of the paper	121
2	Statistical and Probabilistic framework	122
2.1	Notation	122
2.2	The statistical diffusion model	122
3	Estimation procedure	124
4	Upper bounds and hyper-parameter choice	128
4.1	General results	129
4.2	Analysis of the convergence rate	131
4.3	Comments	133
5	Bernstein inequality and adaptive choice of the hyperparameters	133
5.1	Bernstein inequality	133
5.2	Choice of the hyperparameters	134
6	Numerical Analysis	135
7	Probabilistic tools: the Markovian structure of the preaveraged process	137

8	Invertibility of A	139
9	Proof of Proposition 4.3	139
10	Proof of Proposition 4.4	151
11	Proof of Proposition 4.5	152
11.1	Structure and completion of the proof	152
11.2	Proof of Lemma 4.26	158
11.3	Proof of Lemma 4.27	159
11.4	Proof of Lemma 4.28	160
11.5	Proof of Lemma 4.29	161
12	Proof of the results of Section 4.1	163
12.1	Preliminary results	163
12.2	Proof of Proposition 4.7	164
12.3	Proof of Proposition 4.8	165
12.4	Proof of Proposition 4.10	166
12.5	Proof of Proposition 4.11	167
13	Proof of the results of Section 5	168
13.1	Proof of Theorem 4.12	168
13.2	Proof of Proposition 4.14	170
14	Proof of the results of Section 7	175
14.1	Proof of Lemma 4.17	176
14.2	Proof of Lemma 4.18	179
14.3	Proof of Corollary 4.19	179
14.4	Proof of Lemma 4.20	179

1 Introduction

1.1 Setting

In this paper, we revisit the classical problem of estimating the distribution of a signal blurred by additive noise. We focus on a d -dimensional deconvolution model

$$Y_{i,n} = X_{i,n} + \tau_n \xi_{i,n}, \quad i = 0, \dots, n, \quad (4.1)$$

where the variables $(X_{i,n})_i$ are stationary with common distribution μ and $(\xi_{i,n})_{i,n}$ is an i.i.d sequence independent of X . The sequence (τ_n) represents the noise intensity and is assumed known. When the $(X_{i,n})_i$ are independent, the noise is Gaussian and the target density is α -Hölder regular, the best achievable pointwise quadratic rate of estimation is $\log(n)^{-\alpha/2} \tau_n^\alpha$, see [Fan91, CL13] for more details. When τ_n is of order 1, the resulting convergence rate becomes logarithmic, and the estimators cannot be used in practice. In this work, we show the situation improves when the process $(X_{i,n})_i$ exhibits Markovian properties. Such structure can significantly enhance the efficiency of denoising the observations.

Throughout this paper, we consider a d -dimensional stochastic process X , defined on an appropriate probability space $(\Omega, \mathcal{A}, \mathbb{P})$, governed by the following dynamics

$$dX_t = b(X_t) dt + \sigma(X_t) dW_t. \quad (4.2)$$

Here, W represents a d -dimensional Brownian motion, and the functions $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ are the transport and diffusion coefficients of X , respectively. We specifically consider the case where σ is a known constant diagonal matrix. Under mild conditions on b and σ , the process X is ergodic and admits a unique stationary distribution, denoted by $\bar{\mu}^b$. Furthermore, $\bar{\mu}^b$ is absolutely continuous with respect to the Lebesgue measure on \mathbb{R}^d and has a density also denoted by $\bar{\mu}^b$.

We aim at estimating $\bar{\mu}^b$ from the $(Y_i)_{i=1, \dots, n}$ defined in Equation (4.1), with

$$X_{i,n} = X_{i\Delta_n}, \quad i = 0, \dots, n, \quad (4.3)$$

for some positive sequence $(\Delta_n)_n$. If Δ_n is of order 1 or larger, the variables $(X_{i,n})$ are exponentially β -mixing and we would retrieve results similar to the i.i.d case. When $\Delta_n \rightarrow 0$ and $n\Delta_n \rightarrow \infty$, we can use the Markovian and ergodic structure of (4.2) to build an estimator of $\bar{\mu}^b$ with polynomial rate even when $\tau_n = 1$.

1.2 Motivation

Historically, diffusion models were first introduced as approximations of discrete Markov chains. Over time, their relevance has significantly expanded across various domains of applied mathematics [Pap95, Ber93, Hul03, Bai57, Ric77]. Statistical inference for diffusion processes has attracted extensive study due to the model's significance in many applied fields. This research first included both parametric and

non-parametric estimation of the parameters b and σ . The estimation of the invariant measure $\bar{\mu}^b$ is also studied [NZ79, Del80, Bos98], pushed by its association with various numerical methods, such as Markov Chain Monte Carlo [LP02, Pan08]. Note also that the non-parametric estimation of $\bar{\mu}^b$ and the estimation of the transport b are intertwined [Sch13].

Initiated by [Ros56, Par62], non-parametric density estimation has been extensively studied in the context of i.i.d observations [Tsy09]. A natural estimator of the common density $\bar{\mu}$ of i.i.d. observations $(X_{i,n})_i$ is given by

$$\hat{\mu}_n(x) = \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{K}_{\mathbf{h}}(x - X_k)$$

where we write $\mathbf{K}_{\mathbf{h}}(y) = \prod_i h_i^{-1} K(y_i h_i^{-1})$ for any $\mathbf{h} = (h_1, \dots, h_d) \in (0, +\infty)^d$ and $y = (y_1, \dots, y_d) \in \mathbb{R}^d$, and where $K : \mathbb{R} \rightarrow \mathbb{R}$ is a bounded kernel. Oracle inequalities show that this estimator can achieve the convergence rate $n^{-\alpha/(2\alpha+d)}$ where α is the Hölder regularity of $\bar{\mu}$. This rate is optimal, see e.g. [Tsy09] for details. The tuning parameter \mathbf{h} needs to be chosen in an adaptive way to achieve this rate, see [GL08, GL09, GL11]. The extension to ergodic processes is not trivial and earlier statistical studies focus on two different asymptotic regimes: continuous observations of the process $(X_t)_{t \leq T}$ with $T \rightarrow \infty$, or low frequency observations given by $(X_{i\Delta})_i$ with $0 \leq i \leq n$. Since X is exponentially β -mixing under mild assumptions on b , low frequency observations naturally relate to the i.i.d. case and present similar convergence rate. On the other hand, continuous observations are can be studied using fine probabilistic tools for ergodic continuous-time Markovian dynamics [BCG08, CG08, Lez01, Pau15] and precise estimates on the transition densities [CW97, QRZ03, QZ04]. The seminal works of [DR06, DR07] first established convergence rates of kernel estimators under continuous observations of X over a time interval $[0, T]$. In these works, the invariant density is still estimated through a kernel based estimator as in the i.i.d. case. [DR06, DR07] also obtain the convergence rate

$$T^{-\alpha/(2\alpha+d-2)}$$

when $\bar{\mu}^b$ is α -Hölder and $d \geq 3$. When considering an anisotropic framework, the convergence rate depends on the effective average smoothness [Str18] and minimax rates are derived in [AG21].

Two significant limitations still need to be addressed in this setup: First, what happens if we access discrete high-frequency observations of X , i.e. we observe $(X_{i\Delta_n})_i$ for $0 \leq i \leq n$ when $\Delta_n \rightarrow 0$ and $n\Delta_n \rightarrow \infty$. Secondly, can these methods be applied if these observations are polluted with a noise.

The question of discrete observations naturally arise with the advent of high-frequency data collection, and in particular in finance [ASJ14]. Hence, understanding the estimation rates under different asymptotic conditions becomes crucial. This includes specifying conditions on Δ_n that determine when continuous or low-frequency observations are more analytically pertinent. While this topic has only recently received substantial attention, pioneering works like [GHR04] and [CT16], which explore random sampling times, stand out. However, this question has recently been addressed in full generalities in [AG23] where the breakeven point between the high frequency observations similar to continuous

observations and low frequency observations similar to the i.i.d. case is identified and studied. The continuous rate is known to be optimal [AG21] in all dimensions, but the question of the optimality of the low frequency rate is still an open question, with the exception of the one dimensional case, solved in [AG22].

The second limitation in previous studies is to effectively incorporate noise into these analyses. Noise is an often unavoidable element in practical applications, such as financial modeling, biological experiments, and sensor data analysis due to measurement errors or external noise. Underestimating noise can lead to biased and unreliable estimations. This question was extensively studied in the context of noisy observations of i.i.d random variables. When noise is assumed to have an additive structure, existing literature uses Fourier inversion and kernel-based methods to recover the distribution of interest [Dev89, LT89, SC90]. Later works [CH88, Fan91, Fan93] establish minimax optimality of this procedure under the assumptions that the noise distribution is known and has a non-vanishing Fourier transform. It is important to note that in this setup, the convergence rates are slow. For instance, when the distribution of interest is α -Hölder regular and the noises are independent standard Gaussian variables, the optimal rate of convergence for any estimator is only $\log(n)^{-\alpha/2}$. The rates of convergences in such framework has been studied under different set of assumptions depending on both regularities of the noise and the density of interest in [CL13]. However, literature on noisy ergodic setups remains sparse, with a few notable contributions including [Sch11, Sch12]. Unlike the i.i.d. case, the structure of (4.2) and (4.3) allows better extraction of the information hidden by the noise, leading to improved estimation rates. Indeed, using the Hölder regularity of X , we can denoise high-frequency data while preserving the analytical properties of the signal. We achieve this through a pre-averaging technique, as in [JLM⁺09]. This approach has been widely studied and could be generalized to other noises, as demonstrated in [JPV10, HP13, JM15] for high-frequency statistics and [Sch11, Sch12] within an ergodic framework. In this paper, we show that when the noise is relatively small (see Section 4.1 for details) we can estimate the invariant density rates to the non-noisy case studied in [AG23]. When the noise τ_n is relatively large, we can still estimate the invariant density with convergence rate given by

$$(\tau_n^2 \Delta_n)^{\frac{\alpha}{2+2\alpha}}.$$

This rate is polynomial in Δ_n even for τ_n constant, which is a great improvement compared to the i.i.d. case. Note also that this rate does not depend on the dimension, see Section 4.1 for insights.

1.3 Organization of the paper

In Section 2, we present the statistical model, including the underlying assumptions and the probabilistic framework of our analysis. We provide clear definitions for our observation scheme and outline the requirements for the noise source. In Section 3, we systematically construct the kernel estimator for the invariant density $\hat{\mu}_{n,h,p}$. This estimator differs from the standard kernel estimator due to the need to preprocess the data, which arises from the presence of noise. To reduce the impact of noise, we use a preaveraging strategy. Specifically, for some integer $p \geq 1$, we divide our number of observations by p by averaging them over a range of size p . Section 4 presents the upper bounds for the quadratic risk

$\mathbb{E}[|\hat{\mu}_{n,h,p}(x) - \bar{\mu}^b(x)|^2]$, providing insights into the optimal hyperparameters selection. Although the obtained rates may not achieve minimax optimality, they align with the expected non-parametric estimation rates in similar contexts. Furthermore, it enables us to precisely understand the noise intensity threshold above which averaging is required to achieve a better convergence rate, given by $\Delta_n^{1/\alpha}$, whenever $\bar{\mu}^b$ is assumed to be α -Hölder. Section 5 includes the derivation of a Bernstein-type concentration inequality allowing adaptive selection of the bandwidth h . We then proceed to a numerical analysis section (Section 6), which encompasses experiments and discussions on the estimation procedure. This is followed by a section consolidating essential probabilistic results instrumental in variance control of our estimator, see Section 7. Lastly, all proofs are gathered in the Appendix.

2 Statistical and Probabilistic framework

2.1 Notation

For all $x \in \mathbb{R}^d$, we denote by $|x|^2 := x \cdot x$ the Euclidian norm. Throughout the paper, we denote by $\mathcal{P}(\mathbb{R}^d)$ the space of probability measures on \mathbb{R}^d . Moreover, for any differentiable function, $f : \mathbb{R}^d \rightarrow \mathbb{R}$, ∇f stands for the gradient of f . Similarly, if f admits k derivative in the i -th component, we denote this derivative by $\partial_i^k f$. Finally, for any $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$, $\nabla \cdot g$ denotes the divergence of g .

For any σ -finite measure ν on \mathbb{R}^d , for any $q \geq 1$, we say that $f : \mathbb{R}^d \rightarrow \mathbb{R}$ belongs to $L^q(\nu)$ whenever

$$\|f\|_{L^q(\nu)}^q := \int_{\mathbb{R}^d} |f(x)|^q \nu(dx) < +\infty.$$

When ν is the Lebesgue measure on \mathbb{R}^d , we only denote L^q and the associated norm $\|\cdot\|_q$. Finally, when $q = +\infty$ and ν is the Lebesgue measure, we define

$$\|f\|_\infty := \sup_{x \in \mathbb{R}^d} |f(x)|.$$

2.2 The statistical diffusion model

We consider a stochastic process X defined on a rich enough probability space $(\Omega, \mathcal{A}, \mathbb{P})$. For a given Lipschitz continuous function b and probability measure μ on \mathbb{R}^d , we can define a probability \mathbb{P}_μ^b under which X is solution of the following stochastic differential equation

$$dX_t = b(X_t) dt + dW_t, \quad \mathcal{L}(X_0) = \mu, \quad (4.4)$$

where W is a d -dimensional \mathbb{P}_μ^b -Brownian motion. We denote by $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ the filtration generated by W . We write \mathbb{E}_μ^b for the expectation with respect to \mathbb{P}_μ^b . We also use \mathbb{P}_x^b and \mathbb{E}_x^b instead of $\mathbb{P}_{\delta_x}^b$ and $\mathbb{E}_{\delta_x}^b$. In the paper, for any probability measure $\nu \in \mathcal{P}(\mathbb{R}^d)$, for any $q > 1$, we say that $f : \mathbb{R}^d \rightarrow \mathbb{R}$ belongs to $L^q(\nu)$ whenever

$$\|f\|_{L^q(\nu)}^q := \int_{\mathbb{R}^d} |f(x)|^q \nu(dx) < +\infty.$$

When ν is the Lebesgue measure on \mathbb{R}^d , we only denote L^q and the associated norm $\|\cdot\|_q$. In the following, we always assume the following conditions on b .

Assumption 7. *The function b is differentiable and satisfies*

$$|b(0)| \leq b_0 \quad \text{and} \quad \forall i \in \{1, \dots, d\}, \quad \|\partial_i b\|_\infty \leq b_1/d,$$

where b_0 and b_1 are positive constants. This ensures in particular that $\|\nabla \cdot b\|_\infty \leq b_1$.

Assumption 8. *There exists a function $V : \mathbb{R}^d \rightarrow \mathbb{R}$ differentiable and bounded below by a constant V_0 such that $b = -\nabla V$ and $V(0) = 0$.*

Assumption 9. *There exists $\tilde{C}_b > 0$ and $\tilde{\rho}_b > 0$ such that $\langle x, b(x) \rangle \leq -\tilde{C}_b|x|$, $\forall x : |x| \geq \tilde{\rho}_b$.*

By definition, X solution of (4.2) is a Markov process. Assumption 7 shows that the drift force exhibits at most linear growth, which implies that there exists a constant $C_0 > 0$ such that the transition density $p_t^b(x, y)$ for all $t > 0$ and for all $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ with $|x - y|^2 < t$ is satisfies

$$p_t^b(x, y) \leq C_0(t^{-d/2} + t^{3d/2}), \quad (4.5)$$

see for example [CW97]. This inequality is crucial to obtain robust upper bounds when estimating the invariant measure of a stochastic process, as demonstrated, for instance in [DR07, Str18]. Assumption 8 is also usual when studying the asymptotic behaviour of a diffusion process. However, it is not a necessary condition [Bha78, DR06], since the existence and uniqueness of the invariant measure primarily depend on the asymptotic behaviour of the drift force b and its growth at infinity. In the present paper, Assumption 8 is needed at a later stage in order to derive upper bounds similar to (4.5) for the transition density of the pre-averaged process, see Section 14.1.

Under Assumptions 8 and 9, the process X defined by (4.4) admits a unique stationary distribution denoted by $\bar{\mu}^b$, absolutely continuous with respect to the Lebesgue measure on \mathbb{R}^d whose density, also denoted by $\bar{\mu}^b$, is explicitly given for all $x \in \mathbb{R}^{2d}$ by

$$\bar{\mu}^b(x) = Z_V^{-1} \exp(-2V(x)) \quad \text{where} \quad Z_V = \int_{\mathbb{R}^d} \exp(-2V(y)) dy.$$

Assumption 9 is common to guarantee an exponentially fast convergence towards equilibrium. In the case where the drift force is derived from a potential V , a direct link exists between the classical Poincaré inequality and Equation 4.6, and Assumption 9 implies that X satisfies a Poincaré inequality as shown in [BGL14]. More precisely, there exists $C_{PI} > 0$, depending only on \tilde{C}_b , such that for any $f \in L^2(\bar{\mu}^b)$ satisfying $\mathbb{E}_{\bar{\mu}^b}^b[f(X_0)] = 0$ and any $t \geq 0$, we have

$$\text{Var}_{\bar{\mu}^b}^b[P_t^b f(X_0)] \leq e^{-2tC_{PI}^{-1}} \mathbb{E}_{\bar{\mu}^b}^b[f(X_0)^2], \quad (4.6)$$

where $(P_t^b)_{t \geq 0}$ is the semi-group associated to the process X , acting on any measurable functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$ by

$$\forall t \geq 0, \forall x \in \mathbb{R}^d, P_t^b f(x) = \mathbb{E}_x^b[f(X_t)]. \quad (4.7)$$

It is well known that the accuracy of the estimation of the invariant density $\bar{\mu}^b$ strongly depends on the regularity of $\bar{\mu}^b$ [DR07, Str18, AG21, AG23]. Therefore, we assume that $\bar{\mu}^b$ belongs to the anisotropic Hölder class $\mathcal{H}_d(\alpha, \mathcal{L})$, which is defined below.

Definition 4.1. Let $\alpha = (\alpha_1, \dots, \alpha_d) \in (0, \infty)^d$ and $\mathcal{L} = (\mathcal{L}_1, \dots, \mathcal{L}_d) \in (0, \infty)^d$. A function $g : \mathbb{R}^d \rightarrow \mathbb{R}$ is said to belong to the anisotropic Hölder class $\mathcal{H}_d(\alpha, \mathcal{L})$ of functions if, for all $1 \leq i \leq d$, g is $\lfloor \alpha_i \rfloor$ -differentiable in the i -th variable and the partial derivatives satisfy for $0 \leq k \leq \lfloor \alpha_i \rfloor$

$$\|\partial_i^k g\|_\infty \leq \mathcal{L}_i \quad \text{and} \quad \forall t \in \mathbb{R}, \|\partial_i^{\lfloor \alpha_i \rfloor} g(\cdot + te_i) - \partial_i^{\lfloor \alpha_i \rfloor} g(\cdot)\|_\infty \leq \mathcal{L}_i |t|^{\alpha_i - \lfloor \alpha_i \rfloor}$$

where (e_1, \dots, e_d) is the canonical basis of \mathbb{R}^d .

Definition 4.2. Let $\alpha = (\alpha_1, \dots, \alpha_d) \in (0, \infty)^d$, $\mathcal{L} = (\mathcal{L}_1, \dots, \mathcal{L}_d) \in (0, \infty)^d$ and $b_0, b_1 > 0$. We write $\mathbf{b} = (\alpha, \mathcal{L}, b_0, b_1, V_0)$ and $\Sigma(\mathbf{b})$ the set of functions $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ satisfying Assumption 7 with constants b_0 and b_1 , Assumption 8 with constant V_0 , Assumption 9 and such that $\bar{\mu}^b$ belongs to the anisotropic Hölder class $\mathcal{H}_d(\alpha, \mathcal{L})$.

In this paper, we always assume that b belongs to $\Sigma(\mathbf{b})$, for some $\mathbf{b} = (\alpha, \mathcal{L}, b_0, b_1, V_0)$, with $\alpha_1 \leq \dots \leq \alpha_d$. Moreover, X is assumed to be observed at discrete times $i\Delta_n$, for $0 \leq i \leq n$, and blurred by a noise composed of independent standard Gaussian variables. We observe

$$Y_{i,n} = X_{i\Delta_n} + \tau_n \xi_{i,n},$$

where $\xi_{i,n}$ are i.i.d Gaussian variables, $\Delta_n \rightarrow 0$ and $n\Delta_n \rightarrow \infty$.

3 Estimation procedure

In this section, we plan to use a kernel type estimation procedure. We consider a bounded kernel with compact support $K : \mathbb{R} \rightarrow \mathbb{R}$, that is a measurable function such that $\int_{\mathbb{R}} K(y) dy = 1$. We assume that it is of order $l \geq 1$, i.e. that for all $1 \leq k \leq l - 1$, we have

$$\int_{\mathbb{R}} y^k K(y) dy = 0. \tag{4.8}$$

Now for any $\mathbf{h} = (h_1, \dots, h_d) \in (0, +\infty)^d$ and $y = (y_1, \dots, y_d) \in \mathbb{R}^d$, we define

$$\mathbf{K}_{\mathbf{h}}(y) = \prod_{i=1}^d h_i^{-1} K(y_i h_i^{-1}). \tag{4.9}$$

In the case where the density of interest $\bar{\mu}^b$ belongs to the anisotropic Hölder class $\mathcal{H}_d(\alpha, \mathcal{L})$, we always assume that $l \geq \lfloor \alpha_d \rfloor$.

First, note that the natural kernel based estimator of the invariant density in absence of noise does

not work in our context. Indeed, since $(Y_{k,n})$ is a stationary process, the estimator

$$\hat{\mu}_{KB,n,h,p}(x) = \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{K}_h(x - Y_{k,n}) \quad (4.10)$$

estimates the density of $Y_{1,n}$, which is given by

$$\bar{\mu}^b * \varphi_{\tau_n}(x) = \int_{\mathbb{R}^d} \bar{\mu}^b(x - y) \varphi_{\tau_n}(y) dy = \int_{\mathbb{R}^d} \bar{\mu}^b(x - \tau_n y) \varphi_1(\tau_n y) dy$$

where $\varphi_{\tau}(x) = (2\pi\tau^2)^{-d/2} \exp(-|x|^2/(2\tau^2))$. Therefore, the presence of noise in (4.10) creates an additional bias satisfying

$$|\bar{\mu}^b * \varphi_{\tau_n}(x) - \mu(x)| \leq \int_{\mathbb{R}^d} |\bar{\mu}^b(x - \tau_n y) - \bar{\mu}^b(x)| \varphi_1(\tau_n y) dy \quad (4.11)$$

which is of order τ_n . This bias is dominating the usual bias of kernel based estimators when τ_n is large. In particular, when τ_n is of order 1, the estimator is not consistent. Therefore, we need to reduce the influence of the noise in the observation.

To that extent, we implement in this paper a preaveraging approach and we compute local average the observations over batches of size p . This yields to the following modified observations $(p^{-1} \sum_{\ell=0}^{p-1} Y_{kp+\ell,n})_{1 \leq k \leq \lfloor n/p \rfloor}$. Intuitively, this approach should work because of the regularity of X . Indeed, we have

$$\frac{1}{p} \sum_{\ell=0}^{p-1} Y_{kp+\ell,n} = X_{kp\Delta_n} + \frac{1}{p} \sum_{\ell=0}^{p-1} (X_{(kp+\ell)\Delta_n} - X_{kp\Delta_n}) + \frac{\tau_n}{p} \sum_{\ell=0}^{p-1} \xi_{kp+\ell,n}. \quad (4.12)$$

This expression can be seen as $X_{kp\Delta_n} + \text{noise}$ and the effective sampling frequency becomes $(p\Delta_n)^{-1}$. Moreover, the noise now comes from two sources: in addition to the noise $(\xi_{i,n})_{i,n}$, we now have a preaveraging error $p^{-1} \sum_{\ell=0}^{p-1} (X_{(kp+\ell)\Delta_n} - X_{kp\Delta_n})$. Using the pathwise regularity of X , this error should remain small when p is not too large. Moreover, since the random variable $(\xi_{i,n})_{i,n}$ are independent standard Gaussian variables, we can rewrite $p^{-1} \sum_{\ell=0}^{p-1} \xi_{kp+\ell,n} = \tau_n p^{-1/2} \tilde{\xi}_{k,n}$ where $\tilde{\xi}_{k,n}$ is also a standard Gaussian variable. We next define a kernel estimator based on the preaveraged observations. For any non negative integer p and any $x \in \mathbb{R}^d$, we define

$$\hat{\nu}_{n,h,p}(x) = \frac{1}{\lfloor n/p \rfloor} \sum_{k=0}^{\lfloor n/p \rfloor - 1} \mathbf{K}_h\left(x - p^{-1} \sum_{\ell=0}^{p-1} Y_{kp+\ell,n}\right). \quad (4.13)$$

Remark that when $p = 1$, this estimator is the usual estimator used for estimating the invariant density from discrete observations, see [Str18, AG23]. From what precedes, it first seems natural that $\hat{\nu}_{n,h,p}$ estimates $\bar{\mu}^b * \varphi_{\tau_n p^{-1/2}}$ which is the invariant density of $X_{kp\Delta_n} + \tau_n p^{-1/2} \tilde{\xi}_{k,n}$. However, the preaveraging

error $p^{-1} \sum_{\ell=0}^{p-1} (X_{(kp+\ell)\Delta_n} - X_{kp\Delta_n})$ induces an additional term. Indeed, we have

$$\frac{1}{p} \sum_{\ell=0}^{p-1} (X_{(kp+\ell)\Delta_n} - X_{kp\Delta_n}) = \frac{1}{p} \sum_{\ell=0}^{p-1} \left(\int_{kp\Delta_n}^{(kp+\ell)\Delta_n} b(X_s) ds \right) + \frac{1}{p} \sum_{\ell=0}^{p-1} (W_{(kp+\ell)\Delta_n} - W_{kp\Delta_n}). \quad (4.14)$$

The first sum is negligible compared to the second one. Moreover, $p^{-1} \sum_{\ell=0}^{p-1} (W_{(kp+\ell)\Delta_n} - W_{kp\Delta_n})$ is a centered Gaussian variable with variance $(12p)^{-1}(p-1)(2p-1)\Delta_n$, independent of $X_{kp\Delta_n}$. Therefore, this term has an effect on the estimation comparable to that of $\tau_n p^{-1/2} \tilde{\xi}_{k,n}$. Combining these two terms, we deduce that $\hat{\nu}_{n,h,p}$ estimates $\bar{\mu}^b * \varphi_{\tilde{\tau}_{n,p}}$ where

$$\tilde{\tau}_{n,p} = \left(\frac{\tau_n^2}{p} + \frac{(p-1)(2p-1)\Delta_n}{12p} \right)^{1/2}. \quad (4.15)$$

This analysis is formalised in the following proposition.

Proposition 4.3. *Under Assumptions 7, 8 and 9, there exists $C_B > 0$ such that for any $p \in \{1, \dots, \lfloor \Delta_n^{-1/2} \rfloor\}$ and any $x \in \mathbb{R}^d$, we have*

$$\left| \mathbb{E}_{\bar{\mu}^b}[\hat{\nu}_{n,h,p}(x)] - \bar{\mu}^b * \varphi_{\tilde{\tau}_{n,p}}(x) \right| \leq C_B \begin{cases} \sum_i h_i^{\alpha_i} & \text{if } p = 1, \\ \sum_i h_i^{\alpha_i} + \sqrt{p\Delta_n} & \text{if } p \geq 2. \end{cases} \quad (4.16)$$

Although bounding the bias of kernel based estimators is usually an easy task, this is not the case here. The proof of Proposition 4.3 is indeed quite delicate and can be found in Section 9. Indeed, we cannot use Itô's formula because, from Equation 4.9, we see that for any $i \in \{1, \dots, d\}$, $\partial_i \mathbf{K}_h(x) = h_i^{-1} \prod_j h_j^{-1} K(h_j^{-1} x_j)$ which would interfere and create a contribution of order $\prod_i h_i^{-1}$. Instead, using crucially Assumption 8, we use Girsanov's theorem to remove the contribution of the drift term in (4.14). We then control the likelihood introduced via the change of measure which introduce the additional term $\sqrt{p\Delta_n}$ in (4.16) when $p \geq 2$.

Proceeding as in (4.11), we see that (4.16) implies

$$\left| \mathbb{E}_{\bar{\mu}^b}[\hat{\nu}_{n,h,p}(x)] - \bar{\mu}^b(x) \right| \leq \tilde{C}_B \left(\sum_i h_i^{\alpha_i} + \tilde{\tau}_{n,p} \right) \quad (4.17)$$

for some $\tilde{C}_B \geq C_B$. This can be improved by a deconvolution procedure. To do so, note that

$$\bar{\mu}^b * \varphi_{\tilde{\tau}_{n,p}}(x) = \mathbb{E}[\bar{\mu}^b(x - \tilde{\tau}_{n,p}\zeta)]$$

where ζ is a standard d -dimensional Gaussian variable. Using the regularity of $\bar{\mu}^b$, we can proceed to a Taylor expansion of $\bar{\mu}^b(x + \tilde{\tau}_{n,p}(\gamma - \zeta))$ around x for all $\gamma = (\gamma_1, \dots, \gamma_d) \in \{0, \dots, l\}^d$. Computing explicitly the moments of $(\gamma - \zeta)$ appearing in this expression, we get an explicit expansion of $\bar{\mu}^b * \varphi_{\tilde{\tau}_{n,p}}(x + \tilde{\tau}_{n,p}\gamma)$ around $\bar{\mu}^b(x)$. We can isolate $\bar{\mu}^b(x)$ from this expression. Specifically, we introduce

the matrix $A = (a_{k,i})_{0 \leq k, i \leq l}$, with coefficients given for any $k, i \in \{0, \dots, l\}$ by

$$a_{k,i} = \sum_{j=0}^k \binom{k}{j} (-1)^j m_j i^{k-j} \quad (4.18)$$

where m_j stands for the j -th moment of a standard Gaussian variable. The matrix A is invertible, as shown in Appendix 8. We denote its inverse by A^{-1} , and $\mathbf{u} = (u_0, \dots, u_l)$ stands for its first column. Then, for all $k \in \{0, \dots, l\}$,

$$\sum_{i=0}^l u_i \left(\sum_{j=0}^k \frac{(-1)^j m_j i^{k-j}}{j!(k-j)!} \right) = \begin{cases} 1 & \text{if } k = 0, \\ 0 & \text{otherwise.} \end{cases} \quad (4.19)$$

For any multi-index $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_d) \in \{0, \dots, l\}^d$, we define $\mathbf{u}_{\boldsymbol{\gamma}} = \prod_{i=1}^d u_{\gamma_i}$, and the following point-wise estimator

$$\hat{\mu}_{n, \mathbf{h}, p}(x) = \sum \mathbf{u}_{\boldsymbol{\gamma}} \hat{\nu}_{n, \mathbf{h}, p}(x + \boldsymbol{\gamma} \tilde{\tau}_{n, p}) \quad (4.20)$$

where the sum holds over all $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_d) \in \{0, \dots, l\}^d$. The bias of $\hat{\mu}_{n, \mathbf{h}, p}(x)$ is precised in the following proposition, proved in Section 10

Proposition 4.4. *Suppose that $b \in \Sigma(\mathbf{b})$, with $\boldsymbol{\alpha}$ such that $\alpha_1 \leq \dots \leq \alpha_d$. Then there exists a constant \tilde{C}_B depending only in \mathbf{b} so that for any $x \in \mathbb{R}^d$, any $p \in \{1, \dots, \lceil \Delta_n^{-1/2} \rceil\}$ and any $\mathbf{h} \in \mathbb{R}^d$, we have*

$$|\mathbb{E}_{\bar{\mu}^b} [\hat{\mu}_{n, \mathbf{h}, p}(x)] - \bar{\mu}^b(x)| \leq C_B \begin{cases} \tau_n^{\alpha_1} + \sum_{i=1}^d h_i^{\alpha_i} & \text{if } p = 1, \\ \sqrt{p \Delta_n} + \frac{\tau_n^{\alpha_1}}{p^{\alpha_1/2}} + \sum_{i=1}^d h_i^{\alpha_i} & \text{if } p \geq 2. \end{cases} \quad (4.21)$$

We now turn to the variance of $\hat{\mu}_{n, \mathbf{h}, p}(x)$. Before stating the upper bound of the variance let us now define $k_0 := k_0(\boldsymbol{\alpha})$ such that $\alpha_1 = \alpha_2 = \dots = \alpha_{k_0} < \alpha_{k_0+1} \leq \dots \leq \alpha_d$. Let us also define

$$\begin{aligned} D_1 &= \{(\boldsymbol{\alpha}, k_0), k_0 = 1 \text{ or } k_0 = 2 \text{ and } \alpha_2 < \alpha_3\}; \\ D_2 &= \{(\boldsymbol{\alpha}, k_0), k_0 \geq 3\}; \\ D_3 &= \{(\boldsymbol{\alpha}, k_0), k_0 = 1 \text{ and } \alpha_2 = \alpha_3\}. \end{aligned} \quad (4.22)$$

The upper bound is stated in the following proposition.

Proposition 4.5. *Suppose that $b \in \Sigma(\mathbf{b})$. Suppose that $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)$ satisfies $\alpha_1 = \alpha_2 = \dots = \alpha_{k_0} < \alpha_{k_0+1} \leq \dots \leq \alpha_d$, for some $k_0 \in \{1, \dots, d\}$. If $\hat{\mu}_{n, \mathbf{h}, p}$ is the estimator proposed in (4.20), then there exist $C_V > 0$ uniform over $\Sigma(\mathbf{b})$ and $n_0 > 0$ such that, for $n \geq n_0$, the following holds true for all $x \in \mathbb{R}^d$ and all $\mathbf{h} \in (0, 1]^d$.*

- If $d = 1$, then

$$\text{Var}_{\bar{\mu}^b}(\hat{\mu}_{n, \mathbf{h}, p}(x)) \leq \frac{C_V}{T_n} \left(p \Delta_n h_1^{-1} + |\log(h_1)| \right). \quad (4.23)$$

- If $d = 2$, then

$$\text{Var}_{\bar{\mu}^b}^b(\hat{\mu}_{n,\mathbf{h},p}(x)) \leq \frac{C_V}{T_n} \left(p\Delta_n h_1^{-1} h_2^{-1} + |\log(p\Delta_n)| + |\log(h_1 h_2)| \right). \quad (4.24)$$

- If $d \geq 3$ and $(k_0, \boldsymbol{\alpha}) \in D_1$, then

$$\text{Var}_{\bar{\mu}^b}^b(\hat{\mu}_{n,\mathbf{h},p}(x)) \leq \frac{C_V}{T_n} \left(p\Delta_n \prod_{i=1}^d h_i^{-1} + \sum_{i=1}^d |\log(h_i)| \prod_{i=3}^d h_i^{-1} \right). \quad (4.25)$$

- If $d \geq 3$ and $(k_0, \boldsymbol{\alpha}) \in D_2$, then

$$\text{Var}_{\bar{\mu}^b}^b(\hat{\mu}_{n,\mathbf{h},p}(x)) \leq \frac{C_V}{T_n} \left(\left(\prod_{i=1}^{k_0} (h_i)^{(2-k_0)/k_0} \prod_{i=k_0+1}^d h_i^{-1} + p\Delta_n \prod_{i=1}^d h_i^{-1} + \sum_{i=1}^d |\log h_i| \right) \right). \quad (4.26)$$

- If $d \geq 3$ and $(k_0, \boldsymbol{\alpha}) \in D_3$, then

$$\text{Var}_{\bar{\mu}^b}^b(\hat{\mu}_{n,\mathbf{h},p}(x)) \leq \frac{C_V}{T_n} \left(\sum_{i=1}^d |\log h_i| + p\Delta_n \prod_{i=1}^d h_i^{-1} + (h_2 h_3)^{-1/2} \prod_{i=4}^d h_i^{-1} \right). \quad (4.27)$$

The proof of Proposition 4.5 is delayed to Appendix 11. It is important to point out that our results yield the same upper bounds for the variance term similar to those found in Proposition 1 of [AG23]. The only difference is the presence of a factor p in front of $\Delta_n (T_n \prod_{i=1}^d h_i)^{-1}$. This change is natural as this term directly comes from the discretisation of the process. Here, the preaveraging induces a sub-sampling of the data, grouping the observations on windows of length p . The discretisation step Δ_n therefore becomes $p\Delta_n$. As seen in [AG23], this term does not contribute, except when Δ_n is large. Here, the break-even point also depends on the noise intensity and is detailed in Section 4.1. Note that our proof and the proof Proposition 1 of [AG23] follows the same paths. However, the introduction of additive noise structure requires new bounds on transition densities of the preaveraged process, see Section 7. Note also that although Proposition 4.5 is stated on the final estimator $\hat{\mu}_{n,\mathbf{h},p}(x)$, the same result holds for $\hat{\nu}_{n,\mathbf{h},p}(x)$. Indeed, in Appendix 11, we prove it for $\hat{\nu}_{n,\mathbf{h},p}(x)$ and the proof for $\hat{\mu}_{n,\mathbf{h},p}(x)$ follows from

$$\text{Var}_{\bar{\mu}^b}^b(\hat{\mu}_{n,\mathbf{h},p}(x)) \leq \left(\sum_{\gamma} |u_{\gamma}| \right)^{2d} \sup_{y \in \mathbb{R}^d} \text{Var}_{\bar{\mu}^b}^b(\hat{\nu}_{n,\mathbf{h},p}(y)) \quad (4.28)$$

and the fact that $(\sum_{\gamma} |u_{\gamma}|)^{2d}$ is a constant independent of n .

4 Upper bounds and hyper-parameter choice

We quantify the quality of this estimation procedure by deriving an upper bound for the quadratic error

$$\mathcal{R}(\hat{\mu}, b; x) = \mathbb{E}_{\bar{\mu}^b}^b[|\hat{\mu}(x) - \bar{\mu}^b(x)|^2].$$

Using the classical Bias-Variance decomposition of the quadratic error

$$\mathcal{R}(\hat{\mu}_{n,\mathbf{h},p}, b; x) = B_{n,\mathbf{h},p}^b(x)^2 + V_{n,\mathbf{h},p}^b(x), \quad (4.29)$$

where for all $x \in \mathbb{R}^d$, we have

$$B_{n,\mathbf{h},p}^b(x) = |\mathbb{E}_{\bar{\mu}^b}[\hat{\mu}_{n,\mathbf{h},p}(x)] - \bar{\mu}^b(x)| \quad \text{and} \quad V_{n,\mathbf{h},p}^b(x) = \text{Var}_{\bar{\mu}^b}[\hat{\mu}_{n,\mathbf{h},p}(x)].$$

Upper bounds for both $B_{n,\mathbf{h},p}^b(x)$ and $V_{n,\mathbf{h},p}^b(x)$ are given by Propositions 4.4 and 4.5 respectively. The variance obtained in Proposition 4.5 depends on many case and the choice of the hyper-parameters p and \mathbf{h} naturally depends on these cases. In this section, we detail each case and the convergence rate that can be obtained using the best choices p^* and \mathbf{h}^* . For future use, we write

$$\bar{\alpha} = \left(\frac{1}{d} \sum_{i=1}^d \alpha_i^{-1} \right)^{-1} \quad \text{and} \quad \bar{\alpha}_3 = \left(\frac{1}{d-2} \sum_{i=3}^d \alpha_i^{-1} \right)^{-1}. \quad (4.30)$$

4.1 General results

Before studying the quadratic risk in this model, we first recall the results from [AG23] where the estimation of $\bar{\mu}^b$ is studied from discrete non-noisy observations. In [AG23], the authors distinguish two regimes depending on Δ_n . The breaking point w_n^{HF} between these two regimes is defined by

$$w_n^{HF} = \begin{cases} \log(T_n)T_n^{-1} & \text{if } d = 1, 2, \\ \log(T_n) \left(\frac{\log(T_n)}{T_n} \right)^{\frac{\bar{\alpha}_3}{(2\bar{\alpha}_3+d-2)} \left(\frac{1}{\alpha_1} + \frac{1}{\alpha_2} \right)} & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_1, \\ T_n^{-\frac{2\bar{\alpha}_3}{2\bar{\alpha}_3+d-2} \alpha_1^{-1}} & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_2, \\ T_n^{\frac{-\bar{\alpha}_3}{(2\bar{\alpha}_3+d-2)} \left(\frac{1}{\alpha_1} + \frac{1}{\alpha_2} \right)} & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_3. \end{cases}$$

They show that, when $\Delta_n \leq w_n^{HF}$, the estimator of the invariant measure behaves as in the continuous observation case and therefore exhibits the convergence rate $(v_n^{HF})^{1/2}$ defined as

$$v_n^{HF} = \begin{cases} \log(T_n)T_n^{-1} & \text{if } d = 1, 2, \\ \left(\frac{\log(T_n)}{T_n} \right)^{\frac{2\bar{\alpha}_3}{2\bar{\alpha}_3+d-2}} & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_1, \\ T_n^{-\frac{2\bar{\alpha}_3}{2\bar{\alpha}_3+d-2}} & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_2, \\ T_n^{-\frac{2\bar{\alpha}_3}{2\bar{\alpha}_3+d-2}} & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_3. \end{cases}$$

In order to get this convergence rate, when $d \geq 3$ [AG23] identifies the optimal bandwidth choice given by

$$h_i^{*,HF} = \begin{cases} \left(\frac{\log(T_n)}{T_n} \right)^{\frac{\bar{\alpha}_3}{\alpha_i(2\bar{\alpha}_3+d-2)}} & \text{if } \alpha_2 = \alpha_3, \\ T_n^{-\frac{\bar{\alpha}_3}{\alpha_i(\bar{\alpha}_3+d-2)}} & \text{if } \alpha_2 < \alpha_3. \end{cases}$$

When $d = 1, 2$, the choice of the bandwidth in the high frequency framework is given by $h_i^{*,HF} = T_n^{-1/2}$, for $i \in \{1, \dots, d\}$. When $\Delta_n \geq w_n^{HF}$, the estimator of the invariant measure behaves as in the i.i.d. case and exhibits a convergence rate $n^{\frac{\alpha}{2\bar{\alpha}+d}}$. Here, we want to reproduce this behaviour while also incorporating noise. In order to get this convergence rate, [AG23] also identifies the optimal bandwidth choice $\mathbf{h}^{*,LF}$ in that case, given by

$$h_i^{*,LF} = n^{\frac{-\alpha}{\alpha_i(2\bar{\alpha}+d)}}.$$

More over, the convergence rate of the low frequency estimator is $(v_n^{LF})^{1/2}$ where

$$v_n^{LF} = n^{-\frac{2\bar{\alpha}}{2\bar{\alpha}+d}}.$$

In our case, the effective discretization step is $p\Delta_n$ and therefore, we expect that the switch between the two regimes of [AG23] appear when $p\Delta_n \approx w_n^{HF}$. However, p also needs to be chosen to minimize the quadratic error bound. Applying the same analysis as in [AG23], we then define $\mathbf{h}^{*,p}$ for each p by

$$h_i^{*,p} = \begin{cases} h_i^{*,HF} & \text{if } p\Delta_n \leq w_n^{HF}, \\ \left(\frac{p}{n}\right)^{\frac{\alpha}{\alpha_i(2\bar{\alpha}+d)}} & \text{if } p\Delta_n \geq w_n^{HF}. \end{cases}$$

Then we get for each $p \geq 1$,

$$\mathcal{R}(\hat{\mu}_{n,\mathbf{h}^{*,p}}, b; x) \lesssim \begin{cases} p\Delta_n \mathbb{1}_{p \geq 2} + \frac{\tau_n^{2\alpha_1}}{p^{\alpha_1}} + v_n^{HF} & \text{if } p\Delta_n \leq w_n^{HF}, \\ p\Delta_n \mathbb{1}_{p \geq 2} + \frac{\tau_n^{2\alpha_1}}{p^{\alpha_1}} + \left(\frac{p}{n}\right)^{\frac{2\bar{\alpha}}{\alpha_i(2\bar{\alpha}+d)}} & \text{if } p\Delta_n \geq w_n^{HF}. \end{cases} \quad (4.31)$$

We then want to simplify the case $p\Delta_n \geq w_n^{HF}$ by removing one of the dependence in p . To do so, we use that the condition $p\Delta_n \geq w_n^{HF}$ implies that $p\Delta_n \gtrsim \left(\frac{p}{n}\right)^{\frac{2\bar{\alpha}}{\alpha_i(2\bar{\alpha}+d)}}$. This statement is formalised in the following Lemma, proved in Section 12.1.

Lemma 4.6. *For all $c_1 > 0$, there exists $c_2 > 0$ such that if $p\Delta_n \geq c_1 w_n^{HF}$, then $p\Delta_n \geq c_2 \left(\frac{p}{n}\right)^{\frac{2\bar{\alpha}}{\alpha_i(2\bar{\alpha}+d)}}$*

Using Lemma 4.6, we simplify (4.31). If $p = 1$, we have

$$\mathcal{R}(\hat{\mu}_{n,\mathbf{h}^{*,1}}, b; x) \lesssim \begin{cases} \tau_n^{2\alpha_1} + v_n^{HF} & \text{if } \Delta_n \leq w_n^{HF}, \\ \tau_n^{2\alpha_1} + n^{-\frac{2\bar{\alpha}}{2\bar{\alpha}+d}} & \text{if } \Delta_n \geq w_n^{HF}. \end{cases} \quad (4.32)$$

and for $p \geq 2$, we have

$$\mathcal{R}(\hat{\mu}_{n,\mathbf{h}^{*,p}}, b; x) \lesssim \begin{cases} p\Delta_n + \frac{\tau_n^{2\alpha_1}}{p^{\alpha_1}} + v_n^{HF} & \text{if } p\Delta_n \leq w_n^{HF}, \\ p\Delta_n + \frac{\tau_n^{2\alpha_1}}{p^{\alpha_1}} & \text{if } p\Delta_n \geq w_n^{HF}. \end{cases} \quad (4.33)$$

From these, we see that the optimal choice of p is given by

$$p^* = \left[\left(\tau_n^{2\alpha_1} \Delta_n^{-1} \right)^{1/(1+\alpha_1)} \right] \vee 1 \quad (4.34)$$

and from this choice, we take $\mathbf{h}^* = \mathbf{h}^{*,p^*}$. Note that this choice ensure that $p^* \Delta_n^{1/2}$ is bounded so that

Proposition 4.4 applies. Four asymptotic regimes can be observed.

Proposition 4.7 (Small noise intensity, high sampling frequency). *Suppose that $p^* \Delta_n \leq w_n^{HF}$ and that $\tau_n^{2\alpha_1} \leq \Delta_n$. In that case, we have*

$$p^* = 1 \quad \text{and} \quad \mathbf{h}^* = \mathbf{h}^{*,HF}$$

and we get

$$\mathcal{R}(\hat{\mu}_{n,\mathbf{h}^*,p^*}, b; x) \lesssim v_n^{HF} \vee \tau_n^{\alpha_1}.$$

Proposition 4.8 (Large noise intensity, high sampling frequency). *Suppose that $p^* \Delta_n \leq w_n^{HF}$ and that $\tau_n^{2\alpha_1} \geq \Delta_n$. In that case, we have*

$$p^* = \left\lceil (\tau_n^{2\alpha_1} \Delta_n^{-1})^{1/(1+\alpha_1)} \right\rceil \quad \text{and} \quad \mathbf{h}^* = \mathbf{h}^{*,HF}$$

and we get

$$\mathcal{R}(\hat{\mu}_{n,\mathbf{h}^*,p^*}, b; x) \lesssim v_n^{HF} \vee (\tau_n^2 \Delta_n)^{\frac{\alpha_1}{1+\alpha_1}}.$$

Remark 4.9. *In the previous proposition, note that since $\tau_n^{2\alpha_1} \geq \Delta_n$, the condition $p^* \Delta_n \lesssim w_n^{HF}$ is equivalent to $\Delta_n \lesssim \tau_n^{-2} (w_n^{HF})^{\frac{1+\alpha_1}{\alpha_1}}$.*

Proposition 4.10 (Small noise intensity, Low sampling frequency). *Suppose that $p^* \Delta_n \geq w_n^{HF}$ and that $\tau_n^{2\alpha_1} \leq \Delta_n$. In that case, we have*

$$p^* = 1 \quad \text{and} \quad \mathbf{h}^* = \mathbf{h}^{*,1}$$

and we get

$$\mathcal{R}(\hat{\mu}_{n,\mathbf{h}^*,p^*}, b; x) \lesssim n^{-\frac{2\bar{\alpha}}{2\bar{\alpha}+d}} \vee \tau_n^{2\alpha_1}.$$

Proposition 4.11 (Large noise intensity, Low sampling frequency). *Suppose that $p^* \Delta_n \geq w_n^{HF}$ and that $\tau_n^{2\alpha_1} \geq \Delta_n$. In that case, we have*

$$p^* = \left\lceil (\tau_n^{2\alpha_1} \Delta_n^{-1})^{1/(1+\alpha_1)} \right\rceil \quad \text{and} \quad \mathbf{h}^* = \mathbf{h}^{*,p^*}$$

and we get

$$\mathcal{R}(\hat{\mu}_{n,\mathbf{h}^*,p^*}, b; x) \lesssim (\tau_n^2 \Delta_n)^{\frac{\alpha_1}{1+\alpha_1}}.$$

4.2 Analysis of the convergence rate

In this section, we study the convergence rates found previously in the setup $\Delta_n = n^{-\theta}$ and $\tau_n = n^{-\kappa}$. Of course, since $\Delta_n \rightarrow 0$ and $T_n = n\Delta_n \rightarrow \infty$, we must impose $0 < \theta < 1$. We also require $\kappa \geq 0$ for convenience, although a deeper analysis shows that we can still estimate the invariant measure when $\kappa > -\theta/(2\alpha_1)$. For conciseness, we also ignore the logarithmic factors in the rates w_n^{HF} and v_n^{HF} . We also introduce the following notations

$$\bar{\beta} = \frac{2\bar{\alpha}}{2\bar{\alpha} + d} \quad \text{and} \quad \bar{\beta}_3 = \frac{2\bar{\alpha}_3}{2\bar{\alpha}_3 + d - 2}.$$

Rewriting the conditions for each proposition shows that

- Proposition 4.7 applies when $\kappa \geq \theta/(2\alpha_1)$ and $\theta \geq \bar{\beta}(\alpha_1^{-1} + \alpha_2^{-1})/2$,
- Proposition 4.8 applies when $\kappa \leq \theta/(2\alpha_1)$ and $\kappa \geq (1 - \theta)\bar{\beta}_3(1 + \alpha_1)(\alpha_1^{-1} + \alpha_2^{-1})/2 - \theta\alpha_1$,
- Proposition 4.10 applies when $\kappa \geq \theta/(2\alpha_1)$ and $\theta \leq \bar{\beta}(\alpha_1^{-1} + \alpha_2^{-1})/2$,
- Proposition 4.11 applies when $\kappa \leq \theta/(2\alpha_1)$ and $\kappa \leq (1 - \theta)\bar{\beta}_3(1 + \alpha_1)(\alpha_1^{-1} + \alpha_2^{-1})/2 - \theta\alpha_1$.

We can then identify the domains where each convergence rates operate and we refer to Figure ?? for an illustration of the different regimes

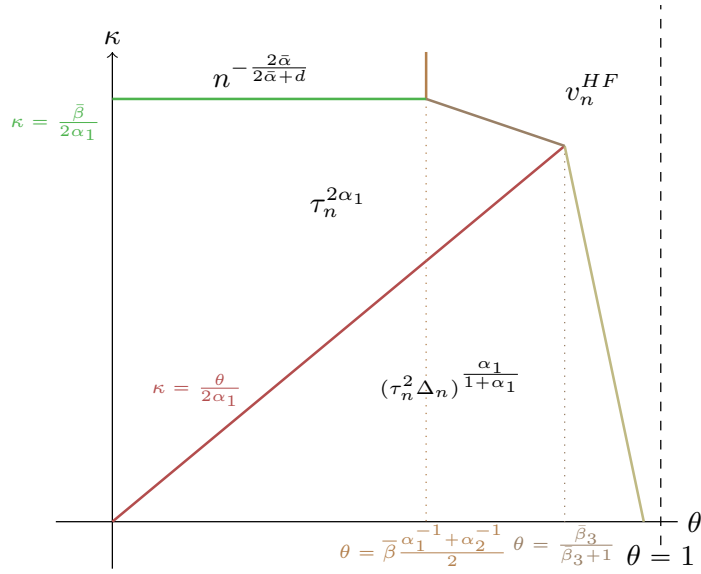


Figure 4.1: Rates of convergence of $\bar{\mu}_{n,h,p}(x)$.

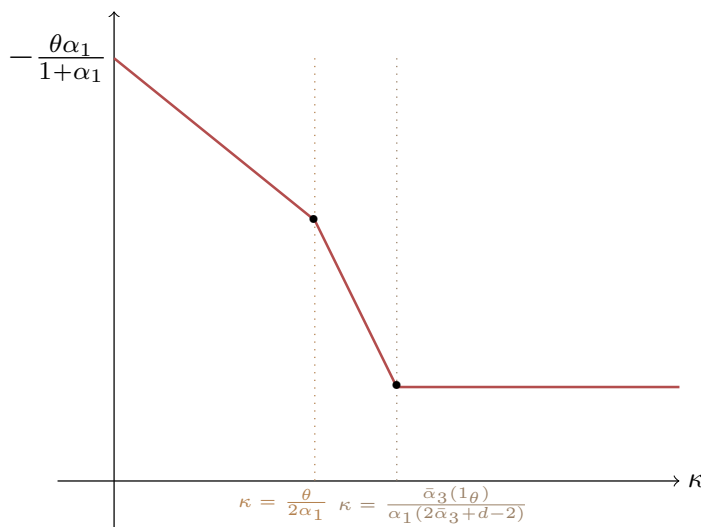


Figure 4.2: Evolution of the rate of convergence as a function of κ with fixed θ in a logarithmic scale.

4.3 Comments

The rate w_n^{HF} is the same in the cases $d \geq 3, k_0 \geq 3$ and $d \geq 3, k_0 = 1$ and $\alpha_2 = \alpha_3$. The same remark holds for v_n^{HF} and in these two cases, the convergence rate of the estimator is therefore the same. We still distinguish these cases here because the proofs differ slightly. As noted in [AG23], it is not clear that the high-frequency and the low-frequency rates meet when $\Delta_n \approx w_n^{HF}$.

5 Bernstein inequality and adaptive choice of the hyperparameters

In this section, we introduce a concentration inequality of Bernstein's type for the kernel estimator built in Section 3. This inequality provides a robust framework for analyzing the estimator's performance and variance. Leveraging this concentration inequality, we develop an adaptive approach for hyperparameter selection in a data-driven way.

5.1 Bernstein inequality

Recalling that D_1, D_2 and D_3 are defined in Equation (4.22), we get the following result,

Theorem 4.12 (Bernstein inequality). *Suppose that $b \in \Sigma(\mathbf{b})$. Then there exists C, τ positive and uniformly bounded over $\Sigma(\mathbf{b})$ such that for all n, \mathbf{h}, p satisfying $p\Delta_n \leq 1$, we have*

$$\mathbb{P}_{\bar{\mu}^b}^b \left(|\hat{\mu}_{n, \mathbf{h}, p}(x) - \mathbb{E}_{\bar{\mu}^b}^b[\hat{\mu}_{n, \mathbf{h}, p}(x)]| > \varepsilon \right) \leq K \exp \left(- \frac{n_p^2 \varepsilon^2 \beta^2}{32n_p v^2(\alpha, n, \mathbf{h}, p) + \tau \beta \varepsilon \|\mathbf{K}_{\mathbf{h}}\|_{\infty} n_p \log n_p} \right)$$

where $n_p = \lfloor n/p \rfloor$, $\beta = 1/\sqrt{d\ell} \|u\|_2$, $\|u\|_2^2 = \sum_{\gamma} |u_{\gamma}|^2$ and

$$\begin{aligned} v^2(\alpha, n, \mathbf{h}, p) &= \text{Var}_{\bar{\mu}^b}^b \left(\mathbf{K}_{\mathbf{h}}(x - p^{-1} \sum_{\ell=0}^{p-1} Y_{\ell, n}) \right) \\ &\quad + 2 \sum_{k=1}^{\infty} \text{Cov}_{\bar{\mu}^b}^b \left(\mathbf{K}_{\mathbf{h}}(x - p^{-1} \sum_{\ell=0}^{p-1} Y_{\ell, n}), \mathbf{K}_{\mathbf{h}}(x - p^{-1} \sum_{\ell=0}^{p-1} Y_{k p + \ell, n}) \right). \end{aligned}$$

Moreover, we have the following estimates for $v^2(\alpha, n, \mathbf{h}, p)$:

$$v^2(\alpha, n, \mathbf{h}, p) \leq C \begin{cases} (h_1 h_2)^{-1} + \frac{|\log(p\Delta_n)|}{p\Delta_n} + \frac{|\log(h_1 h_2)|}{p\Delta_n} & \text{if } d = 1, 2, \\ \prod_{i=1}^d h_i^{-1} + \frac{\sum_{i=1}^d |\log(h_i)|}{p\Delta_n \prod_{i=3}^d h_i} & \text{if } d \geq 3 \text{ and } (\alpha, k_0) \in D_1, \\ \prod_{i=1}^d h_i^{-1} + \frac{\prod_{i=1}^{k_0} (h_i)^{2/k_0}}{p\Delta_n \prod_{i=1}^d h_i} + \frac{\sum_{i=1}^d |\log h_i|}{p\Delta_n} & \text{if } d \geq 3 \text{ and } (\alpha, k_0) \in D_2, \\ \prod_{i=1}^d h_i^{-1} + \frac{\sum_{i=1}^d |\log h_i|}{p\Delta_n} + \frac{1}{p\Delta_n (h_2 h_3)^{1/2} \prod_{i=4}^d h_i^2} & \text{if } d \geq 3 \text{ and } (\alpha, k_0) \in D_3. \end{cases}$$

for some constant $C > 0$ uniform over $\Sigma(\mathbf{b})$ and independent of n, \mathbf{h} and p .

The Bernstein Inequality presented in Theorem 4.12 is standard and based on [Lem21]. Our proof requires the derivation of an exponential rate of convergence toward the invariant measure for the Markov process embedding the preaveraged process $(p^{-1} \sum_{\ell=0}^{p-1} Y_{kp+\ell, n})_k$. We refer to Section 13 for details.

5.2 Choice of the hyperparameters

Bernstein inequality can be used to tune adaptively the parameter \mathbf{h} , following the ideas of [GL08, GL09, GL11]. Fix $p \geq 1$. As previously discussed, for dimensions $d = 1$ and $d = 2$, the variance's upper bound does not rely on the smoothness. Consequently, there is no advantage to adopting a data-oriented adaptive strategy for $d < 3$.

For $d \geq 3$, the strategy is to consider a range of possible bandwidths and to select the one that minimizes the error. For this aim, we introduce a heuristic representation of the bias and a penalty term proportional to the variance bound of Proposition ... The optimal bandwidths are those minimizing the sum of these two latter. Let us begin by introducing the grid

$$\mathcal{H}_n^p \subset \left\{ \mathbf{h} \in (0, 1]^d : h_1 \leq \dots \leq h_d, \forall l = 1, \dots, d \quad \left(\frac{\log(n_p)^3}{n_p} \right)^{1/d} \leq h_l \leq 1 \right\},$$

where we recall that n_p stands for $\lfloor n/p \rfloor$ for the sake of clarity. We also assume that $\#\mathcal{H}_n^p \leq T_n$. According to the set of candidate bandwidths, we can introduce the set of candidate estimators:

$$\mathcal{F}(\mathcal{H}_n^p) := \left\{ \hat{\mu}_{n, \mathbf{h}, p}(x) : x \in \mathbb{R}^d, \mathbf{h} \in \mathcal{H}_n^p \right\}.$$

The goal of this section is to choose an estimator in the family $\mathcal{F}(\mathcal{H}_n^p)$, in a completely data-driven way. Following the idea in [GL08, GL09, GL11], our selection procedure relies on the introduction of auxiliary convolution estimators, $\hat{\mu}_{n, (\mathbf{h}, \boldsymbol{\eta}), p}$ for $(\mathbf{h}, \boldsymbol{\eta}) \in (\mathcal{H}_n^p)^2$ which is the same estimator as the one introduced in Section 3 but with kernel $\mathbf{K}_{\mathbf{h}} * \mathbf{K}_{\boldsymbol{\eta}}$. The first important remark is that when the regularity of the function of interest is unknown, the upper bound on the variance for p fixed can be rewritten for any $\mathbf{h} \in \mathcal{H}_n^p$,

$$v_n^p(\mathbf{h}) = \frac{1}{T_n} \left(p \Delta_n \prod_{i=1}^d h_i^{-1} + \min \left(\sum_{i=1}^d |\log(h_i)| \prod_{i=3}^d h_i^{-1}, (h_2 h_3)^{-1/2} \prod_{i=4}^d h_i^{-1}, \min_{k_0 \geq 3} \prod_{i=1}^{k_0} h_i^{\frac{2-k_0}{k_0}} \prod_{i=k_0+1}^d h_i^{-1} \right) \right).$$

Moreover, one can write:

$$\begin{aligned} \sum_{i=1}^d |\log(h_i)| \prod_{i=3}^d h_i^{-1} &\leq \prod_{i=1}^3 h_i^{-\frac{1}{3}} \prod_{i=4}^d h_i^{-1} \\ \iff h_3^2 &\geq h_1 h_2 \sum_{i=1}^d |\log(h_i)|. \end{aligned}$$

Using the fact that $h_1 \leq \dots \leq h_d$, we obtain

$$v_n^p(\mathbf{h}) = \frac{1}{T_n} \left(p\Delta_n \prod_{i=1}^d h_i^{-1} + \min \left(\sum_{i=1}^d |\log(h_i)| \prod_{i=3}^d h_i^{-1}, (h_2 h_3)^{-1/2} \prod_{i=4}^d h_i^{-1} \right) \right).$$

With this purpose in mind, we introduce the following penalty function

$$V_n^p(\mathbf{h}) = \bar{\omega} \log(n_p) v_n^p(\mathbf{h}),$$

for some positive constant $\bar{\omega}$ which has to be taken large. Now, following once again the procedure of [GL08, GL09, GL11], we define for any $\mathbf{h} \in \mathcal{H}_n^p$,

$$A_n^p(\mathbf{h}) = \max_{\boldsymbol{\eta} \in \mathcal{H}_n^p} \left\{ \left| \hat{\mu}_{n,(\mathbf{h},\boldsymbol{\eta}),p}(x) - \hat{\mu}_{n,\boldsymbol{\eta},p}(x) \right|^2 - V_n^p(\boldsymbol{\eta}) \right\}_+. \quad (4.35)$$

Finally, we define the choice procedure

$$\mathbf{h}^* \in \operatorname{argmin}_{\mathbf{h} \in \mathcal{H}_n^p} \{ A_n^p(\mathbf{h}) + V_n^p(\mathbf{h}) \}.$$

Remark 4.13. *The choice of the penalty $A_n^p(\mathbf{h})$ and the threshold $V_n^p(\mathbf{h})$ are standard in the Goldenshulger-Lepski methodology: $A_n^p(\mathbf{h})$ is a kind of proxy for the estimation of the squared bias of $\hat{\mu}_{n,\mathbf{h},p}(x)$ while $V_n^p(\mathbf{h})$ is the exact penalty needed to balance the size of the variance of the estimator in \mathbf{h} , inflated by a logarithmic term and tuned with $\bar{\omega} > 0$. This enables one to control all the stochastic deviation terms. See section 13 for details.*

Setting this methodology up, we obtain for any pre-averaging level p , the following Oracle inequality

Proposition 4.14 (Oracle inequality). *Assume that Assumptions 7, 8 and 9 hold and that $d \geq 3$, then there exists $n_0 \in \mathbb{N}$, such that for any $n \geq n_0$,*

$$\mathbb{E} \left[\left| \hat{\mu}_{n,\mathbf{h}^*,p} - \bar{\mu}^b(x) \right|^2 \right] \leq c \inf_{\mathbf{h} \in \mathcal{H}_n^p} \left\{ B_{n,\mathbf{h},p}^b(x)^2 + V_n^p(\mathbf{h}) \right\} + c n_p^{-\gamma} + p\Delta_n \mathbb{1}_{p \geq 2} + \frac{\tau_n^{2\alpha_1}}{p_1^\alpha},$$

for some $c > 0$ and $\gamma > 1$, and where we recall that $B_{n,\mathbf{h},p}^b(x)$ is defined in Equation (4.29).

The proof of Proposition 4.14 is postponed to Section 13.

6 Numerical Analysis

In this section, we study the estimators derived in Section 3. We first focus on the case $d = 1$. Following the results of Section 4 bandwidth h is given by $h = T_n^{-1}$. The main challenge is thus to choose the preaveraging parameter p . Theoretically, the best approach would be to consider the bias-corrected estimator $\hat{\mu}_{n,\mathbf{h},p}(x)$ defined in (4.20). However, this estimator is unstable, due to the constant $(\sum_\gamma |u_\gamma|)^{2d}$ appearing in (4.28). Indeed, consider for instance the case

$$V(x) = x^2/4 \quad \text{and} \quad b(x) = -x/2.$$

Taking $n = 2^{14}$, $\Delta_n = n^{-1/2} = 2^{-7}$ and $\tau_n = 1$, we can compute the bias and the variance of both $\hat{\nu}_{n,h,p}(x)$ and $\hat{\mu}_{n,h,p}(x)$. The results are presented in Table 4.1 with $p = \lfloor (\tau_n^2 \Delta_n^{-1})^{1/2} \rfloor \vee 1$. In this example, we can see clearly that the variance increases by a non-negligible factor when doing the bias-correction procedure and consequently the mean squared error is higher, even if the bias is substantially smaller.

	Without bias correction	With bias correction
Error	1.61e-3	8.2e-3
Bias	1.21e-3	3.48e-4
Variance	1.67 e-3	9.77e-3

Table 4.1: Impact of the bias correction.

Therefore, in most practical applications, the use of the initial estimator $\hat{\nu}_{n,h,p}(x)$ leads to better results compared to $\hat{\mu}_{n,h,p}(x)$. Therefore, in the following, we focus on $\hat{\nu}_{n,h,p}(x)$ in our simulations. In that case, we can choose p with easy asymptotics: indeed, the optimal choice for p minimises the bias in (4.17), and thus we take $p^* = \lfloor (\tau_n^2 \Delta_n^{-1})^{1/2} \rfloor \vee 1$.

We now illustrate the effectiveness of the pre-averaging strategy and stack it up against the straightforward method where $p = 1$. We still consider the case

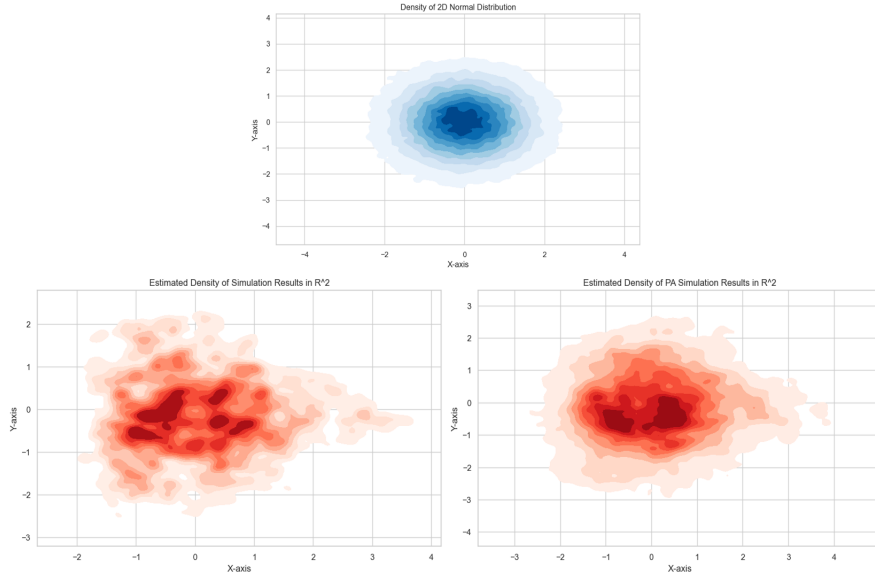
$$V(x) = x^2/4 \quad \text{and} \quad b(x) = -x/2.$$

with $n = 2^{14}$, $\Delta_n = n^{-1/2} = 2^{-7}$ and $\tau_n = 1$. The results are presented in Table 4.2 where the pointwise quadratic error is presented for different values of p and at different points x .

p	$x = 0$	$x = 0.25$	$x = 0.5$	$x = 1$	$x = 0.75$
1	1.29e-1	1.20e-1	9.75e-2	6.80e-2	4.12e-2
16	6.31e-2	5.62e-2	4.36e-2	2.57e-2	1.08e-2
p^*	1.04e-2	1.08e-2	6.49e-3	2.83e-3	1.70e-3
1024	7.07e-2	5.05e-2	4.06e-2	1.67e-2	2.02e-2
4096	7.49e-1	5.39e-1	3.25e-1	7.27e-2	5.22e-2

Table 4.2: Pointwise quadratic error.

We also illustrate the effectiveness of this method in dimension 2, with the potential $V(x) = |x|^2/4$. In that case $\bar{\mu}^b$ is the density of a standard Gaussian distribution. The following plots provides a visual comparison between the desired distribution and the outcomes of two estimation techniques in dimension $d = 2$, with the same parameters as in the unidimensional case. The first plot is the target density. The bottom left plot shows the estimated distribution using no preaveraging method while the bottom right plot shows the results with preaveraging. As expected, the second method exhibits better results, confirming that pre-averaging refines the estimation process when the noise intensity is constant.



Remark 4.15 (Non-Gaussian noise). *In this paper, we assume for simplicity that the noise is composed of independent d -dimensional standard Gaussian variables, also independent of the underlying process X . However, this assumption could be softened. Indeed, we only need that the noise is centered, independent from the diffusion X , identically distributed and that it has finite moment of order l . In the case that its distribution is that of a variable ξ , we should replace estimate (4.16) of Proposition 4.3 by*

$$\left| \mathbb{E}_{\tilde{\mu}^b}[\hat{\nu}_{n,h,p}(x)] - \mathbb{E}[\tilde{\mu}^b(x - p^{-1/2}\tau_n\tilde{\xi}_p - \tilde{\tau}_{n,p}\zeta)] \right| \leq C_B \begin{cases} \sum_i h_i^{\alpha_i} & \text{if } p = 1, \\ \sum_i h_i^{\alpha_i} + \sqrt{p\Delta_n} & \text{if } p \geq 2 \end{cases}$$

where ζ is a standard Gaussian variable, $\tilde{\tau}_{n,p}^2 = (p-1)(2p-1)\Delta_n/12p$ and $\tilde{\xi}_p = p^{-1/2}\sum_{k=1}^p \xi_k$ with ξ_1, \dots, ξ_n i.i.d with common distribution ξ . In that case, the bias correction procedure should be changed accordingly and should be split into two parts. First, we repeat the same procedure with $\tilde{\tau}_{n,p}^2 = (p-1)(2p-1)\Delta_n/12p$ to ensure the first bias corrected estimator centres around $\mathbb{E}[\tilde{\mu}^b(x - p^{-1/2}\tau_n\tilde{\xi}_p)]$. Then we repeat the same procedure with $p^{-1/2}\tau_n$ instead of $\tilde{\tau}_{n,p}^2$ and where m_j stands for the moment of order j of $\tilde{\xi}_p$.

7 Probabilistic tools: the Markovian structure of the preaveraged process

In this section, we gather some estimates about transition probabilities that will be very useful in the following proofs. Indeed, the proof of Proposition 4.5 is based on an in-depth study of the Markov chain

$$\left(p^{-1} \sum_{\ell=0}^{p-1} Y_{kp+\ell,n}, X_{(k+1)p\Delta_n} \right)_{k \in \mathbb{N}}. \quad (4.36)$$

First, note that the stationarity of X ensures that this Markov chain is also stationary under $\mathbb{P}_{\bar{\mu}^b}^b$. We write $\bar{\pi}^b$ for its stationary distribution on \mathbb{R}^{2d} . We now summarize the notations used in the following.

Notation 4.16. For any $t > 0$, $x, y \in \mathbb{R}^d$,

- $p_t^b(x; \cdot)$ stands for the transition density of X , that is density of X_t conditionally on $X_0 = x$;
- $\mathfrak{p}_{p,n,t}^b(x; \cdot)$ stands for the density of $p^{-1} \sum_{\ell=0}^{p-1} X_{\ell\Delta_n+t}$ conditionally on $X_0 = x$;
- $\mathfrak{p}_{p,n}^b(x; \cdot) = \mathfrak{p}_{p,n,0}^b(x; \cdot)$ stands for the density of $p^{-1} \sum_{\ell=0}^{p-1} X_{\ell\Delta_n}$ conditionally on $X_0 = x$;
- $\mathfrak{p}_{p,n}^b(x; \cdot, \cdot)$ stands for the joint density of $(p^{-1} \sum_{\ell=0}^{p-1} X_{\ell\Delta_n}, X_{p\Delta_n})$ conditionally on $X_0 = x$;
- $\bar{\pi}^b(\cdot, \cdot)$ is an invariant density of the Markov process $(p^{-1} \sum_{\ell=0}^{p-1} X_{(kp+\ell)\Delta_n}, X_{(k+1)p\Delta_n})_{k \in \mathbb{N}}$.

We now derive some bounds for these densities. The proofs of these results are relegated to Appendix 14. We start with a short time control of the transition density for the pre-averaged process.

Lemma 4.17. Suppose that Assumptions 7 and 8 hold. Then there exists positive constants κ_1 , λ_1 and η_1 , such that if $p\Delta_n \leq \eta_1$, we have for any $x, y \in \mathbb{R}^d$,

$$\mathfrak{p}_{p,n}^b(x; y) \leq \frac{\kappa_1}{(p\Delta_n)^{d/2}} \exp\left(-\lambda_1 \frac{|y-x|^2}{p\Delta_n} + V(x)\right). \quad (4.37)$$

This lemma is arguably the most technical result in this section. Its proof, postponed to Appendix 14.1, is strongly inspired by Theorem 4 in [GG08] where a similar result is shown for a unidimensional diffusion. As noted in [GG08], their approach cannot readily apply in a multidimensional setting because of a non-trivial time change. In our case the diffusion coefficient is assumed to be a real constant, and the outcomes can be extended. Our motivation for undertaking this proof also lies in the requirement to ensure uniformity of the bounds with respect to p, n . Moreover, here we want to get rid of the uniform boundedness of the drift term b . Its proof is based on the Girsanov theorem to remove the drift contribution. Then we condition at each time i/n so that on each interval $[i/n, (i+1)/n]$, the Brownian Motion is a Brownian bridge. Then sharp estimates can be obtained using the properties of the Brownian Bridge.

We now study the transition density of X in short time.

Lemma 4.18. Under Assumptions 7, 8 and 9, there exists a positive constant κ_3 , such that for any $t \in (0, 1)$, $b \in \Sigma(\mathbf{b})$, and $z \in \mathbb{R}^d$:

$$p_t^b(x; z) \leq \frac{\kappa_3}{t^{d/2}} \exp\left(-\frac{|x-z|^2}{2t} + V(x) - V(z)\right). \quad (4.38)$$

This Lemma is particularly useful when $t \ll 1$. In that case, it is usually stated in the weaker form (4.5). The sharper bound depending explicitly in the potential V allow for better estimates in Proposition 4.5. Combining Lemmas 4.17 and 4.18, we easily get the following corollary.

Corollary 4.19. *Under Assumption 7 and 8, let us assume that $p\Delta_n \leq \eta_1$, where $\eta_1 > 0$ is defined in Lemma 4.17. Then, for each $x, y \in \mathbb{R}^d$, and any $t \in (0, 1)$*

$$\mathfrak{p}_{p,n,t}^b(x, y) \leq \kappa_1 (2\pi)^{-d/2} e^{b_1/2} (p\Delta_n + 2t\lambda_1)^{-d/2} \exp\left(-\frac{\lambda_1}{2t + p\Delta_n} |x - y|^2\right).$$

We finish this section with a result concerning the invariant density of (4.36).

Lemma 4.20. *For all $(x, y) \in \mathbb{R}^{2d}$, there exists a constant κ_2 independent of p, n and b such that*

$$\bar{\pi}^b(y, z) \leq \frac{\kappa_2}{(p\Delta_n)^{d/2}} \exp\left(-V(z) - \frac{\lambda_1}{p\Delta_n} |y - z|^2\right).$$

In the following sections, we assume that Assumptions 7, 8 and 9 hold and that $p\Delta_n \leq 1$. Furthermore $a \lesssim b$ stands for $a \leq Cb$ where C can change from line to line and depending only on \mathfrak{b} .

8 Invertibility of A

We show that the matrix A defined in Section 3 is invertible. To that extent, we compute explicitly $\det(A)$ and we show that it does not vanish. First, we write $a_{k,i,j} = \binom{k}{j} (-1)^j m_j i^{k-j}$ for $0 \leq j \leq k$ so that the (k, i) -th coefficient of A defined in Equation (4.18) is given by $a_{k,i} = \sum_{j=0}^k a_{k,i,j}$. By multilinearity of the determinant, we have then

$$\begin{aligned} \det(A) &= \sum_{j_0=0}^0 \sum_{j_1=0}^1 \cdots \sum_{j_\ell=0}^\ell \det((a_{k,i,j_k})_{k,i}) \\ &= \sum_{j_0=0}^0 \sum_{j_1=0}^1 \cdots \sum_{j_\ell=0}^\ell \det((i^{k-j_k})_{k,i}) \prod_{k=0}^\ell \left(\binom{k}{j_k} (-1)^{j_k} m_{j_k} \right). \end{aligned}$$

Note now that $\det((i^{k-j_k})_{k,i}) = 0$ when two lines are the same. This happens when $k - j_k = k' - j_{k'}$ for some $k \neq k'$. Thus $\det((i^{k-j_k})_{k,i}) = 0$ except when the $k - j_k$, $0 \leq k \leq \ell$ are all distinct. Since $0 \leq j_k \leq k$, we show by induction that the only possibility is $j_0 = j_1 = \cdots = j_\ell = 0$ so that

$$\det(A) = \det((i^k)_{k,i}) \prod_{k=0}^\ell \left(\binom{k}{0} (-1)^0 m_0 \right) = \prod_{0 \leq k < i \leq \ell} (i - k)$$

where the last equality is obtained using the expression of the Vandermonde determinant and using $m_0 = 1$ (by definition).

9 Proof of Proposition 4.3

In this Section, we aim at studying the expectation $\mathbb{E}_{\bar{\mu}^b}^b[\widehat{\nu}_{n,h,p}(x)]$. Let us assume in all the proof that there exists $L > 0$ such that

$$p^2 \Delta_n \leq L, \tag{4.39}$$

for all $n \geq 1$. By definition of $\hat{\nu}_{n,h,p}$ in Equation (4.13) and by stationarity of X under $\mathbb{P}_{\mu^b}^b$, we have

$$\mathbb{E}_{\mu^b}^b[\hat{\nu}_{n,h,p}(x)] = \int_{\mathbb{R}^d} \bar{\mu}^b(y) \mathbb{E}_y^b \left[\mathbf{K}_h \left(x - \frac{\tau_n}{p^{1/2}} \tilde{\xi}_{i,n} - \frac{1}{p} \sum_{\ell=0}^{p-1} X_{\ell\Delta_n} \right) \right] dy.$$

Recall that \mathbb{P}_y^b is the probability measure under which $X_0 = y$ almost surely. When $p = 1$, we observe that $p^{-1} \sum_{\ell=0}^{p-1} X_{\ell\Delta_n} = y$ holds almost surely under \mathbb{P}_y^b . This is not the case when $p > 1$. Instead, we have

$$\frac{1}{p} \sum_{\ell=0}^{p-1} X_{\ell\Delta_n} = y + \frac{1}{p} \sum_{\ell=0}^{p-1} (X_{(kp+\ell)\Delta_n} - X_{kp\Delta_n}).$$

The second part of the right hand side can be seen as a noise term that we decompose as

$$\frac{1}{p} \sum_{\ell=0}^{p-1} (X_{(kp+\ell)\Delta_n} - X_{kp\Delta_n}) = \frac{1}{p} \sum_{\ell=0}^{p-1} \left(\int_{kp\Delta_n}^{(kp+\ell)\Delta_n} b(X_s) ds \right) + \frac{1}{p} \sum_{\ell=0}^{p-1} (W_{(kp+\ell)\Delta_n} - W_{kp\Delta_n}).$$

We remove the effects of the drift b through the Girsanov Theorem. For all admissible drift b , we define

$$N_t^b = \int_0^t b(X_s) \cdot dW_s.$$

The following technical lemma is needed to apply the Girsanov Theorem.

Lemma 4.21. *There exists $\delta_0 > 0$ such that for any $T > 0$ and $x \in \mathbb{R}^d$,*

$$\sup_{t \in [0, T-\delta]} \mathbb{E}_x^b \left[\exp \left(\frac{1}{2} (\langle N^b \rangle_{t+\delta} - \langle N^b \rangle_t) \right) \right] \leq C_1,$$

for every $0 \leq \delta \leq \delta_0$ and some C_1 that depends on b .

Proof. Using Assumption 7, we get that for any $t \in [0, T]$,

$$|X_t| \leq |X_0| + b_0 T + |b|_{lip} \int_0^t |X_s| ds + \sup_{t \in [0, T]} |B_t|.$$

Therefore, from Gronwall Lemma, we have for all $p \geq 1$ and all $0 \leq t \leq T$

$$\mathbb{E}[|X_t|^{2p}] \leq 3^{2p-1} e^{2pt|b|_{lip}} (|x|^{2p} + T^p + (b_0 T)^{2p}). \quad (4.40)$$

Moreover, for any $\delta > 0$ and $t \in [0, T - \delta]$, using Jensen inequality,

$$\mathbb{E}_x^b \left[\exp \left(\frac{1}{2} \int_t^{t+\delta} |b(X_s)|^2 ds \right) \right] \leq \frac{1}{\delta} \int_t^{t+\delta} \mathbb{E}_x^b \left[\exp \left(\frac{\delta}{2} |b(X_s)|^2 \right) \right] ds.$$

Using Assumption 7, we get the following bound

$$\begin{aligned} \mathbb{E}_x^b \left[\exp \left(\frac{1}{2} \int_t^{t+\delta} |b(X_s)|^2 ds \right) \right] &\leq \frac{1}{\delta} \int_t^{t+\delta} \mathbb{E}_x^b \left[\exp \left(\frac{\delta}{2} (|b_0|^2 + |b|_{lip}^2 |X_s|^2) \right) \right] ds \\ &\leq \frac{e^{\delta |b_0|^2}}{\delta} \int_t^{t+\delta} \mathbb{E}_x^b \left[\exp \left(\frac{\delta}{2} |b|_{lip} |X_s|^2 \right) \right] ds. \end{aligned}$$

Now, for $s \in [t, t + \delta]$, we have

$$\mathbb{E}_x^b \left[\exp \left(\frac{\delta}{2} |b|_{lip} |X_s|^2 \right) \right] = 1 + \sum_{k \geq 1} \left(\frac{\delta |b|_{lip}}{2} \right)^k \frac{\mathbb{E}_x^b [|X_s|^{2k}]}{k!}.$$

From Equation (4.40), we get that $\mathbb{E}_x^b \left[\exp \left(\frac{\delta}{2} |b(X_s)|^2 \right) \right] < +\infty$, for all $\delta > 0$, which proves Lemma 4.21. \square

By Novikov's criterion – in its version developed in the classical textbook [KS91], Lemma 5.14, p.198 – Lemma 4.21 shows that the local martingale $\mathcal{E}_t^b(N) = \exp(-N_t^b - 1/2 \langle N^b \rangle_t)$ is indeed a (non-local) martingale under \mathbb{P}_x^b so we can apply Girsanov theorem. Let \mathbb{Q}_x^b be the probability defined by its restriction to \mathcal{F}_t by

$$\frac{d\mathbb{Q}_x^b}{d\mathbb{P}_x^b} \Big|_t = \mathcal{E}_t^b(N).$$

Note that under \mathbb{Q}_x^b , the process $W^{\mathbb{Q}^b}$ defined for $t \geq 0$ by

$$W_t^{\mathbb{Q}^b} = W_t + \int_0^t b(X_s) ds = X_t - X_0 \quad (4.41)$$

is a d -dimensional Brownian motion. Then we have, using the Markov property

$$\begin{aligned} \mathbb{E}_{\bar{\mu}^b}^b [\hat{\nu}_{n,h,p}(x)] &= \int_{\mathbb{R}^d} \bar{\mu}^b(y) \mathbb{E}_y^b [\hat{\nu}_{n,h,p}(x)] dy \\ &= \int_{\mathbb{R}^d} \bar{\mu}^b(y) \mathbb{E}_y^{\mathbb{Q}^b} \left[\mathbf{K}_h \left(x - \frac{\tau_n}{p^{1/2}} \tilde{\xi}_{i,n} - \frac{1}{p} \sum_{\ell=0}^{p-1} X_{\ell \Delta_n} \right) \frac{d\mathbb{P}_y^b}{d\mathbb{Q}_y^b} \Big|_{p\Delta_n} \right] dy. \end{aligned}$$

Moreover, we have $p^{-1} \sum_{\ell=0}^{p-1} X_{\ell \Delta_n} = X_0 + p^{-1} \sum_{\ell=0}^{p-1} W_{\ell \Delta_n}^{\mathbb{Q}^b}$ and thus for all $t \geq 0$

$$\frac{d\mathbb{P}_y^b}{d\mathbb{Q}_y^b} \Big|_t = \exp \left(\int_0^t b(y + W_s^{\mathbb{Q}^b}) \cdot dW_s^{\mathbb{Q}^b} - \frac{1}{2} \int_0^t |b(y + W_s^{\mathbb{Q}^b})|^2 ds \right).$$

and

$$\mathbb{E}_{\bar{\mu}^b}^b [\hat{\nu}_{n,h,p}(x)] = \int_{\mathbb{R}^d} \bar{\mu}^b(y) \mathbb{E}_y^{\mathbb{Q}^b} \left[\mathbf{K}_h \left(x - y - \frac{\tau_n}{p^{1/2}} \tilde{\xi}_{i,n} - \frac{1}{p} \sum_{\ell=0}^{p-1} W_{\ell \Delta_n}^{\mathbb{Q}^b} \right) \frac{d\mathbb{P}_y^b}{d\mathbb{Q}_y^b} \Big|_{p\Delta_n} \right] dy.$$

Note that in the expectation on the right hand side, $\frac{d\mathbb{P}_y^b}{d\mathbb{Q}_y^b} \Big|_{p\Delta_n}$ is entirely determined by $W^{\mathbb{Q}^b}$ which is a

standard Brownian motion under \mathbb{Q}^b . Therefore, we have

$$\begin{aligned} & \mathbb{E}_y^{\mathbb{Q}^b} \left[\mathbf{K}_h \left(x - y - \frac{\tau_n}{p^{1/2}} \tilde{\xi}_{i,n} - \frac{1}{p} \sum_{\ell=0}^{p-1} W_{\ell\Delta_n}^{\mathbb{Q}^b} \right) \frac{d\mathbb{P}_y^b}{d\mathbb{Q}_y^b} \Big|_{p\Delta_n} \right] \\ &= \mathbb{E} \left[\mathbf{K}_h \left(x - y - \frac{\tau_n}{p^{1/2}} \tilde{\xi}_{0,n} - \frac{1}{p} \sum_{\ell=0}^{p-1} W_{\ell\Delta_n} \right) M_{p\Delta_n}^b(y) \right] \end{aligned}$$

where W is a standard Brownian motion, $\tilde{\xi}_{0,n}$ is a standard Gaussian variable and where $M_t^b(y)$ is defined by

$$M_t^b(y) = \exp \left(\int_0^t b(y + W_s) \cdot dW_s - \frac{1}{2} \int_0^t |b(y + W_s)|^2 ds \right).$$

Note that $(M_t^b(y))_t$ can be seen as the solution of the stochastic differential equation

$$M_t^b(y) = 1 + \int_0^t M_s^b(y) b(y + W_s) \cdot dW_s.$$

Using Assumption 8 and Ito's formula, we have

$$M_t^b(y) = \exp \left(V(y) - V(y + W_t) - \frac{1}{2} \int_0^t |b(y + W_s)|^2 + \nabla \cdot b(y + W_s) ds \right). \quad (4.42)$$

This expression ensures that $y \mapsto M_t^b(y)$ is continuous, and therefore $M_t^b(Y)$ is measurable for any random variable Y . Then we have

$$\begin{aligned} \mathbb{E}_{\bar{\mu}^b}^b[\hat{\nu}_{n,h,p}(x)] &= \int_{\mathbb{R}^d} \bar{\mu}^b(y) \mathbb{E} \left[\mathbf{K}_h \left(x - y - \frac{\tau_n}{p^{1/2}} \tilde{\xi}_{0,n} - \frac{1}{p} \sum_{\ell=0}^{p-1} W_{\ell\Delta_n} \right) M_{p\Delta_n}^b(y) \right] dy \\ &= B_1(x) + B_2(x) \end{aligned}$$

where

$$B_1(x) := \int_{\mathbb{R}^d} \bar{\mu}^b(y) \mathbb{E} \left[\mathbf{K}_h \left(x - y - \frac{\tau_n}{p^{1/2}} \tilde{\xi}_{0,n} - \frac{1}{p} \sum_{\ell=0}^{p-1} W_{\ell\Delta_n} \right) \right] dy, \quad (4.43)$$

$$B_2(x) := \int_{\mathbb{R}^d} \bar{\mu}^b(y) \mathbb{E} \left[\mathbf{K}_h \left(x - y - \frac{\tau_n}{p^{1/2}} \tilde{\xi}_{0,n} - \frac{1}{p} \sum_{\ell=0}^{p-1} W_{\ell\Delta_n} \right) (M_{p\Delta_n}^b(y) - 1) \right] dy \quad (4.44)$$

and we study each term separately. More precisely, we will prove the bounds

$$|B_1(x) - \mathbb{E}[\bar{\mu}^b(x - \tilde{\tau}_{n,p}\zeta)]| \lesssim \sum_{i=1}^d h_i^{\alpha_i} \quad \text{and} \quad |B_2| \lesssim \sqrt{p\Delta_n}$$

which prove Proposition 4.3.

Control of $B_1(x)$.

Note that $p^{-1} \sum_{\ell=0}^{p-1} W_{\ell \Delta_n}$ is a centred Gaussian variable with covariance matrix $(p-1)(2p-1)/(12p) \Delta_n I_d$, independent from $\tilde{\xi}_{0,n}$. Therefore, we get:

$$B_1(x) = \int_{\mathbb{R}^d} \bar{\mu}^b(y) \mathbb{E} \left[\mathbf{K}_h \left(x - y - \tilde{\tau}_{n,p} \zeta \right) \right] dy,$$

where $\tilde{\tau}_{n,p}^2 = \tau_n^2/p + (p-1)(2p-1)\Delta_n/12p$ is defined in Equation (4.15), and ζ is a standard Gaussian random variable on \mathbb{R}^d . Moreover, we have

$$\left| B_1(x) - \mathbb{E}[\bar{\mu}^b(x - \tilde{\tau}_{n,p} \zeta)] \right| = \left| \mathbb{E} \left[\int_{\mathbb{R}^d} K(z) [\bar{\mu}^b(x - h \odot z - \tilde{\tau}_{n,p} \zeta) - \bar{\mu}^b(x - \tilde{\tau}_{n,p} \zeta)] dz \right] \right|,$$

where $h \odot z = (h_1 z_1, \dots, h_d z_d)$. We then use a Taylor expansion and the Hölder regularity of the density $\bar{\mu}^b$ and the level of the kernel K , we get

$$\left| \mathbb{E} \left[\int_{\mathbb{R}^d} \mathbf{K}(z) [\bar{\mu}^b(x - h \odot z - \tilde{\tau}_{n,p} \zeta) - \bar{\mu}^b(x - \tilde{\tau}_{n,p} \zeta)] dz \right] \right| \lesssim \sum_{i=1}^d \frac{\int |z_j|^{\alpha_j} |\mathbf{K}(z)| dz}{|\alpha_i|!} h_i^{\alpha_i} \lesssim \sum_{i=1}^d h_i^{\alpha_i}.$$

Control of $B_2(x)$.

First, when $p = 1$, we have

$$\int_{\mathbb{R}^d} \bar{\mu}^b(y) \mathbb{E} \left[\mathbf{K}_h \left(x - y - \frac{\tau_n}{p^{1/2}} \tilde{\xi}_{0,n} \right) (M_{p\Delta_n}^b(y) - 1) \right] dy = 0$$

since $(M_t^b(y))_t$ is a martingale independent of $\tilde{\xi}_{0,n}$. For $p \geq 2$, the situation becomes more intricate. We first state the following lemma.

Lemma 4.22. *Under Assumptions 7, 8 and 9, there exists a constant $M > 0$ which is uniform over $\Sigma(b)$, such that for all $x \in \mathbb{R}^d$*

$$|b(x)e^{-V(x)}| \leq M.$$

Proof. First, note that the function $x \mapsto b(x)e^{-V(x)}$ is continuous and hence bounded on any compact set. Precisely, this function is bounded on the ball centered to 0 and of radius $2\tilde{\rho}_b$, where $\tilde{\rho}_b$ is defined in Assumption 9. We now consider x , such that $|x| \geq 2\tilde{\rho}_b$. With a classical Taylor expansion argument, we get that $V(x) \geq V_0 + \tilde{C}_b|x|/2$.

Moreover, Assumption 7 ensures that

$$|b(x)| \leq C_1(1 + |x|),$$

for some constant $C_1 > 0$ so we get

$$|b(x)e^{-V(x)}| \leq C_1(1 + |x|)e^{-\tilde{C}_b|x|+V_0},$$

which is bounded, concluding the proof. \square

We can now come back to the control of $B_2(x)$, we write $\overline{W}_p^n = p^{-1} \sum_{k=0}^{p-1} W_{k\Delta_n}$ for conciseness. Note that since $\tilde{\xi}_{i,n}$ and \overline{W}_p^n are independent, we have

$$\begin{aligned}
 B_2(x) &= \int_{\mathbb{R}^d} \overline{\mu}^b(y) \mathbb{E} \left[\mathbf{K}_h \left(x - y - \frac{\tau_n}{p^{1/2}} \tilde{\xi}_{0,n} - \overline{W}_p^n \right) (M_{p\Delta_n}^b(y) - 1) \right] dy \\
 &= \mathbb{E} \left[\int_{\mathbb{R}^d} \overline{\mu}^b(y) \mathbf{K}_h \left(x - y - \frac{\tau_n}{p^{1/2}} \tilde{\xi}_{0,n} - \overline{W}_p^n \right) (M_{p\Delta_n}^b(y) - 1) dy \right] \\
 &= \mathbb{E} \left[\int_{\mathbb{R}^d} \mathbf{K}_h \left(x - y - \frac{\tau_n}{p^{1/2}} \tilde{\xi}_{0,n} \right) \overline{\mu}^b \left(y - \overline{W}_p^n \right) (M_{p\Delta_n}^b(y - \overline{W}_p^n) - 1) dy \right] \\
 &= \int_{\mathbb{R}^d} \mathbb{E} \left[\mathbf{K}_h \left(x - y - \frac{\tau_n}{p^{1/2}} \tilde{\xi}_{0,n} \right) \right] \mathbb{E} \left[\overline{\mu}^b \left(y - \overline{W}_p^n \right) (M_{p\Delta_n}^b(y - \overline{W}_p^n) - 1) \right] dy.
 \end{aligned} \tag{4.45}$$

Since

$$\int_{\mathbb{R}^d} |\mathbf{K}_h(x)| dx = \int_{\mathbb{R}^d} |\mathbf{K}(x)| dx,$$

the proof of Proposition 4.3 is down to proving that

$$|\mathbb{E}[\overline{\mu}^b(y - \overline{W}_p^n)(M_{p\Delta_n}^b(y - \overline{W}_p^n) - 1)]| \lesssim \sqrt{p\Delta_n}$$

uniformly for $y \in \mathbb{R}^d$.

Step 1. The main idea of the proof is to perform a Taylor expansion of our quantity of interest around y . To do so, let us introduce for any $y \in \mathbb{R}^d$, $s \geq 0$ and $j \in \{1, \dots, d\}$,

$$Y_{y,s}^j : \mathbb{R} \rightarrow \mathbb{R}^d \\
 u \mapsto (y_1 + W_s^1, \dots, y_{j-1} + W_s^{j-1}, u + W_s^j, y_{j+1} + W_s^{j+1}, \dots, y_d + W_s^d).$$

We also define for all $s \geq 0$, $j \in \{1, \dots, d\}$ and $y \in \mathbb{R}^d$ the function $\phi_{y,s}^j$ for $u \in \mathbb{R}$ by

$$\phi_{y,s}^j(u) = M_s^b(y_1, \dots, y_{j-1}, u, y_{j+1}, \dots, y_d).$$

The purpose of this first step is to show that for any $s \geq 0$, $j \in \{1, \dots, d\}$ and $y \in \mathbb{R}^d$, the function $\phi_{y,s}^j$ is continuously differentiable. Since $\overline{\mu}^b$ is continuously differentiable and for any $z \in \mathbb{R}^d$ $M_t^b(z)$ is positive, it is equivalent to prove that $\log(\phi_{y,s}^j)$ is continuously differentiable. In fact for any $y \in \mathbb{R}^d$, we write

$$\log(M_t^b(y)) = - \sum_{i=1}^d \int_0^t b^i(y + W_s) dW_s^i + \frac{1}{2} \int_0^t |b^i(y + W_s)|^2 ds,$$

where W^i and b^i stand for the i -th component of W and b respectively. Moreover, using the fact that b is continuously differentiable in any direction, we obtain

$$b^i(Y_{y,s}^j(y_j)) = \int_0^{y_j} \partial_j b^i(Y_{y,s}^j(u)) du + b^i(Y_{y,s}^j(0)).$$

Then we can rewrite the previous equation

$$\begin{aligned} \log(M_t^b(y)) &= - \sum_{i=1}^d \left\{ \int_0^t \left(\int_0^{y_j} \partial_j b^i(Y_{y,s}^j(u)) du + b^i(Y_{y,s}^j(0)) \right) dW_s^i \right\} \\ &\quad - \frac{1}{2} \sum_{i=1}^d \left\{ \int_0^t \left| \int_0^{y_j} \partial_j b^i(Y_{y,s}^j(u)) du + b^i(Y_{y,s}^j(0)) \right|^2 ds \right\}. \end{aligned}$$

It is clear that the second double integral is continuously differentiable with respect to y_j . The situation is less straightforward for the first one due to the stochastic integral. We plan to use the following Fubini's Theorem for stochastic integrals.

Theorem 4.23 (Theorem 2.2 in [Ver12]). *Let (X, Σ, μ) be a σ -finite measure space. Let $S = M + A$ be a continuous semimartingale. Let $\psi : X \times [0, T] \rightarrow \mathbb{R}$ be progressively measurable and such that almost surely, we have,*

$$\begin{aligned} \int_X \left(\int_0^T |\psi(x, t)|^2 d\langle M \rangle_t \right)^{\frac{1}{2}} d\mu(x) &< \infty, \\ \int_X \int_0^T |\psi(x, t)| dA_t d\mu(x) &< \infty. \end{aligned}$$

Then, for all $t \in [0, T]$, one has that almost surely

$$\int_X \int_0^t \psi(r, x) dS_r d\mu(x) = \int_0^t \int_X \psi(r, x) d\mu(x) dS_r.$$

Using the boundedness of the partial derivatives, we can easily see that this result applies in our case. Therefore, we get

$$\begin{aligned} \log(M_t^b(y)) &= - \sum_{i=1}^d \left\{ \int_0^{y_j} \left(\int_0^t \partial_j b^i(Y_{y,s}^j(u)) dW_s^i \right) du + \int_0^t b^i(Y_{y,s}^j(0)) dW_s^i \right\} \\ &\quad - \frac{1}{2} \sum_{i=1}^d \left\{ \int_0^t \left| \int_0^{y_j} \partial_j b^i(Y_{y,s}^j(u)) du + b^i(Y_{y,s}^j(0)) \right|^2 ds \right\}. \end{aligned}$$

Note also that $Y_{y,s}^j(u)$ and $Y_{y,s}^j(0)$ do not depend on y_j by definition. Therefore, for all $1 \leq j \leq d$, and for all $y \in \mathbb{R}^d$, the function $\phi_{y,s}^j$ is differentiable. Moreover, we have

$$\partial_j \log(M_t^b(y)) = - \sum_{i=1}^d \left\{ \int_0^t \partial_j b^i(y + W_s) dW_s + \int_0^t (b^i \partial_j b^i)(y + W_s) ds \right\}, \quad (4.46)$$

as $y + W_s = Y_{y,s}^j(y_j)$. We now want to prove that these partial derivatives are continuous. To do so, we plan to apply component by component the following Kolmogorov–Chentsov theorem, due to [AL14]

Theorem 4.24 (Theorem 3.1. of [AL14]). *Let $D \subset \mathbb{R}^d$ be a bounded domain of cone type and let $X : \Omega \times D \rightarrow \mathbb{R}$ be a random field on D . Assume that there exist $m \geq 1$, $p > 1$, $\epsilon \in (0, p]$, and $C > 0$ such*

that the weak derivatives $\partial^\beta X$ are in $L^p(\Omega \times D)$ and

$$\mathbb{E} \left(\left| \partial^\beta X(x) - \partial^\beta X(y) \right|^p \right) \leq C|x - y|^{d+\epsilon}$$

for all $x, y \in D$ and any multi-index $\beta \in \mathbb{N}^n$ with $|\beta| \leq d$. Then X has a modification that is locally of class \bar{C}^t for all $t < d + \min\{\epsilon/p, 1 - d/p\}$.

For all $u, v \in \mathbb{R}$, $p > 1$, and $t \geq 0$, we obtain

$$\begin{aligned} \left| \frac{d}{du} \log(\phi_{y,t}(u)) - \frac{d}{du} \log(\phi_{y,t}(v)) \right|^p &\lesssim \left| \int_0^t (\partial_j b(Y_{y,s}^j(u)) - \partial_j b(Y_{y,s}^j(v))) dW_s \right|^p \\ &+ \left| \int_0^t (b(Y_{y,s}^j(u)) - b(Y_{y,s}^j(v))) \partial_j b(Y_{y,s}^j(u)) ds \right|^p \\ &+ \left| \int_0^t (\partial_j b(Y_{y,s}^j(u)) - \partial_j b(Y_{y,s}^j(v))) b(Y_{y,s}^j(v)) ds \right|^p. \end{aligned}$$

Now, taking the expectation, we get

$$\mathbb{E} \left[\left| \frac{d}{du} \log(\phi_{y,s}^j(u)) - \frac{d}{du} \log(\phi_{y,s}^j(v)) \right|^p \right] \lesssim \mathbf{I} + \mathbf{II} + \mathbf{III},$$

where

$$\begin{aligned} \mathbf{I} &= \mathbb{E} \left[\left| \int_0^t (\partial_j b(Y_{y,s}^j(u)) - \partial_j b(Y_{y,s}^j(v))) dW_s \right|^p \right]; \\ \mathbf{II} &= \mathbb{E} \left[\left| \int_0^t (b(Y_{y,s}^j(u)) - b(Y_{y,s}^j(v))) \partial_j b(Y_{y,s}^j(u)) ds \right|^p \right]; \\ \mathbf{III} &= \mathbb{E} \left[\left| \int_0^t (\partial_j b(Y_{y,s}^j(u)) - \partial_j b(Y_{y,s}^j(v))) b(Y_{y,s}^j(v)) ds \right|^p \right]. \end{aligned}$$

We first control the term \mathbf{I} using the Burkholder-Davis-Gundy's inequality and we get

$$\mathbf{I} \leq \left| \int_0^t \mathbb{E} [|\partial_j b(Y_{y,s}^j(u)) - \partial_j b(Y_{y,s}^j(v))|^2] ds \right|^{p/2}.$$

Moreover, we know that $\partial_j b$ is $(\alpha_j - 1) \wedge 1$ -Hölder in the j -th variable. Using that $\alpha_j \geq 2$, we know that $\partial_j b$ is Lipschitz continuous. Then,

$$\mathbf{I} \lesssim t^{p/2} |u - v|^p.$$

For the second term, we use the boundedness of the partial derivatives to get

$$\mathbf{II} \lesssim t^p |u - v|^p.$$

Moreover, for the last term, using the fact that b is at most linear and the control of moments of order

p of the Brownian motion, we get

$$\text{III} \lesssim |u - v|^p,$$

where the constant here may depend on $(y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_d)$ and t . Finally, as $\alpha_j \geq 2$, we have that $p \geq 1/\varepsilon$, for some $\varepsilon \in (0, p]$. Applying Theorem 4.24, we get that $\phi_{y,t}^j$ is continuously differentiable on \mathbb{R} .

Step 2. For any $t \geq 0$, we define $\psi : \mathbb{R}^d \ni y \mapsto \bar{\mu}^b(y)(M_t^b(y) - 1)$. Moreover, recall that we write $\bar{W}_p^n = p^{-1} \sum_{\ell=0}^{p-1} W_{\ell\Delta_n}$. From the previous step, we write for any $y \in \mathbb{R}^d$,

$$\psi(y - \bar{W}_p^n) - \psi(y) = \sum_{j=1}^d E_j,$$

where

$$\begin{aligned} E_j &= \psi((y - \bar{W}_p^n)_1, \dots, (y - \bar{W}_p^n)_j, y_{j+1}, \dots, y_d) \\ &\quad - \psi((y - \bar{W}_p^n)_1, \dots, (y - \bar{W}_p^n)_{j-1}, y_j, \dots, y_d), \end{aligned}$$

for any $j \in \{2, \dots, d-1\}$, and

$$\begin{aligned} E_1 &= \psi((y - \bar{W}_p^n)_1, y_2, \dots, y_d) - \psi(y_1, \dots, y_d), \\ E_d &= \psi((y - \bar{W}_p^n)_1, \dots, (y - \bar{W}_p^n)_d) - \psi((y - \bar{W}_p^n)_1, \dots, (y - \bar{W}_p^n)_{d-1}, y_d). \end{aligned}$$

Then, using Step 1, we have for any $j \in \{1, \dots, d\}$

$$E_j = \int_0^1 (\partial_j \psi)((y - \bar{W}_p^n)_1, \dots, (y - \bar{W}_p^n)_{j-1}, (y - \lambda \bar{W}_p^n)_j, y_{j+1}, \dots, y_d) (\bar{W}_p^n)_j \, d\lambda.$$

Then, we have

$$\begin{aligned} & (p\Delta_n)^{-1/2} |\mathbb{E}[\psi(y - \bar{W}_p^n) - \psi(y)]| \\ & \leq \sum_{j=1}^d \mathbb{E} \left[\left| \int_0^1 \partial_j \psi((y - \bar{W}_p^n)_1, \dots, (y - \bar{W}_p^n)_{j-1}, (y - \lambda \bar{W}_p^n)_j, y_{j+1}, \dots, y_d) (\bar{W}_p^n)_j \, d\lambda \right|^2 \right]^{1/2}. \end{aligned}$$

We now introduce the following notation for any $\lambda \in (0, 1)$,

$$Y_{y,n,p}^j(\lambda) = ((y - \bar{W}_p^n)_1, \dots, (y - \bar{W}_p^n)_{j-1}, (y - \lambda \bar{W}_p^n)_j, y_{j+1}, \dots, y_d)$$

Moreover, we have the following decomposition

$$\mathbb{E} \left[\left| \int_0^1 \partial_j \psi(Y_{y,n,p}^j(\lambda)) \, d\lambda \right|^2 \right] \lesssim \text{I} + \text{II},$$

where

$$\begin{aligned} \mathbf{I} &= \int_0^1 \mathbb{E} \left[|\partial_j \bar{\mu}^b(Y_{y,n,p}^j(\lambda)) (M_{p\Delta_n}^b(Y_{y,n,p}^j(\lambda)) - 1)|^2 \right] d\lambda \\ \mathbf{II} &= \int_0^1 \mathbb{E} \left[|\mu^b(Y_{y,n,p}^j(\lambda)) \partial_j (M_{p\Delta_n}^b(Y_{y,n,p}^j(\lambda)) - 1)|^2 \right] d\lambda. \end{aligned}$$

In order to control the term \mathbf{I} , we write

$$\mathbf{I} = Z_V^{-1} \int_0^1 \mathbb{E} \left[\left| b^j(Y_{y,n,p}^j(\lambda)) \exp(-2Y_{y,n,p}^j(\lambda)) \left(M_{p\Delta_n}^b(Y_{y,n,p}^j(\lambda)) - 1 \right) \right|^2 \right] d\lambda.$$

Moreover, using Lemma 4.22, we get that

$$\mathbf{I} \lesssim 1 + \int_0^1 \mathbb{E} \left[|M_{p\Delta_n}^b(Y_{y,n,p}^j(\lambda))|^2 \right] d\lambda.$$

Finally, using Equation (4.42) we obtain for any $\lambda \in (0, 1)$,

$$\begin{aligned} M_{p\Delta_n}^b(Y_{y,n,p}^j(\lambda)) &= \exp \left(V(Y_{y,n,p}^j(\lambda)) - V(Y_{y,n,p}^j(\lambda) + W_t) \right. \\ &\quad \left. - \frac{1}{2} \int_0^t |b(Y_{y,n,p}^j(\lambda) + W_s)|^2 + \nabla \cdot b(Y_{y,n,p}^j(\lambda) + W_s) ds \right). \end{aligned}$$

Finally, leveraging on the fact that V is bounded from below, and both $\bar{\mu}^b$ and $\nabla \cdot b$ are bounded, we get

$$\mathbf{I} \lesssim 1.$$

Step 3. We now move on to the control of term \mathbf{II} . First, for any $j \in \{1, \dots, d\}$ we have

$$\frac{\partial_j M_{p\Delta_n}^b(Y_{y,n,p}^j(\lambda))}{M_{p\Delta_n}^b(Y_{y,n,p}^j(\lambda))} = - \sum_{i=1}^d \left(\int_0^{p\Delta_n} \partial_j b^i(Y_{y,n,p}^j(\lambda) + W_s) dW_s^i + \int_0^{p\Delta_n} b^i \partial_j b^i(Y_{y,n,p}^j(\lambda) + W_s) ds \right)$$

Then we have $\mathbf{II} \leq \mathbf{II}_A + \mathbf{II}_B$, where

$$\begin{aligned} \mathbf{II}_A &= 2 \int_0^1 \mathbb{E} \left[|\mu^b(Y_{y,n,p}^j(\lambda)) M_{p\Delta_n}^b(Y_{y,n,p}^j(\lambda)) \sum_{i=1}^d \int_0^{p\Delta_n} \partial_j b^i(Y_{y,n,p}^j(\lambda) + W_s) dW_s^i|^2 \right] d\lambda \\ \mathbf{II}_B &= 2 \int_0^1 \mathbb{E} \left[|\bar{\mu}^b(Y_{y,n,p}^j(\lambda)) M_{p\Delta_n}^b(Y_{y,n,p}^j(\lambda)) \sum_{i=1}^d \int_0^{p\Delta_n} (b^i \partial_j b^i)(Y_{y,n,p}^j(\lambda) + W_s)|^2 \right] d\lambda \end{aligned}$$

Moreover using consecutively Jensen and Cauchy-Schwarz inequalities, we obtain

$$\begin{aligned} \mathbf{II}_A &\lesssim \sum_{i=1}^d \int_0^1 \mathbb{E} \left[|\bar{\mu}^b(Y_{y,n,p}^j(\lambda)) M_{p\Delta_n}^b(Y_{y,n,p}^j(\lambda)) \int_0^{p\Delta_n} \partial_j b^i(Y_{y,n,p}^j(\lambda) + W_u) dW_u^i|^2 \right] d\lambda \\ &\lesssim \sum_{i=1}^d \int_0^1 \mathbb{E} \left[|\bar{\mu}^b(Y_{y,n,p}^j(\lambda)) M_{p\Delta_n}^b(Y_{y,n,p}^j(\lambda))|^4 \right]^{1/2} \mathbb{E} \left[\left| \int_0^{p\Delta_n} \partial_j b^i(Y_{y,n,p}^j(\lambda) + W_u) dW_u^i \right|^4 \right]^{1/2} d\lambda. \end{aligned}$$

Moreover, from Equation (4.42) and the boundedness of $\bar{\mu}^b$, we get once again

$$\int_0^1 \mathbb{E} \left[\left| \bar{\mu}^b(Y_{y,n,p}^j(\lambda)) M_{p\Delta_n}^b(Y_{y,n,p}^j(\lambda)) \right|^4 \right]^{1/2} \lesssim 1.$$

Then, we only need to deal for each $i \in \{1, \dots, d\}$ with

$$\mathbb{E} \left[\left| \int_0^{p\Delta_n} \partial_j b^i(Y_{y,n,p}^j(\lambda) + W_u) dW_u^i \right|^4 \right].$$

Let $\mathcal{G}_{p,\Delta_n} = \sigma(W_{k\Delta_n}, 0 \leq k \leq p)$, the σ -field generated by the discrete observation of W at times $k\Delta_n$ for $k \in \{0, \dots, p\}$, so that conditionally on \mathcal{G}_{p,Δ_n} ,

$$\begin{cases} dW_u = I_k du + dW_u^{*,k}, \\ dW_u^{*,k} = \frac{-W_u^{*,k}}{(k+1)\Delta_n - u} du + dB_u^k, \end{cases} \quad (4.47)$$

in distribution, where for each $k \in \{0, \dots, p-1\}$

$$I_k := \frac{W_{(k+1)\Delta_n} - W_{k\Delta_n}}{\Delta_n},$$

$W^{*,k}$ is a Brownian Bridge on $[k\Delta_n; (k+1)\Delta_n]$, and B^k is a d -dimensional Brownian motion independent of \mathcal{G}_{p,Δ_n} . Then, we write

$$\begin{aligned} & \mathbb{E}^b \left[\left| \int_0^{p\Delta_n} \partial_j b^i(Y_{y,n,p}^j(\lambda) + W_u) dW_u^i \right|^4 \right] = \mathbb{E}^b \left[\mathbb{E}^b \left[\left| \int_0^{p\Delta_n} \partial_j b^i(Y_{y,n,p}^j(\lambda) + W_u) dW_u^i \right|^4 \middle| \mathcal{G}_{p,\Delta} \right] \right] \\ &= \mathbb{E}^b \left[\mathbb{E}^b \left[\left| \sum_{\ell=0}^{p-1} \int_{\ell\Delta_n}^{(\ell+1)\Delta_n} \partial_j b^i(Y_{y,n,p}^j(\lambda) + W_u) dW_u^i \right|^4 \middle| \mathcal{G}_{p,\Delta} \right] \right] \\ &\lesssim \left(\mathbb{E}^b \left[\mathbb{E}^b \left[\left| \sum_{\ell=0}^{p-1} \int_{\ell\Delta_n}^{(\ell+1)\Delta_n} \partial_j b^i(Y_{y,n,p}^j(\lambda) + W_u) I_\ell^i du \right|^4 \middle| \mathcal{G}_{p,\Delta} \right] \right] \right. \\ &\quad \left. + \mathbb{E}^b \left[\mathbb{E}^b \left[\left| \sum_{\ell=0}^{p-1} \int_{\ell\Delta_n}^{(\ell+1)\Delta_n} \partial_j b^i(Y_{y,n,p}^j(\lambda) + W_u) \frac{W_u^{*,\ell,i}}{(\ell+1)\Delta_n - u} du \right|^4 \middle| \mathcal{G}_{p,\Delta} \right] \right] \right. \\ &\quad \left. + \mathbb{E}^b \left[\mathbb{E}^b \left[\left| \sum_{\ell=0}^{p-1} \int_{\ell\Delta_n}^{(\ell+1)\Delta_n} \partial_j b^i(Y_{y,n,p}^j(\lambda) + W_u) dB_u^{\ell,i} \right|^4 \middle| \mathcal{G}_{p,\Delta} \right] \right] \right). \end{aligned}$$

Using both boundedness of the partial derivatives and Jensen inequality, we get

$$\mathbb{E}^b \left[\mathbb{E}^b \left[\left| \sum_{\ell=0}^{p-1} \int_{\ell\Delta_n}^{(\ell+1)\Delta_n} \partial_j b^i(Y_{y,n,p}^j(\lambda) + W_u) I_\ell^i du \right|^4 \middle| \mathcal{G}_{p,\Delta} \right] \right] \lesssim p^3 \Delta_n^4 \mathbb{E} \left[\sum_{\ell=0}^{p-1} I_\ell^4 \right] \lesssim p^4 \Delta_n^2.$$

For the second term, let us remark that by definition of the Brownian Bridge, for any $i \in \{1, \dots, d\}$, $\ell \in \{0, \dots, p-1\}$ and $u \in [\ell\Delta_n, (\ell+1)\Delta_n]$, we have $W_u^{*,\ell,i}$ is gaussian variable with $\mathbb{E}[W_u^{*,\ell,i}] = 0$

and $\text{Var}[W_u^{*,\ell,i}] = u$, as $W_{\ell\Delta_n}^{*,\ell,i} = W_{(\ell+1)\Delta_n}^{*,\ell,i} = 0$. This allows us to obtain

$$\mathbb{E}[|W_u^{*,\ell,i}|^4] = \frac{((\ell+1)\Delta_n - u)^4(u - \ell\Delta_n)^4}{\Delta_n^4}.$$

Using once again the boundedness of $\bar{\mu}^b$, we obtain

$$\mathbb{E}^b \left[\mathbb{E}^b \left[\left| \sum_{\ell=0}^{p-1} \int_{\ell\Delta_n}^{(\ell+1)\Delta_n} \partial_j b^i(Y_{y,n,p}^j(\lambda) + W_u) \frac{W_u^{*,\ell,i}}{(\ell+1)\Delta_n - u} du \right|^4 \middle| \mathcal{G}_{p,\Delta} \right] \right] \lesssim p^4 \Delta_n^2.$$

Lastly, we use the Burkholder-Davis-Gundy Inequality to handle the final term. Every previous step has paved the way for this one. Conditionally to \mathcal{G}_{p,Δ_n} , we have that for each $i \in \{1, \dots, d\}$

$$\left(\int_0^t \partial_j b^i(Y_{y,n,p}^j(\lambda) + W_u) dB_u^{\ell,i} \right)_{t \geq 0},$$

is a martingale. This would not have been true without the conditioning. Then, we get

$$\begin{aligned} \mathbb{E}^b \left[\mathbb{E}^b \left[\left| \sum_{\ell=0}^{p-1} \int_{\ell\Delta_n}^{(\ell+1)\Delta_n} \partial_j b^i(Y_{y,n,p}^j(\lambda) + W_u) dB_u^{\ell,i} \right|^4 \middle| \mathcal{G}_{p,\Delta} \right] \right] &\lesssim \left(\int_0^{p\Delta_n} \partial_j b^i(Y_{y,n,p}^j(\lambda) + W_u)^2 du \right)^2 \\ &\lesssim p^2 \Delta_n^2. \end{aligned}$$

We obtain

$$\mathbb{E}^b \left[\left| \int_0^{p\Delta_n} \partial_j b^i(Y_{y,n,p}^j(\lambda) + W_u) dW_u \right|^4 \right] \lesssim p^4 \Delta_n^2.$$

Combining this to (4.47) gives $\Pi_A \lesssim p^2 \Delta_n$ which is bounded, thanks to Equation (4.39).

Step 4. Control of Π_B . Using Cauchy-Schwarz inequality, we get

$$\Pi_B \lesssim \sum_{i=1}^d \int_0^1 \mathbb{E}^b \left[\left| \int_0^{p\Delta_n} \sqrt{\bar{\mu}^b(Y_{y,n,p}^j(\lambda))} (b^i \partial_j b^i)(Y_{y,n,p}^j(\lambda) + W_u) du \right|^4 \right]^{1/2} d\lambda. \quad (4.48)$$

Moreover for $i \in \{1, \dots, d\}$ and $\lambda \in (0, 1)$,

$$\begin{aligned} &\mathbb{E} \left[\left| \int_0^{p\Delta_n} \sqrt{\bar{\mu}^b(Y_{y,n,p}^j(\lambda))} (b^i \partial_j b^i)(Y_{y,n,p}^j(\lambda) + W_u) du \right|^4 \right] \\ &\lesssim (p\Delta_n)^3 \int_0^{p\Delta_n} \mathbb{E} [\bar{\mu}^b(Y_{y,n,p}^j(\lambda))^2 (b^i \partial_j b^i)(Y_{y,n,p}^j(\lambda) + W_u)^4] du \\ &\lesssim (p\Delta_n)^3 \int_0^{p\Delta_n} \mathbb{E} [\bar{\mu}^b(Y_{y,n,p}^j(\lambda))^2 b^i(Y_{y,n,p}^j(\lambda) + W_u)^4] du \\ &\lesssim (p\Delta_n)^4, \end{aligned}$$

where, at the last line we used a similar argument as the one of the proof of Lemma 4.22 to bound $\mathbb{E}^b [\bar{\mu}^b(Y_{y,n,p}^j(\lambda))^2 b^i(Y_{y,n,p}^j(\lambda) + W_u)^4]$ uniformly in n, p, y, u, i and λ . Therefore $\Pi_B \lesssim (p\Delta_n)^2$, and

$\Pi \lesssim 1$.

Conclusion. Combining the previous four steps, we have

$$|\mathbb{E}[\bar{\mu}^b(y - \bar{W}_p^n)(M_{p\Delta_n}^b(y - \bar{W}_p^n) - 1)]| \lesssim \sqrt{p\Delta_n}$$

and therefore, for all $x \in \mathbb{R}^d$,

$$B_2(x) \lesssim \sqrt{p\Delta_n}, \quad (4.49)$$

which concludes the proof.

10 Proof of Proposition 4.4

We now plan to use Proposition 4.3 to prove the estimate (4.21). Recall that for all $x \in \mathbb{R}^d$, $\hat{\mu}_{n,h,p}(x)$ is defined in Equation (4.20) by

$$\hat{\mu}_{n,h,p}(x) = \sum \mathbf{u}_\gamma \hat{\nu}_{n,h,p}(x + \gamma \tilde{\tau}_{n,p})$$

where the sum holds over all $\gamma = (\gamma_1, \dots, \gamma_d) \in \{0, \dots, \ell\}^d$. Thus Proposition 4.3 ensures that (4.21) is proved once we control the following quantity

$$\left| \sum \mathbf{u}_\gamma \mathbb{E}[\bar{\mu}^b(x + \tilde{\tau}_{n,p}(\gamma - \zeta))] - \bar{\mu}^b(x) \right| \quad (4.50)$$

We first provide a technical Lemma in dimension $d = 1$.

Lemma 4.25. *Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a α -Hölder function and ζ be a standard real Gaussian variable. Then we have*

$$\left| \sum_{\gamma=0}^l u_\gamma \mathbb{E}[f(x + \tau(\gamma - \zeta))] - f(x) \right| \leq \tau^\alpha \frac{\mathcal{L}}{[\alpha]} \sum_{\gamma=0}^l \mathbb{E}[|\gamma - \zeta|^\alpha].$$

Proof. By Taylor formula and using the α -Hölder property of f , we get the existence of $\varepsilon \in [0, 1]$ such that

$$\begin{aligned} \sum_{\gamma=0}^l u_\gamma \mathbb{E}[f(x + \tau(\gamma - \zeta))] &= \sum_{\gamma=0}^l u_\gamma \sum_{k=0}^{[\alpha]-1} \frac{f^{(k)}(x)}{k!} \mathbb{E}[(\gamma - \zeta)^k] + R_{\alpha,l}(\varepsilon) \\ &= \sum_{k=0}^{[\alpha]-1} \frac{f^{(k)}(x)}{k!} \sum_{\gamma=0}^l u_\gamma \sum_{\beta=0}^k \frac{k!}{\beta!(k-\beta)!} (-1)^\beta m_\beta \gamma^{k-\beta} + R_{\alpha,l}(\varepsilon), \end{aligned}$$

where

$$R_{\alpha,l}(\varepsilon) = \mathbb{E} \left[\sum_{\gamma=0}^l \frac{f^{([\alpha])}(x + \varepsilon\tau(\gamma - \zeta))}{([\alpha])!} (\tau\varepsilon)^{[\alpha]} (\gamma - \xi)^{[\alpha]} \right].$$

Then, from Equation (4.19), we get that $\sum_{\gamma=0}^l u_\gamma \sum_{k=0}^{[\alpha]-1} f^{(k)}(x) \mathbb{E}[(\gamma - \zeta)^k]/k! = f(x)$. Moreover as f is α -Hölder, we conclude that

$$R_{\alpha,l}(\varepsilon) \leq \tau^\alpha \frac{\mathcal{L}}{[\alpha]} \sum_{\gamma=0}^l \mathbb{E}[|\gamma - \zeta|^\alpha].$$

□

We now extend Lemma 4.25 to our setup using an appropriate induction. We first introduce $\theta = \theta(x, \gamma, n, p) = x + \gamma \tilde{\tau}_{n,p} \in \mathbb{R}^d$. We prove by induction on $0 \leq I \leq d$ that for any $0 \leq \gamma_{I+1}, \dots, \gamma_d \leq l$,

$$\left| \sum_{\gamma_{I+1}=0}^l \dots \sum_{\gamma_1=0}^l \prod_{i=1}^I u_{\gamma_i} \mathbb{E}[\bar{\mu}^b(\theta - \tilde{\tau}_{n,p}\zeta)] - \mathbb{E}[\bar{\mu}^b(x_1, \dots, x_I, \theta_{I+1} - \tilde{\tau}_{n,p}\zeta_{I+1}, \dots, \theta_d - \tilde{\tau}_{n,p}\zeta_d)] \right|,$$

is bounded up to some constant by $\sum_{i=1}^I \tilde{\tau}_{n,p}^{\alpha_i}$. For $I = 0$, this is trivial and for $I = d$, the second expectation does not contain any random variable any-more and thus equals $\bar{\mu}^b(x)$, which proves Equation (4.50). Let $I < d$ at which this property is proved. We fix $0 \leq \gamma_{I+2}, \dots, \gamma_d \leq l$. We have

$$\sum_{\gamma_{I+1}=0}^l \dots \sum_{\gamma_1=0}^l \prod_{i=1}^{I+1} u_{\gamma_i} \mathbb{E}[\bar{\mu}^b(\theta - \tilde{\tau}_{n,p}\zeta)] = \sum_{\gamma_{I+1}=0}^l u_{\gamma_{I+1}} \sum_{\gamma_I=0}^l \dots \sum_{\gamma_1=0}^l \prod_{i=1}^I u_{\gamma_i} \mathbb{E}[\bar{\mu}^b(\theta - \tilde{\tau}_{n,p}\zeta)]$$

Recall also that by Assumption, $\bar{\mu}^b$ is α_{I+1} -Hölder regular in the $(I+1)$ -th variable. Therefore, by induction and using Lemma 4.25 we get

$$\left| \sum_{\gamma_{I+1}=0}^l u_{\gamma_{I+1}} \mathbb{E}[\bar{\mu}^b(x_1, \dots, x_I, \theta_{I+1} - \tilde{\tau}_{n,p}\zeta_{I+1}, \dots, \theta_d - \tilde{\tau}_{n,p}\zeta_d) | \zeta_{I+2}, \dots, \zeta_d] - \bar{\mu}^b(x_1, \dots, x_{I+1}, \theta_{I+2} - \tilde{\tau}_{n,p}\zeta_{I+2}, \dots, \theta_d - \tilde{\tau}_{n,p}\zeta_d) \right| \lesssim \tilde{\tau}_{n,p}^{\alpha_{I+1}}.$$

Note that Lemma 4.25 ensures that the last inequality holds up to a constant uniformly in n, p, b .

11 Proof of Proposition 4.5

11.1 Structure and completion of the proof

First note that by (4.28), it enough to prove Proposition 4.5 with $\hat{\nu}_{n,\mathbf{h},p}(x)$ instead of $\hat{\mu}_{n,\mathbf{h},p}(x)$. We now introduce the main notations used in this proof. Let $n_p = \lfloor n/p \rfloor$. For each j , we write

$$\tilde{\xi}_j = \frac{1}{\sqrt{p}} \sum_{\ell=0}^{p-1} \xi_{j+\ell}$$

so that $(\tilde{\xi}_j)_j$ is a sequence of i.i.d standard d -dimensional Gaussian variables. Then we write

$$\mathfrak{C}_p^b(k; x) = \text{Cov}_{\bar{\mu}^b} \left(\mathbf{K}_\mathbf{h} \left(x - p^{-1} \sum_{\ell=0}^{p-1} X_{(kp+\ell)\Delta_n} + \frac{\tau_n}{\sqrt{p}} \tilde{\xi}_k \right), \mathbf{K}_\mathbf{h} \left(x - p^{-1} \sum_{\ell=0}^{p-1} X_{\ell\Delta_n} + \frac{\tau_n}{\sqrt{p}} \tilde{\xi}_0 \right) \right). \quad (4.51)$$

for each k and x , and

$$\mathfrak{C}_p^b(k; x, u, v) = \text{Cov}_{\bar{\mu}^b}^b \left(\mathbf{K}_h \left(x - p^{-1} \sum_{\ell=0}^{p-1} X_{(kp+\ell)\Delta_n} + u \right), \mathbf{K}_h \left(x - p^{-1} \sum_{\ell=0}^{p-1} X_{\ell\Delta_n} + v \right) \right) \quad (4.52)$$

$$\mathfrak{D}_p^b(k; x, u, v) = \mathbb{E}_{\bar{\mu}^b}^b \left[\mathbf{K}_h \left(x - \frac{1}{p} \sum_{\ell=0}^{p-1} X_{(kp+\ell)\Delta_n} + u \right) \mathbf{K}_h \left(x - \frac{1}{p} \sum_{\ell=0}^{p-1} X_{\ell\Delta_n} + v \right) \right] \quad (4.53)$$

for each k , x , u and v . Note that for $k > 0$, the noises $\tilde{\xi}_0$ and $\tilde{\xi}_k$ are independent and therefore

$$\mathfrak{C}_p^b(k; x) \leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathfrak{C}_p^b(k; x, u, v) \chi_{p,n}(du) \chi_{p,n}(dv) \quad (4.54)$$

where $\chi_{p,n}$ stands for the law of $\tau_n p^{-1/2} \tilde{\xi}_0$ so that $\chi_{p,n}(du) = \varphi_{\tau_n p^{-1/2}}(u) du$. Moreover, we also have

$$\mathfrak{C}_p^b(k; x) \leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathfrak{D}_p^b(k; x, u, v) \chi_{p,n}(du) \chi_{p,n}(dv) \quad (4.55)$$

By definition of $\hat{\nu}_{n,\mathbf{h},p}(x)$, we have

$$\begin{aligned} \text{Var}_{\bar{\mu}^b}^b(\hat{\nu}_{n,\mathbf{h},p}(x)) &= \text{Var}_{\bar{\mu}^b}^b \left(\frac{1}{n_p} \sum_{k=0}^{n_p-1} \mathbf{K}_h \left(x - \frac{1}{p} \sum_{\ell=0}^{p-1} Y_{kp+\ell,n} \right) \right) \\ &= \frac{1}{n_p^2} \sum_{i,j=0}^{n_p-1} \text{Cov}_{\bar{\mu}^b}^b \left(\mathbf{K}_h \left(x - \frac{1}{p} \sum_{\ell=0}^{p-1} X_{(ip+\ell)\Delta_n} + \frac{\tau_n}{\sqrt{p}} \tilde{\xi}_i \right), \mathbf{K}_h \left(x - \frac{1}{p} \sum_{\ell=0}^{p-1} X_{(jp+\ell)\Delta_n} + \frac{\tau_n}{\sqrt{p}} \tilde{\xi}_j \right) \right). \end{aligned}$$

Since X and ξ are both ergodic and mutually independent, we deduce that

$$\text{Var}_{\bar{\mu}^b}^b \left(\frac{1}{n_p} \sum_{k=0}^{n_p-1} \mathbf{K}_h \left(x - \frac{1}{p} \sum_{\ell=0}^{p-1} Y_{kp+\ell,n} \right) \right) = \frac{2}{n_p^2} \sum_{k=0}^{n_p-1} (n_p - k) \mathfrak{C}_p^b(k; x). \quad (4.56)$$

We now present several bounds on the coefficients $\mathfrak{C}_p^b(k; x)$.

Lemma 4.26 (Instantaneous correlations). *For each $k \geq 0$, we have*

$$\mathfrak{C}_p^b(k; x) \lesssim \prod_{i=1}^d h_i^{-1}.$$

Lemma 4.27 (Short-term correlations). *For any $1 \leq k_1 \leq d$ and any $k \geq 2$, we have*

$$\mathfrak{C}_p^b(k; x) \lesssim (kp\Delta_n)^{-k_1/2} \prod_{i=k_1+1}^d h_i^{-1}.$$

Lemma 4.28 (Mid-term correlations). *For any $k \geq 1$ such that $(k-1)p\Delta_n \geq 1$, we have*

$$\mathfrak{C}_p^b(k; x) \lesssim 1.$$

Lemma 4.29 (Long-term correlations). *For any $k \geq 1$, we have*

$$\mathfrak{E}_p^b(k; x) \lesssim e^{-C_{PI}^{-1}(k-1)p\Delta_n} \left(\prod_{i=1}^d h_i^{-1} \right)^2.$$

Following [AG23], the rest of the proof consists in identifying the best way to combine these lemmas with (4.56), depending on d and k_0 . Recall that k_0 is such that $\alpha_1 = \dots = \alpha_{k_0} < \alpha_{k_0+1} \leq \dots \leq \alpha_d$. We now introduce $0 \leq j_1 \leq j_2 \leq j_3 \leq j_4$ such that $j_3 = \lceil (p\Delta_n)^{-1} \rceil$ and

$$j_4 = \max \left(\min \left(\left\lfloor \frac{-2C_{PI} \log \prod_{i=1}^d h_i}{p\Delta_n} \right\rfloor, n_p \right), j_3 \right).$$

Then we have

$$\frac{1}{n_p^2} \sum_{k=0}^{n_p-1} (n_p - k) \mathfrak{E}_p^b(k; x) = \frac{1}{n_p^2} \left(\sum_{k=0}^{j_1} + \sum_{k=j_1+1}^{j_2} + \sum_{k=j_2+1}^{j_3} + \sum_{k=j_3+1}^{j_4} + \sum_{k=j_4+1}^{n_p-1} \right) (n_p - k) \mathfrak{E}_p^b(k; x).$$

For the first sum, we use Lemma 4.26 so that we have

$$\frac{1}{n_p^2} \sum_{k=0}^{j_1} (n_p - k) \mathfrak{E}_p^b(k; x) \lesssim \frac{1}{n_p \prod_{i=1}^d h_i} (j_1 + 1).$$

For the second and third sums, we use Lemma 4.27 for some $1 \leq k_1 \leq d$ and $1 \leq k_2 \leq d$ to be chosen later. In that case, we get

$$\begin{cases} n_p^{-2} \sum_{k=j_1+1}^{j_2} (n_p - k) \mathfrak{E}_p^b(k; x) \lesssim n_p^{-1} \sum_{k=j_1+1}^{j_2} (kp\Delta_n)^{-k_1/2} \prod_{i=k_1+1}^d h_i^{-1}, \\ n_p^{-2} \sum_{k=j_2+1}^{j_3} (n_p - k) \mathfrak{E}_p^b(k; x) \lesssim n_p^{-1} \sum_{k=j_2+1}^{j_3} (kp\Delta_n)^{-k_2/2} \prod_{i=k_2+1}^d h_i^{-1}. \end{cases} .$$

Since $j_3 k \Delta_n \geq 1$, we use Lemma 4.28 for the fourth sum and we get

$$\frac{1}{n_p^2} \sum_{k=j_3+1}^{j_4} (n_p - k) \mathfrak{E}_p^b(k; x) \lesssim \frac{1}{n_p} (j_4 - j_3).$$

Using the definition of j_4 , we have

$$\frac{1}{n_p^2} \sum_{k=j_3+1}^{j_4} (n_p - k) \mathfrak{E}_p^b(k; x) \lesssim \min \left(\frac{-2C_{PI} \log \prod_{i=1}^d h_i}{n_p p \Delta_n}, 1 \right).$$

Using also that $n_p p \Delta_n \gtrsim T_n$, we get

$$\frac{1}{n_p^2} \sum_{k=j_3+1}^{j_4} (n_p - k) \mathfrak{E}_p^b(k; x) \lesssim \min \left(\frac{\sum_{i=1}^d |\log h_i|}{T_n}, 1 \right). \quad (4.57)$$

For the last sum, we use Lemma 4.29 and we have

$$\begin{aligned} \frac{1}{n_p^2} \sum_{k=j_4+1}^{n_p-1} (n_p - k) \mathfrak{E}_p^b(k; x) &\lesssim \frac{1}{n_p} \sum_{k=j_4+1}^{n_p-1} \mathfrak{E}_p^b(k; x) \\ &\lesssim \frac{1}{n_p} \sum_{k=j_4+1}^{n_p-1} e^{-C_{PI}^{-1}(k-1)p\Delta_n} \left(\prod_{i=1}^d h_i^{-1} \right)^2 \\ &\lesssim \frac{1}{T_n} e^{-C_{PI}^{-1}j_4 p\Delta_n} \left(\prod_{i=1}^d h_i^{-1} \right)^2. \end{aligned}$$

Using the definition of j_4 , we have

$$\frac{1}{n_p^2} \sum_{k=j_4+1}^{n_p-1} (n_p - k) \mathfrak{E}_p^b(k; x) \lesssim \frac{1}{T_n} + \frac{1}{T_n} e^{-C_{PI}^{-1}T_n} \left(\prod_{i=1}^d h_i^{-1} \right)^2. \quad (4.58)$$

Case $d = 1, 2$. Taking $j_1 = 0, j_2 = j_3$ and $k_1 = d$, we obtain

$$\begin{aligned} \frac{2}{n_p^2} \sum_{k=j_1+1}^{j_2} \mathfrak{E}_p^b(k; x) &\lesssim \frac{2}{n_p^2} \sum_{k=1}^{j_2} (n_p - k) \frac{1}{kp\Delta_n} \\ &\lesssim \frac{\log(j_2)}{n_p p \Delta_n} = \frac{|\log(p\Delta_n)|}{T_n}. \end{aligned}$$

and therefore

$$\text{Var}_{\mu^b}(\hat{\nu}_{n, \mathbf{h}, p}(x)) \lesssim \frac{p\Delta_n}{T_n} \prod_{i=1}^d h_i^{-1} + \frac{|\log(p\Delta_n)|}{T_n} + \sum_{i=1}^d \frac{|\log(h_i)|}{T_n} + \frac{1}{T_n} e^{-CT_n} \left(\prod_{i=1}^d h_i^{-1} \right)^2.$$

Case $d \geq 3, k_0 = 1$ and $\alpha_2 < \alpha_3$ or $k_0 = 2$. We take $k_1 = 2$ and $k_2 = d$ and

$$j_1 = \left\lfloor \frac{h_1 h_2}{p\Delta_n} \right\rfloor \quad \text{and} \quad j_2 = \left\lfloor \frac{(\prod_{i \geq 3} h_i)^{\frac{2}{d-2}}}{p\Delta_n} \right\rfloor.$$

Therefore we have

$$\begin{aligned} \frac{1}{n_p^2} \sum_{k=0}^{j_1} (n_p - k) \mathfrak{E}_p^b(k; x) &\lesssim \frac{1}{n_p \prod_{i=1}^d h_i} \frac{h_1 h_2}{p\Delta_n} + \frac{1}{n_p} \prod_{i=1}^d h_i^{-1} \\ &\lesssim \frac{1}{T_n \prod_{i=3}^d h_i} + \frac{1}{n_p} \prod_{i=1}^d h_i^{-1} \end{aligned}$$

and

$$\begin{aligned}
 n_p^{-2} \sum_{k=j_1+1}^{j_2} (n_p - k) \mathfrak{E}_p^b(k; x) &\lesssim n_p^{-1} \sum_{k=j_1+1}^{j_2} (kp\Delta_n)^{-1} \prod_{i=3}^d h_i^{-1} \\
 &\lesssim n_p^{-1} \log(j_2/j_1) (p\Delta_n)^{-1} \prod_{i=3}^d h_i^{-1} \\
 &\lesssim T_n^{-1} \sum_{i=1}^d |\log(h_i)| \prod_{i=3}^d h_i^{-1}
 \end{aligned}$$

and

$$\begin{aligned}
 n_p^{-2} \sum_{k=j_2+1}^{j_3} (n_p - k) \mathfrak{E}_p^b(k; x) &\lesssim n_p^{-1} \sum_{k=j_2+1}^{j_3} (kp\Delta_n)^{-d/2} \\
 &\lesssim n_p^{-1} j_2^{1-d/2} (p\Delta_n)^{-d/2} \\
 &\lesssim T_n^{-1} \left(\prod_{i \geq 3} h_i \right)^{-1}.
 \end{aligned}$$

Combining these three bounds with (4.57) and (4.58), we get

$$\text{Var}_{\bar{\mu}^b}(\hat{\nu}_{n, \mathbf{h}, p}(x)) \lesssim \frac{1}{T_n \prod_{i=3}^d h_i} + \frac{1}{n_p} \prod_{i=1}^d h_i^{-1} + \frac{1}{T_n} \sum_{i=1}^d |\log(h_i)| \prod_{i=3}^d h_i^{-1}$$

which proves (4.25).

Case $d \geq 3, k_0 \geq 3$. We take $k_1 = k_0$ and $k_2 = d$ and

$$j_1 = \left\lfloor \frac{\left(\prod_{i=1}^{k_0} h_i \right)^{2/k_0}}{p\Delta_n} \right\rfloor \quad \text{and} \quad j_2 = \left\lfloor \frac{1}{p\Delta_n} \right\rfloor.$$

Therefore, we have

$$\begin{aligned}
 \frac{1}{n_p^2} \sum_{k=0}^{j_1} (n_p - k) \mathfrak{E}_p^b(k; x) &\lesssim \frac{1}{n_p \prod_{i=1}^d h_i} \frac{\left(\prod_{i=1}^{k_0} h_i \right)^{2/k_0}}{p\Delta_n} + \frac{1}{n_p} \prod_{i=1}^d h_i^{-1} \\
 &\lesssim \frac{1}{T_n} \prod_{i=1}^{k_0} (h_i)^{(2-k_0)/k_0} \prod_{i=k_0+1}^d h_i^{-1} + \frac{1}{n_p} \prod_{i=1}^d h_i^{-1},
 \end{aligned}$$

and

$$\begin{aligned}
 n_p^{-2} \sum_{k=j_1+1}^{j_2} (n_p - k) \mathfrak{C}_p^b(k; x) &\lesssim n_p^{-1} \sum_{k=j_1+1}^{j_2} (kp\Delta_n)^{-k_0/2} \prod_{i=k_0+1}^d h_i^{-1} \\
 &\lesssim n_p^{-1} j_1^{1-k_0/2} (p\Delta_n)^{-k_0/2} \prod_{i=k_0+1}^d h_i^{-1} \\
 &\lesssim T_n^{-1} \prod_{i=1}^{k_0} (h_i)^{(2-k_0)/k_0} \prod_{i=k_0+1}^d h_i^{-1},
 \end{aligned}$$

and using that $j_3 \leq j_2 + 1$

$$n_p^{-2} \sum_{k=j_2+1}^{j_3} (n_p - k) \mathfrak{C}_p^b(k; x) \lesssim n_p^{-1} (j_2 + 1)^{-d/2} (p\Delta_n)^{-d/2} \lesssim \frac{1}{n_p}.$$

Combining these three bounds with (4.57) and (4.58), we get

$$\text{Var}_{\bar{\mu}^b}(\hat{\nu}_{n, \mathbf{h}, p}(x)) \lesssim T_n^{-1} \prod_{i=1}^{k_0} (h_i)^{(2-k_0)/k_0} \prod_{i=k_0+1}^d h_i^{-1} + \frac{1}{n_p} \prod_{i=1}^d h_i^{-1} + \frac{\sum_{i=1}^d |\log h_i|}{T_n}$$

which proves (4.26).

Case $d \geq 3$, $k_0 = 1$ and $\alpha_2 = \alpha_3$ We take $k_1 = 1$ and $k_2 = 3$,

$$j_1 = 0 \quad \text{and} \quad j_2 = \left\lfloor \frac{h_2 h_3}{p\Delta_n} \right\rfloor.$$

Therefore, we have

$$\frac{1}{n_p^2} \sum_{k=0}^{j_1} (n_p - k) \mathfrak{C}_p^b(k; x) \lesssim \frac{1}{n_p \prod_{i=1}^d h_i}$$

and

$$\begin{aligned}
 n_p^{-2} \sum_{k=j_1+1}^{j_2} (n_p - k) \mathfrak{C}_p^b(k; x) &\lesssim n_p^{-1} \sum_{k=j_1+1}^{j_2} (kp\Delta_n)^{-1/2} \prod_{i=2}^d h_i^{-1} \\
 &\lesssim n_p^{-1} j_2^{1/2} (p\Delta_n)^{-1/2} \prod_{i=2}^d h_i^{-1} \\
 &\lesssim T_n^{-1} (h_2 h_3)^{-1/2} \prod_{i=4}^d h_i^{-1}
 \end{aligned}$$

and

$$\begin{aligned}
 n_p^{-2} \sum_{k=j_2+1}^{j_3} (n_p - k) \mathfrak{C}_p^b(k; x) &\lesssim n_p^{-1} \sum_{k=j_2+1}^{j_3} (kp\Delta_n)^{-3/2} \prod_{i=4}^d h_i^{-1} \\
 &\lesssim n_p^{-1} j_2^{-1/2} (p\Delta_n)^{-3/2} \prod_{i=4}^d h_i^{-1} \\
 &\lesssim T_n^{-1} (h_2 h_3)^{-1/2} \prod_{i=4}^d h_i^{-1}
 \end{aligned}$$

Combining these three bounds with (4.57) and (4.58), we get

$$\text{Var}_{\bar{\mu}^b}^b(\hat{\nu}_{n, \mathbf{h}, p}(x)) \lesssim \frac{\sum_{i=1}^d |\log h_i|}{T_n} + \frac{1}{n_p \prod_{i=1}^d h_i} + \frac{1}{T_n \sqrt{h_2 h_3}} \prod_{i=4}^d h_i^{-1}$$

which proves (4.27).

11.2 Proof of Lemma 4.26

Since X and ξ are both ergodic and mutually independent we have by Cauchy-Schwarz inequality

$$\begin{aligned}
 \mathfrak{C}_p^b(k; x) &\leq \text{Var}_{\bar{\mu}^b}^b(\mathbf{K}_{\mathbf{h}}(x - p^{-1} \sum_{\ell=0}^{p-1} X_{(kp+\ell)\Delta_n} + \frac{\tau_n}{\sqrt{p}} \tilde{\xi}_k)) \\
 &\leq \mathbb{E}_{\bar{\mu}^b}^b \left[\mathbf{K}_{\mathbf{h}}(x - p^{-1} \sum_{\ell=0}^{p-1} X_{(kp+\ell)\Delta_n} + \frac{\tau_n}{\sqrt{p}} \tilde{\xi}_k)^2 \right].
 \end{aligned}$$

Using the notations of Section 7, we know that $p^{-1} \sum_{\ell=0}^{p-1} X_{(kp+\ell)\Delta_n} + \tau_n p^{-1/2} \tilde{\xi}_k$ has a density under $\mathbb{P}_{\bar{\mu}^b}^p$ given by

$$\bar{\mu}_{p,n}^b(z) := \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \bar{\mu}^b(y) \mathfrak{p}_{p,n}^b(y; u) \varphi_{\tau_n p^{-1/2}}(z - u) \, dy \, du$$

Moreover, for all $z \in \mathbb{R}^d$,

$$\begin{aligned}
 |\bar{\mu}_{p,n}^b(z)| &\leq \|\bar{\mu}^b\|_{\infty} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \mathfrak{p}_{p,n}^b(y; u) \, dy \right) \varphi_{\tau_n p^{-1/2}}(z - u) \, du \\
 &\leq \|\bar{\mu}^b\|_{\infty}.
 \end{aligned}$$

Then, performing a change of variable gives

$$\mathfrak{C}_p^b(k; x) = \int_{\mathbb{R}^d} \bar{\mu}_{p,n}^b(z) \mathbf{K}_{\mathbf{h}}(x - z)^2 \, dz = \left(\prod_{i=1}^d h_i^{-1} \right) \int_{\mathbb{R}^d} \bar{\mu}_{p,n}^b(x + \mathbf{h}z) \mathbf{K}(z)^2 \, dz \lesssim \prod_{i=1}^d h_i^{-1}$$

since $\mathbf{K} \in L^2$.

11.3 Proof of Lemma 4.27

First recall from Section 7 that the process

$$(p^{-1} \sum_{\ell=0}^{p-1} X_{(kp+\ell)\Delta_n}, X_{(k+1)p\Delta_n})_{k \in \mathbb{N}}$$

is stationary with law denoted $\bar{\pi}^b$ on \mathbb{R}^{2d} . Using also that X is a Markov process, we get that for all $(u, v) \in \mathbb{R}^d \times \mathbb{R}^d$, we have

$$\begin{aligned} \mathfrak{D}_p^b(k; x, u, v) &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{E}^b[\mathbf{K}_h(x - p^{-1} \sum_{\ell=0}^{p-1} X_{(kp+\ell)\Delta_n} + v) | X_{p\Delta_n} = z] \mathbf{K}_h(x - \tilde{\omega} + v) \bar{\pi}^b(\tilde{\omega}; z) dz d\tilde{\omega} \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{K}_h(x + v - \omega) \mathbf{K}_h(x + u - \tilde{\omega}) \mathfrak{p}_{p,n,(k-1)p\Delta_n}^b(z; \omega) \bar{\pi}^b(\tilde{\omega}; z) d\tilde{\omega} d\omega dz. \end{aligned}$$

where $\mathfrak{p}_{p,n,(k-1)p\Delta_n}^b$ is also defined in Section 7. Moreover, we know from Lemma 4.20 that there exists $\lambda_1 > 0$ such that

$$\bar{\pi}^b(\tilde{\omega}; z) \lesssim \frac{1}{(p\Delta_n)^{d/2}} e^{-V(z) - \frac{\lambda_1}{p\Delta_n} |\tilde{\omega} - z|^2} \lesssim \frac{1}{(p\Delta_n)^{d/2}} e^{-\frac{\lambda_1}{p\Delta_n} |\tilde{\omega} - z|^2}.$$

From Corollary 4.19, we also have

$$\mathfrak{p}_{p,n,(k-1)p\Delta_n}^b(x, y) \lesssim \frac{1}{(2(k-1)p\lambda_1)^{d/2} (p\Delta_n)^{d/2}} \exp\left(-\frac{\lambda_1}{(1+2(k-1))\Delta_n} |x-y|^2\right).$$

Then, we combine these two estimates using Lemma 4.34. For all $\omega, \tilde{\omega} \in \mathbb{R}^d$, we have

$$\int_{\mathbb{R}^d} \mathfrak{p}_{p,n,(k-1)p\Delta_n}^b(z, \tilde{\omega}) \bar{\pi}^b(\omega, z) dz \lesssim \frac{1}{((1+(k-1)\lambda_1)p\Delta_n)^{d/2}} \exp\left(-\frac{|\omega - \tilde{\omega}|^2}{(1+(k-1)\lambda_1)p\Delta_n}\right).$$

Then,

$$\begin{aligned} \mathfrak{D}_p^b(k; x, u, v) &\lesssim \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |\mathbf{K}_h(x + v - \omega) \mathbf{K}_h(x + u - \tilde{\omega})| s^{-d/2} \exp\left(-\frac{|\omega - \tilde{\omega}|^2}{s}\right) d\tilde{\omega} d\omega \\ &\leq \int_{\mathbb{R}^d} |\mathbf{K}_h(x + v + \omega)| \int_{\mathbb{R}^d} |\mathbf{K}_h(x + u - \tilde{\omega})| s^{-d/2} \exp\left(-\frac{|\omega - \tilde{\omega}|^2}{s}\right) d\tilde{\omega} d\omega. \end{aligned}$$

where $s = (1 + (k-1)\lambda_1)p\Delta_n$. We first focus on the inner integral. We recall that for any $z = (z_1, \dots, z_d) \in \mathbb{R}^d$,

$$\mathbf{K}_h(z) = \prod_{i=1}^d h_i^{-1} K(h_i^{-1} z_i),$$

and using that K is integrable, we obtain for any $i \in \{1, \dots, d\}$,

$$\begin{aligned} \int_{\mathbb{R}} h_i^{-1} \left| K\left(\frac{x_i + u_i + \tilde{\omega}_i}{h_i}\right) \right| \frac{1}{\sqrt{s}} \exp\left(-\frac{|\omega_i - \tilde{\omega}_i|^2}{s}\right) d\tilde{\omega}_i &\lesssim \int_{\mathbb{R}} h_i^{-1} \left| K\left(\frac{x_i + u_i + \tilde{\omega}_i}{h_i}\right) \right| \frac{1}{\sqrt{s}} d\tilde{\omega}_i \\ &\lesssim \frac{1}{\sqrt{s}} \int_{\mathbb{R}} \left| K(\hat{\omega}_i) \right| d\hat{\omega}_i \\ &\lesssim \frac{1}{\sqrt{s}}. \end{aligned} \quad (4.59)$$

Using that K is bounded, we also have

$$\begin{aligned} \int_{\mathbb{R}} h_i^{-1} \left| K\left(\frac{x_i + u_i + \tilde{\omega}_i}{h_i}\right) \right| \frac{1}{\sqrt{s}} \exp\left(-\frac{|\omega_i - \tilde{\omega}_i|^2}{s}\right) d\tilde{\omega}_i &\lesssim \int_{\mathbb{R}} \frac{1}{\sqrt{s}h_i} \exp\left(-\frac{|\omega_i - \tilde{\omega}_i|^2}{s}\right) d\tilde{\omega}_i \\ &\lesssim \frac{1}{h_i}. \end{aligned} \quad (4.60)$$

Combining the bounds given by Equations (4.59) and (4.59), we obtain that for any $1 \leq k_1 \leq d$,

$$\int_{\mathbb{R}^d} |\mathbf{K}_h(x + u - \tilde{\omega})| s^{-d/2} \exp\left(-\frac{|\omega - \tilde{\omega}|^2}{s}\right) d\tilde{\omega} \lesssim \frac{1}{s^{k_1/2} \prod_{i=k_1+1}^d h_i},$$

and therefore we have

$$\mathfrak{D}_p^b(k; x, u, v) \lesssim \frac{1}{s^{k_1/2} \prod_{i=k_1+1}^d h_i} \int_{\mathbb{R}^d} |\mathbf{K}_h(x + v + \omega)| d\omega \lesssim \frac{1}{s^{k_1/2} \prod_{i=k_1+1}^d h_i}.$$

By definition of s , we obtain

$$\mathfrak{D}_p^b(k; x, u, v) \lesssim \frac{1}{(kp\Delta_n)^{k_1/2} \prod_{i=k_1+1}^d h_i}.$$

The results on $\mathfrak{E}_p^b(k; x)$ follows by integrating (4.55).

11.4 Proof of Lemma 4.28

We begin by extending the bound on $\mathfrak{p}_{p,n,t}^b(\cdot, \cdot)$ given by Corollary 4.19 for $t > 1$. In fact, for $x, y \in \mathbb{R}^d$ and $t > 1$, we get using Markov property

$$\mathfrak{p}_{p,n,t}^b(x; y) = \int_{\mathbb{R}^d} p_t^b(x; z) \mathfrak{p}_{p,n}^b(z; y) dz.$$

Moreover, using the fact that X is a Markov process, we have that

$$\mathfrak{p}_{p,n,t}^b(x; y) = \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} p_{1/2}^b(x, \omega) p_{t-1/2}^b(\omega, z) d\omega \right) \mathfrak{p}_{p,n}^b(z; y) dz$$

From Lemma 4.18, we know that

$$p_{1/2}^b(x, \omega) \lesssim \exp\left(-|x - \omega|^2 + V(x) - V(\omega)\right)$$

and therefore using that V is bounded below, we have

$$\mathbf{p}_{p,n,t}^b(x; y) \lesssim e^{V(x)} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} p_{t-1/2}^b(\omega; z) d\omega \right) \mathbf{p}_{p,n}^b(z; y) dz \lesssim e^{V(x)}.$$

Moreover, we know from Lemma 4.20 that there exists $\lambda_1 > 0$ such that

$$\bar{\pi}^b(\tilde{\omega}; z) \lesssim \frac{1}{(p\Delta_n)^{d/2}} e^{-V(z) - \frac{\lambda_1}{p\Delta_n} |\tilde{\omega} - z|^2}.$$

Proceeding as for Lemma 4.27, we have

$$\begin{aligned} & \mathfrak{D}_p^b(k; x, u, v) \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{K}_h(x + v - \omega) \mathbf{K}_h(x + u - \tilde{\omega}) \mathbf{p}_{p,n,(k-1)p\Delta_n}^b(z; \omega) \bar{\pi}^b(\tilde{\omega}; z) d\tilde{\omega} d\omega dz \\ &\lesssim \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{K}_h(x + v - \omega) \mathbf{K}_h(x + u - \tilde{\omega}) \frac{1}{(p\Delta_n)^{d/2}} e^{-\frac{\lambda_1}{p\Delta_n} |\tilde{\omega} - z|^2} d\tilde{\omega} d\omega dz. \end{aligned}$$

Using that

$$\frac{1}{(p\Delta_n)^{d/2}} \int_{\mathbb{R}^d} \exp\left(-\frac{\lambda_1}{p\Delta_n} |z - \omega|^2\right) dz = \left(\frac{4\pi}{\lambda_1}\right)^{d/2},$$

we obtain

$$\mathfrak{D}_p^b(k; x, u, v) \lesssim \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{K}_h(x + v - \omega) \mathbf{K}_h(x + u - \tilde{\omega}) d\tilde{\omega} d\omega$$

and using the usual change of variable and the fact that $|K|$ is integrable, we get

$$\mathfrak{D}_p^b(k; x, u, v) \lesssim 1.$$

The result on $\mathfrak{C}_p^b(k; x)$ follows by integrating (4.55).

11.5 Proof of Lemma 4.29

For $\mathbf{h} = (h_1, \dots, h_d) \in (0, 1)^d$ and $y \in \mathbb{R}^d$, we define

$$\mathfrak{K}_h(y) = \int_{\mathbb{R}^d} \mathbf{K}_h(x - y + u) \chi_{p,n}(du),$$

and

$$\mathfrak{K}_h^c(y) = \mathfrak{K}_h(y) - \mathbb{E}_{\bar{\mu}^b} \left[\mathfrak{K}_h \left(\frac{1}{p} \sum_{\ell=0}^{p-1} X_{\ell\Delta_n} \right) \right].$$

Then we have by definition

$$\begin{aligned}
 \mathfrak{E}_p^b(k; x) &\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathfrak{E}_p^b(k; x, u, v) \chi_{p,n}(du) \chi_{p,n}(dv) \\
 &\leq \mathbb{E}_{\bar{\mu}^b}^b \left[\mathfrak{K}_h \left(\frac{1}{p} \sum_{\ell=0}^{p-1} X_{(kp+\ell)\Delta_n} \right) \mathfrak{K}_h \left(\frac{1}{p} \sum_{\ell=0}^{p-1} X_{\ell\Delta_n} \right) \right] - \mathbb{E}_{\bar{\mu}^b}^b \left[\mathfrak{K}_h \left(\frac{1}{p} \sum_{\ell=0}^{p-1} X_{\ell\Delta_n} \right) \right]^2, \\
 &\leq \mathbb{E}_{\bar{\mu}^b}^b \left[\mathfrak{K}_h^c \left(\frac{1}{p} \sum_{\ell=0}^{p-1} X_{(kp+\ell)\Delta_n} \right) \mathfrak{K}_h^c \left(\frac{1}{p} \sum_{\ell=0}^{p-1} X_{\ell\Delta_n} \right) \right].
 \end{aligned}$$

Moreover, we write $\{\mathcal{A}_k, k \geq 0\}$ for the natural filtration induced by the lifted Markov chain

$$\left(p^{-1} \sum_{\ell=0}^{p-1} Y_{kp+\ell, n}, X_{(k+1)p\Delta_n} \right)_{k \in \mathbb{N}}.$$

Then using Cauchy-Schwarz inequality, we have

$$\begin{aligned}
 &\left| \mathbb{E}_{\bar{\mu}^b}^b \left[\mathfrak{K}_h^c \left(\frac{1}{p} \sum_{\ell=0}^{p-1} X_{(kp+\ell)\Delta_n} \right) \mathfrak{K}_h^c \left(\frac{1}{p} \sum_{\ell=0}^{p-1} X_{\ell\Delta_n} \right) \right] \right| \\
 &= \left| \mathbb{E}_{\bar{\mu}^b}^b \left[\mathbb{E}_{\bar{\mu}^b}^b \left[\mathfrak{K}_h^c \left(\frac{1}{p} \sum_{\ell=0}^{p-1} X_{(kp+\ell)\Delta_n} \right) \middle| \mathcal{A}_0 \right] \mathfrak{K}_h^c \left(\frac{1}{p} \sum_{\ell=0}^{p-1} X_{\ell\Delta_n} \right) \right] \right| \\
 &\leq \mathbb{E}_{\bar{\mu}^b}^b \left[\mathbb{E}_{\bar{\mu}^b}^b \left[\mathfrak{K}_h^c \left(\frac{1}{p} \sum_{\ell=0}^{p-1} X_{(kp+\ell)\Delta_n} \right) \middle| \mathcal{A}_0 \right]^2 \right]^{1/2} \mathbb{E}_{\bar{\mu}^b}^b \left[\mathfrak{K}_h^c \left(\frac{1}{p} \sum_{\ell=0}^{p-1} X_{\ell\Delta_n} \right)^2 \right]^{1/2}.
 \end{aligned}$$

The second expectation is treated as in Lemma 4.26 and we have

$$\mathbb{E}_{\bar{\mu}^b}^b \left[\mathfrak{K}_h^c \left(\frac{1}{p} \sum_{\ell=0}^{p-1} X_{\ell\Delta_n} \right)^2 \right]^{1/2} \lesssim \prod_{i=1}^d h_i^{-1}.$$

For the first one, we use that X is a stationary Markov process to get

$$\begin{aligned}
 &\mathbb{E}_{\bar{\mu}^b}^b \left[\mathbb{E}^b \left[\mathfrak{K}_h^c \left(\frac{1}{p} \sum_{\ell=0}^{p-1} X_{(kp+\ell)\Delta_n} \right) \middle| \mathcal{A}_0 \right]^2 \right] = \mathbb{E}_{\bar{\mu}^b}^b \left[\mathbb{E}^b \left[\mathfrak{K}_h^c \left(\frac{1}{p} \sum_{\ell=0}^{p-1} X_{(kp+\ell)\Delta_n} \right) \middle| \mathcal{F}_{p\Delta_n} \right]^2 \right] \\
 &= \int_{\mathbb{R}^d} \mathbb{E}^b \left[\mathfrak{K}_h^c \left(\frac{1}{p} \sum_{\ell=0}^{p-1} X_{(kp+\ell)\Delta_n} \right) \middle| X_{p\Delta_n} = x \right]^2 \bar{\mu}^b(dx) \\
 &= \int_{\mathbb{R}^d} \mathbb{E}^b \left[\mathbb{E}^b \left[\mathfrak{K}_h^c \left(\frac{1}{p} \sum_{\ell=0}^{p-1} X_{(kp+\ell)\Delta_n} \right) \middle| X_{kp\Delta_n} \right] \middle| X_{p\Delta_n} = x \right]^2 \bar{\mu}^b(dx)
 \end{aligned}$$

Defining $\varphi : y \mapsto \mathbb{E}^b \left[\mathfrak{K}_h^c \left(\frac{1}{p} \sum_{\ell=0}^{p-1} X_{(kp+\ell)\Delta_n} \right) \middle| X_{kp\Delta_n} = y \right]$, we get that for any $x \in \mathbb{R}^d$,

$$\mathbb{E}^b[\varphi(X_{kp\Delta_n}) | X_{p\Delta_n} = x] = \mathbb{E}_x^b[\varphi(X_{(k-1)p\Delta_n})] = P_{(k-1)p\Delta_n}^b \varphi(x),$$

where (P_t^b) is defined in Equation (4.7). Moreover, from the definition of \mathfrak{K}_c , we get $\mathbb{E}_{\mu^b}^b[\varphi(X_0)] = 0$. Then, using Equation (4.6), we have

$$\begin{aligned} \mathbb{E}_{\mu^b}^b \left[\mathbb{E}^b \left[\mathfrak{K}_h^c \left(\frac{1}{p} \sum_{\ell=0}^{p-1} X_{(kp+\ell)\Delta_n} \right) \middle| \mathcal{F}_{p\Delta_n} \right]^2 \right] &= \text{Var}_{\mu^b}^b [P_{(k-1)p\Delta_n}^b \varphi(X_0)] \\ &\leq e^{-2C_{PI}^{-1}(k-1)p\Delta_n} \mathbb{E}_{\mu^b}^b [\varphi(X_0)^2]. \end{aligned}$$

Moreover, for all $x \in \mathbb{R}^d$, $|\varphi(x)| \leq 2\|\mathfrak{K}_h^c\|_\infty \leq 2\|\mathbf{K}_h\|_\infty$. Then, we have $\|\varphi\|_\infty \lesssim \prod_{i=1}^d h_i^{-1}$. Combining all previous inequalities, we get

$$\mathfrak{C}_p^b(k; x) \lesssim e^{-C_{PI}^{-1}(k-1)p\Delta_n} \left(\prod_{i=1}^d h_i^{-1} \right)^2.$$

12 Proof of the results of Section 4.1

12.1 Preliminary results

In this section, we aim at proving Lemma 4.6 and other results used in the proof of the results of Section 4.1. We start by introducing a concise notation. We say that $A(\theta) \lesssim B(\theta)$ implies $C(\theta) \lesssim D(\theta)$, if for all $c_1 > 0$, there exists $c_2 > 0$ such that $A(\theta) \leq c_1 B(\theta)$ implies $C(\theta) \leq c_2 D(\theta)$. We define similarly the equivalence between $A(\theta) \lesssim B(\theta)$ and $C(\theta) \lesssim D(\theta)$. We can now state the main Lemma of this section.

Lemma 4.30. *For $p \geq 1$, the condition $p\Delta_n \lesssim w_n^{HF}$ is equivalent to $p\Delta_n \lesssim \tilde{w}_{n,p}^{HF}$ with*

$$\tilde{w}_{n,p}^{HF} = \begin{cases} \left(\frac{p}{n} \right)^{\frac{1}{2}} \log \left(\frac{n}{p} \right)^{\frac{1}{2}} & \text{if } d = 1, 2 \\ \left(\frac{p}{n} \right)^{\frac{\bar{\alpha}}{2\bar{\alpha}+d} \left(\frac{1}{\alpha_1} + \frac{1}{\alpha_2} \right)} \log \left(\frac{n}{p} \right) & \text{if } d \geq 3 \text{ and } (\alpha, k_0) \in D_1, \\ \left(\frac{p}{n} \right)^{\frac{2\bar{\alpha}}{2\bar{\alpha}+d} \frac{1}{\alpha_1}} & \text{if } d \geq 3 \text{ and } (\alpha, k_0) \in D_2, \\ \left(\frac{p}{n} \right)^{\frac{\bar{\alpha}}{(2\bar{\alpha}+d) \left(\frac{1}{\alpha_1} + \frac{1}{\alpha_2} \right)}} & \text{if } d \geq 3 \text{ and } d \geq 3 \text{ and } (\alpha, k_0) \in D_3. \end{cases}$$

Proof. First note that since $T_n = \frac{np\Delta_n}{p}$, the condition

$$p\Delta_n \lesssim T_n^{-u} \log(T_n)^v$$

for some $u, v \geq 0$, is equivalent to

$$p\Delta_n \lesssim \left(\frac{p}{n} \right)^{\frac{u}{1+u}} \log \left(\frac{n}{p} \right)^{\frac{v}{1+u}}.$$

If $u = \frac{\bar{\alpha}_3}{2\bar{\alpha}_3+d-2} \left(\frac{1}{\alpha_1} + \frac{1}{\alpha_2} \right)$, we have

$$u = \frac{1}{2 + \frac{d-2}{\bar{\alpha}_3}} \frac{\alpha_1 + \alpha_2}{\alpha_1 \alpha_2} = \frac{1}{2 + \frac{d}{\bar{\alpha}} - \frac{1}{\alpha_1} - \frac{1}{\alpha_2}} \frac{\alpha_1 + \alpha_2}{\alpha_1 \alpha_2} = \frac{\bar{\alpha}(\alpha_1 + \alpha_2)}{\alpha_1 \alpha_2 (2\bar{\alpha} + d) - \bar{\alpha}(\alpha_1 + \alpha_2)}$$

and therefore

$$\frac{u}{1+u} = \frac{\bar{\alpha}(\alpha_1 + \alpha_2)}{\alpha_1\alpha_2(2\bar{\alpha} + d)} = \frac{\bar{\alpha}}{2\bar{\alpha} + d} \left(\frac{1}{\alpha_1} + \frac{1}{\alpha_2} \right),$$

which proves the case $d \geq 3$ and $k_0 = 1$ with $\alpha_2 < \alpha_3$ or $k_0 = 2$. The other cases are done similarly. \square

A particular case is the case where $p = 1$, restated as follows

Lemma 4.31. *The condition $\Delta_n \lesssim w_n^{HF}$ is equivalent to $\Delta_n \lesssim \tilde{w}_n^{HF}$ with $\tilde{w}_n^{HF} = \tilde{w}_{n,1}^{HF}$*

Proof of Lemma 4.6. From Lemma 4.30, we know that $p\Delta_n \lesssim w_n^{HF}$ is equivalent to $p\Delta_n \lesssim \tilde{w}_{n,p}^{HF}$. Using that $\alpha_1 > 2$, we check that

$$\tilde{w}_{n,p}^{HF} \geq \left(\frac{p}{n} \right)^{\frac{2\bar{\alpha}}{2\bar{\alpha}+d}}$$

and therefore we have Lemma 4.6. \square

12.2 Proof of Proposition 4.7

Suppose that $p^*\Delta_n \leq w_n^{HF}$ and that $\tau_n^{2\alpha_1} \leq \Delta_n$ and recall that in that case, we take

$$p^* = 1 \quad \text{and} \quad \mathbf{h}^* = \mathbf{h}^{*,HF}$$

From Proposition 4.4, we see by plugging these values that

$$B_{n,\mathbf{h},p}^b(x) \lesssim \tau_n^{2\alpha_1} + v_n^{HF}.$$

Note also that $|\log(h_i^*)|$ is of the order of $\log(T_n)$ and that by definition of \mathbf{h}^* , we have

$$v_n^{HF} \lesssim T_n^{-1} \begin{cases} \sum_{i=1}^d |\log(h_i^*)| & \text{if } d = 1, 2 \\ \sum_{i=1}^d |\log(h_i^*)| \prod_{i=3}^d (h_i^*)^{-1} & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_1 \\ (\prod_{i=1}^{k_0} (h_i^*)^{(2-k_0)/k_0} \prod_{i=k_0+1}^d (h_i^*)^{-1}) & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_2 \\ (h_2^* h_3^*)^{-1/2} \prod_{i=4}^d (h_i^*)^{-1} & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_3. \end{cases}$$

Therefore, by factorizing by v_n^{HF} the variance bound from Proposition 4.5, we have

$$V_{n,\mathbf{h}^*,p^*}^b(x) \lesssim v_n^{HF} (1 + r_n^{HF}),$$

with

$$r_n^{HF} = \begin{cases} 0 & \text{if } d = 1, 2 \\ \Delta_n (h_1^* h_2^*)^{-1} \log(T_n)^{-1} & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_1 \\ \Delta_n \prod_{i=1}^{k_0} (h_i^*)^{-2/k_0} + \log(T_n) \prod_{i=1}^{k_0} (h_i^*)^{-2/k_0} \prod_{i=1}^d h_i^* & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_2 \\ \log(T_n) (h_2^* h_3^*)^{1/2} \prod_{i=4}^d h_i^* + \Delta_n h_1^{-1} (h_2^* h_3^*)^{-1/2} & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_3 \end{cases}$$

and it remains to prove that in each case, the remainder can be ignored so that $V_{n,\mathbf{h},p}^b(x) \lesssim v_n^{HF}$. First

note that the condition $\Delta_n \lesssim w_n^{HF}$ and the definition of \mathbf{h}^* ensures that

$$\Delta_n \lesssim \begin{cases} h_1^* \log(T_n) & \text{if } d = 1 \\ h_1^* h_2^* \log(T_n) & \text{if } d = 2 \\ h_1^* h_2^* \log(T_n) & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_1 \\ \prod_{i=1}^{k_0} (h_i^*)^{2/k_0} & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_2 \\ h_1^* (h_2^* h_3^*)^{1/2} & \text{if } dd \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_3 \end{cases}$$

and therefore

$$r_n^{HF} \lesssim \begin{cases} 0 & \text{if } d = 1, 2 \text{ or } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_1 \\ \log(T_n) \prod_{i=1}^{k_0} (h_i^*)^{-2/k_0} \prod_{i=1}^d h_i^* & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_2 \\ \log(T_n) (h_2^* h_3^*)^{1/2} \prod_{i=4}^d h_i^* & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_3 \end{cases}$$

Then we easily check that if $d \geq 3$ and $k_0 \geq 3$, we have

$$\log(T_n) \prod_{i=1}^{k_0} (h_i^*)^{-2/k_0} \prod_{i=1}^d h_i^* = \log(T_n) T_n^{-\frac{\bar{\alpha}_2}{\bar{\alpha}_3 + d - 2} \sum_{j=k_0+1}^d 1/\alpha_j} \rightarrow 0.$$

Analogously, if $d \geq 3$ and $k_0 = 1$ with $\alpha_2 = \alpha_3$, we have

$$\log(T_n) (h_2^* h_3^*)^{-1/2} \prod_{i=4}^d h_i^* = \log(T_n)^{c+1} T_n^{-c} \rightarrow 0,$$

with

$$c = \frac{\bar{\alpha}_3}{2\bar{\alpha}_3 + d - 2} \left(\frac{1}{2\alpha_2} + \frac{1}{2\alpha_3} + \sum_{i=4}^d \frac{1}{\alpha_i} \right) > 0$$

so that

$$V_{n, \mathbf{h}^*, p^*}^b(x) \lesssim v_n^{HF}$$

which concludes the proof of Proposition 4.7.

12.3 Proof of Proposition 4.8

Suppose that $p^* \Delta_n \leq w_n^{HF}$ and that $\tau_n^{2\alpha_1} \geq \Delta_n$ and recall that in that case, we take

$$p^* = \left[(\tau_n^{2\alpha_1} \Delta_n^{-1})^{1/(1+\alpha_1)} \right] \quad \text{and} \quad \mathbf{h}^* = \mathbf{h}^{*, HF}$$

From Proposition 4.4, we see by plugging these values that

$$B_{n, \mathbf{h}^*, p^*}^b(x) \lesssim \frac{\tau_n^{2\alpha_1}}{(p^*)^{\alpha_1}} + (p^* \Delta_n)^{1/2} + v_n^{HF} \lesssim (\tau_n^2 \Delta_n)^{\frac{\alpha_1}{1+\alpha_1}} + v_n^{HF}.$$

For the variance, we proceed as for Proposition 4.8: using that $p^* \Delta_n \lesssim w_n^{HF}$ and $\mathbf{h}^* = \mathbf{h}^{*, HF}$, we

have from Proposition 4.5 that

$$V_{n, \mathbf{h}^*, p^*}^b(x) \lesssim v_n^{HF}.$$

so we can conclude.

12.4 Proof of Proposition 4.10

Suppose that $\Delta_n \geq w_n^{HF}$ and that $\tau_n^{2\alpha_1} \leq \Delta_n$ and recall that in that case, we take

$$p^* = 1 \quad \text{and} \quad \mathbf{h}^* = \mathbf{h}^{*,1} = \mathbf{h}^{*,LF}$$

From Proposition 4.4, we see by plugging these values that

$$B_{n, \mathbf{h}, p}^b(x) \lesssim \tau_n^{2\alpha_1} + v_n^{HF}.$$

We now study the variance. First, note that

$$T_n^{-1} \Delta_n \prod_{i=1}^d (h_i^*)^{-1} = n^{\frac{-2\bar{\alpha}}{2\bar{\alpha}+d}} = v_n^{LF}.$$

Therefore, by factorizing by v_n^{LF} the variance bound from Proposition 4.5 and using that $|\log(h_i^*)|$ is of the order of $\log(n)$, we have

$$V_{n, \mathbf{h}^*, p^*}^b(x) \lesssim v_n^{LF} (1 + r_n^{LF}),$$

with

$$r_n^{LF} = \begin{cases} 0 & \text{if } d = 1, 2 \\ \Delta_n^{-1} \log(n) h_1^* h_2^* & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_1 \\ \Delta_n^{-1} \prod_{i=1}^{k_0} (h_i^*)^{2/k_0} + \log(n) \Delta_n^{-1} \prod_{i=1}^d h_i^* & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_2 \\ \log(n) \Delta_n^{-1} \prod_{i=1}^d h_i^* + \Delta_n^{-1} h_1 (h_2^* h_3^*)^{1/2} & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_3 \end{cases}$$

Using Lemma 4.31 and the fact that $\Delta_n \gtrsim w_n^{HF}$, we know that $\Delta_n \gtrsim \tilde{w}_n^{HF}$. Using also the definition of \mathbf{h}^* , we check that

$$\Delta_n \gtrsim \begin{cases} h_1^* |\log(h_1^*)| & \text{if } d = 1 \\ h_1^* h_2^* |\log(h_1^* h_2^*)| & \text{if } d = 2 \\ h_1 h_2 \log(n) & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_1 \\ \prod_{i=1}^{k_0} (h_i^*)^{2/k_0} & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_2 \\ h_1^* (h_2^* h_3^*)^{1/2} & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_3. \end{cases} \quad (4.61)$$

and therefore

$$V_{n, \mathbf{h}^*, p^*}^b(x) \lesssim v_n^{LF} \begin{cases} 1 & \text{if } d = 1, 2 \text{ or } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_1 \\ 1 + \log(n) \Delta_n^{-1} \prod_{i=1}^d h_i^* & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_2 \\ 1 + \log(n) \Delta_n^{-1} \prod_{i=1}^d h_i^* & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_3. \end{cases}$$

We conclude the proof of Proposition 4.10 using (4.61) to get

$$\log(n) \Delta_n^{-1} \prod_{i=1}^d h_i^* \lesssim \begin{cases} \log(n) \prod_{i=1}^{k_0} (h_i^*)^{(2-k_0)/k_0} \prod_{i=k_0+1}^d h_i^* & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_2 \\ (h_2^* h_3^*)^{1/2} \prod_{i=4}^d h_i^* & \text{if } d \geq 3 \text{ and } (\boldsymbol{\alpha}, k_0) \in D_3. \end{cases}$$

so that

$$\log(n) \Delta_n^{-1} \prod_{i=1}^d h_i^* \rightarrow 0,$$

as $n \rightarrow \infty$.

12.5 Proof of Proposition 4.11

Suppose that $p^* \Delta_n \geq w_n^{HF}$ and that $\tau_n^{2\alpha_1} \geq \Delta_n$. and recall that in that case, we take

$$p^* = \left[(\tau_n^{2\alpha_1} \Delta_n^{-1})^{1/(1+\alpha_1)} \right] \quad \text{and} \quad \mathbf{h}^* = \mathbf{h}^{*, p^*}$$

From Proposition 4.4, we see by plugging these values that

$$B_{n, \mathbf{h}^*, p^*}^b(x) \lesssim (\tau_n^2 \Delta_n)^{\frac{\alpha_1}{1+\alpha_1}} + \left(\frac{p^*}{n} \right)^{\frac{2\bar{\alpha}}{2\bar{\alpha}+d}}$$

For the variance, we proceed as for Proposition 4.10: using that $p^* \Delta_n \gtrsim w_n^{HF}$ (and therefore that $p^* \Delta_n \gtrsim \tilde{w}_n^{HF}$ by Lemma 4.6) and $\mathbf{h}^* = \mathbf{h}^{*, p^*}$, we have from Proposition 4.5 that

$$V_{n, \mathbf{h}^*, p^*}^b(x) \lesssim \left(\frac{p^*}{n} \right)^{\frac{2\bar{\alpha}}{2\bar{\alpha}+d}}$$

and thus

$$\mathcal{R}(\hat{\mu}_{n, \mathbf{h}^*, p^*}, b; x) \lesssim (\tau_n^2 \Delta_n)^{\frac{\alpha_1}{1+\alpha_1}} + \left(\frac{p^*}{n} \right)^{\frac{2\bar{\alpha}}{2\bar{\alpha}+d}}.$$

It remains to prove that $(\tau_n^2 \Delta_n)^{\frac{\alpha_1}{1+\alpha_1}}$ is always dominating here. This term is indeed of the order of $p^* \Delta_n$. By Lemma 4.30, we have $p^* \Delta_n \gtrsim \tilde{w}_{n, p^*}^{HF}$ so it is enough to prove that

$$\left(\frac{p^*}{n} \right)^{\frac{2\bar{\alpha}}{2\bar{\alpha}+d}} \lesssim \tilde{w}_{n, p^*}^{HF} \quad (4.62)$$

Moreover, since $p^* \Delta_n \rightarrow 0$, it is clear that $p^*/n \rightarrow 0$. Using the definition of \tilde{w}_{n, p^*}^{HF} , we see that (4.62)

is equivalent to

$$\begin{cases} 2\bar{\alpha} \geq d & \text{if } d = 1, 2 \\ \frac{\bar{\alpha}}{2\bar{\alpha}+d} \left(\frac{1}{\alpha_1} + \frac{1}{\alpha_2} \right) \geq \frac{2\bar{\alpha}}{2\bar{\alpha}+d} & \text{if } d \geq 3 \text{ and } (\alpha, k_0) \in D_1 \\ \frac{2\bar{\alpha}}{2\bar{\alpha}+d} \frac{1}{\alpha_1} \geq \frac{2\bar{\alpha}}{2\bar{\alpha}+d} & \text{if } d \geq 3 \text{ and } (\alpha, k_0) \in D_2 \\ \frac{\bar{\alpha}}{(2\bar{\alpha}+d)} \left(\frac{1}{\alpha_1} + \frac{1}{\alpha_2} \right) \geq \frac{2\bar{\alpha}}{2\bar{\alpha}+d} & \text{if } d \geq 3 \text{ and } (\alpha, k_0) \in D_3. \end{cases}$$

These inequalities always hold since $\alpha_1 > 1$ and $\alpha_2 > 1$ so we can conclude.

13 Proof of the results of Section 5

13.1 Proof of Theorem 4.12

The main idea of this proof is to use general results for Markov chains to the lifted Markov chain

$$\left(p^{-1} \sum_{\ell=0}^{p-1} Y_{k p + \ell, n}, X_{(k+1)p \Delta_n} \right)_{k \in \mathbb{N}}. \quad (4.63)$$

We first introduce few definitions for Markov chains and we refer to [KM12] for more details. Consider a Markov chain Υ on a given space (X, \mathcal{X}) . We say that Υ is ψ -irreducible and aperiodic if there exists a σ -finite measure ψ on (X, \mathcal{X}) such that for all $A \in \mathcal{X}$ such that $\psi(A) > 0$, any $x \in X$ and any n large enough, we have

$$P^n(x, A) > 0$$

where P denotes the transition semigroup of Υ . We say that Υ is geometrically ergodic if it admits an invariant measure π and functions $\rho : X \rightarrow (0, 1)$ and $\rho : X \rightarrow [1, \infty)$ such that for all $n \geq 0$ and for π -almost $x \in X$, we have

$$\|P^n(x, \cdot) - \pi\|_{TV} \leq C(x)\rho(x)^n.$$

Intuitively, geometrically ergodic converge fast to their invariant distribution starting from almost any point. Therefore, they must satisfy concentration properties such as Bernstein inequality. In this section, we plan to use the following result, from [Lem21].

Theorem 4.32 (Theorem 1.1 in [Lem21]). *Let Υ be a geometrically ergodic Markov chain with state space \mathcal{X} , and let π be its unique stationary probability measure. Moreover, let $f : \mathcal{X} \rightarrow \mathbb{R}$ be a bounded measurable function such that $\mathbb{E}_\pi f = 0$. Furthermore, let $x \in \mathcal{X}$. Then, we can find constants $K, \tau > 0$ depending only on x and the transition probability $P(\cdot, \cdot)$ such that for all $t > 0$,*

$$\mathbb{P}_x \left(\frac{1}{N} \sum_{i=0}^{N-1} f(\Upsilon_i) > t \right) \leq K \exp \left(- \frac{N^2 t^2}{32N\sigma_{Mrv}^2 + \tau t \|f\|_\infty N \log N} \right),$$

where

$$\sigma_{Mrv}^2 = \text{Var}_\pi(f(\Upsilon_0)) + 2 \sum_{i=1}^{\infty} \text{Cov}_\pi(f(\Upsilon_0), f(\Upsilon_i))$$

denotes the asymptotic variance of the process $(f(\Upsilon_i))_i$.

We want to apply this Theorem to (4.63) and we need to check that this chain is geometrically ergodic. An easy way to check for geometric ergodicity is to use Chapter 15 from [MT93]: any ψ -irreducible and aperiodic Markov chain satisfying the drift condition 10 stated below is geometrically ergodic.

Assumption 10. (V4) *There exists a function $W : X \rightarrow [1, \infty)$, a set $C \in \mathcal{X}$ and constants $\delta > 0$ and $b < \infty$ such that*

$$PW \leq (1 - \delta)W + b\mathbb{1}_C \quad (4.64)$$

and such that there exists $n \geq 1, \varepsilon > 0$ and a probability measure ν on (X, \mathcal{X}) such that

$$P^n(x, A) \geq \varepsilon \mathbb{1}_C(x) \nu(A) \quad (4.65)$$

for all $x \in X$ and $A \in \mathcal{X}$. In that case, we say that W is a Lyapunov function and C is a small set.

Thus the plan is to prove that the lifted chain (4.63) is ψ -irreducible and aperiodic and satisfies Assumption 10.

First remark that the Markov chain $(X_{pk\Delta_n})_k$ is clearly aperiodic and ψ -irreducible. In fact for any $x \in \mathbb{R}^d$ and any Borelian $A \in \mathcal{B}_{\mathbb{R}^d}$ of positive Lebesgue measure, and for any $k \geq 1$, we have

$$\mathbb{P}^b(X_{pk\Delta_n} \in A \mid X_0 = x) > 0. \quad (4.66)$$

We deduce that the lifted Markov chain (4.63) is also aperiodic and ψ -irreducible by using that the density $\mathfrak{p}_{p,n}^b(x; \cdot, \cdot)$ is almost everywhere positive.

Moreover, using Assumption 9, we know that the chain is geometrically ergodic. Combined with aperiodicity and ψ -irreducibility, this ensures that $(X_{pk\Delta_n})_k$ satisfies Assumption 10 and we consider W and C as in Assumption 10. We want to prove that the same property hold for (4.63). We consider $\widetilde{W} : (x, y) \mapsto W(y)$ and $\widetilde{C} = \mathbb{R}^d \times C$. Then, for all $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$, we have

$$\begin{aligned} \mathfrak{P}\widetilde{W}(x, y) &= \mathbb{E}\left[\widetilde{W}\left(p^{-1} \sum_{\ell=0}^{p-1} Y_{p+\ell,n}, X_{2p\Delta_n}\right) \mid p^{-1} \sum_{\ell=0}^{p-1} Y_{p+\ell,n} = x, X_{p\Delta_n} = y\right] \\ &= \mathbb{E}\left[\widetilde{W}\left(p^{-1} \sum_{\ell=0}^{p-1} Y_{p+\ell,n}, X_{2p\Delta_n}\right) \mid X_{p\Delta_n} = y\right]. \end{aligned}$$

and therefore

$$\mathfrak{P}\widetilde{W}(x, y) \leq \lambda \widetilde{W}(x, y) + b\mathbb{1}_C(y).$$

We only need to check that \widetilde{C} is indeed a *small set*. We already know that C is a *small set* and therefore there exists $\delta > 0$ and $\nu \in \mathcal{P}(\mathbb{R}^d)$ such that for some $k \geq 1$,

$$\forall x \in C, \quad \forall B \in \mathcal{B}_{\mathbb{R}^d}, \quad P^k(x, B) \geq \delta \nu(B).$$

We now define the following probability measure on \mathbb{R}^{2d}

$$\tilde{\nu}(\mathrm{d}x, \mathrm{d}y) = \int_{\mathbb{R}^d} \mathbf{p}_{p,n}^b(y; x, z) \mathrm{d}z \nu(\mathrm{d}x) \mathrm{d}y.$$

Then, for all $(\bar{x}, x) \in \tilde{C}$, we get for any Borel set of \mathbb{R}^d , $\bar{B}, B \in \mathcal{B}_{\mathbb{R}^d}$, conditioning with respect to $X_{pk\Delta_n}$ and using the Markov property:

$$\begin{aligned} \mathfrak{P}^{k+1}((\bar{x}, x), \bar{B} \times B) &= \int_{\mathbb{R}^d} p_{kp\Delta_n}(x, x') \int_{\bar{B} \times B} \mathbf{p}_{p,n}^b(x'; y, z) \mathrm{d}y \mathrm{d}z \mathrm{d}x' \\ &\geq \delta \int_{\bar{B} \times B} \int_{\mathbb{R}^d} \mathbf{p}_{p,n}^b(x'; y, z) \nu(\mathrm{d}x') \mathrm{d}y \mathrm{d}z \\ &\geq \delta \tilde{\nu}(\bar{B} \times B), \end{aligned}$$

where we used that $x \in C$ and C is a *small* set.

We are now ready to apply Theorem 4.32 to (4.63). We apply it with $f(x_1, x_2) = \mathbf{K}_{\mathbf{h}}(x - x_1)$ for all $x_1, x_2 \in \mathbb{R}^d$. Thus we get

$$\mathbb{P}_{\bar{\mu}^b}^b \left(|\hat{\nu}_{n,\mathbf{h},p}(x) - \mathbb{E}[\hat{\nu}_{n,\mathbf{h},p}(x)]| > \varepsilon \right) \leq K \exp \left(- \frac{N^2 \varepsilon^2}{32Nv^2(\alpha, n, \mathbf{h}, p) + \tau \varepsilon \|\mathbf{K}_{\mathbf{h}}\|_{\infty} N \log N} \right)$$

where $N = \lfloor n/p \rfloor$, and

$$\begin{aligned} v^2(\alpha, n, \mathbf{h}, p) &= \mathrm{Var}_{\bar{\mu}^b}^b \left(\mathbf{K}_{\mathbf{h}} \left(x - p^{-1} \sum_{\ell=0}^{p-1} Y_{\ell,n} \right) \right) \\ &\quad + 2 \sum_{k=1}^{\infty} \mathrm{Cov}_{\bar{\mu}^b}^b \left(\mathbf{K}_{\mathbf{h}} \left(x - p^{-1} \sum_{\ell=0}^{p-1} Y_{\ell,n} \right), \mathbf{K}_{\mathbf{h}} \left(x - p^{-1} \sum_{\ell=0}^{p-1} Y_{kp+\ell,n} \right) \right). \end{aligned}$$

In order to control the variance term, everything works exactly as in the proof of Proposition 4.5, except for the last term as the sum is infinite.

13.2 Proof of Proposition 4.14

In this section, we will repeatedly use estimates of the form

$$\int_{\nu}^{\infty} \exp(-z^r) \mathrm{d}z \leq 2r^{-1} \nu^{1-r} \exp(-\nu^r), \quad \nu, r > 0, \quad \nu \geq (2/r)^{1/r}. \quad (4.67)$$

and

$$\exp \left(- \frac{az^p}{b + cz^{p/2}} \right) \leq \exp \left(- \frac{az^p}{2b} \right) + \exp \left(- \frac{az^{p/2}}{2c} \right), \quad a, b, c, z > 0. \quad (4.68)$$

Before delving into the proof of Proposition 4.14, let us state and prove the following Lemma

Lemma 4.33. *Under Assumptions 7, 8, 9, there exists $\gamma > 1$ such that for any $\mathbf{h} \in \mathcal{H}_p^n$,*

$$\mathbb{E}_{\bar{\mu}^b}^b[A_n^p(\mathbf{h})] \lesssim V_p^n(\mathbf{h}) + \mathcal{B}_{n,\mathbf{h},p}(x)^2 + n_p^{-\gamma} + \sqrt{p\Delta_n} + \frac{\tau_n^{2\alpha_1}}{p^{\alpha_1}}.$$

Proof of Lemma 4.33. For any $\boldsymbol{\eta} \in \mathcal{H}_n^p$, we have

$$\left\{ \left| \hat{\mu}_{n,(\mathbf{h},\boldsymbol{\eta}),p}(x) - \hat{\mu}_{n,\boldsymbol{\eta},p}(x) \right|^2 - V_n^p(\boldsymbol{\eta}) \right\}_+ \lesssim \mathbf{I}_{\mathbf{h},\boldsymbol{\eta}}^{n,p}(x) + \mathbf{II}_{\mathbf{h},\boldsymbol{\eta}}^{n,p}(x) + \mathbf{III}_{\mathbf{h},\boldsymbol{\eta}}^{n,p}(x). \quad (4.69)$$

where

$$\begin{aligned} \mathbf{I}_{\mathbf{h},\boldsymbol{\eta}}^{n,p}(x) &= \left\{ \left| \hat{\mu}_{n,\boldsymbol{\eta},p}(x) - \mathbb{E}[\hat{\mu}_{n,\boldsymbol{\eta},p}(x)] \right|^2 - V_n^p(\boldsymbol{\eta}) \right\}_+, \\ \mathbf{II}_{\mathbf{h},\boldsymbol{\eta}}^{n,p}(x) &= \left\{ \left| \hat{\mu}_{n,(\mathbf{h},\boldsymbol{\eta}),p}(x) - \mathbb{E}[\hat{\mu}_{n,(\mathbf{h},\boldsymbol{\eta}),p}(x)] \right|^2 - V_n^p(\boldsymbol{\eta}) \right\}_+, \\ \mathbf{III}_{\mathbf{h},\boldsymbol{\eta}}^{n,p}(x) &= \left| \mathbb{E}_{\bar{\mu}^b}^b[\hat{\mu}_{n,\boldsymbol{\eta},p}(x)] - \mathbb{E}_{\bar{\mu}^b}^b[\hat{\mu}_{n,(\mathbf{h},\boldsymbol{\eta}),p}(x)] \right|^2. \end{aligned}$$

Let us begin with the term $\mathbf{I}_{\mathbf{h},\boldsymbol{\eta}}^{n,p}(x)$. We write

$$\begin{aligned} & \mathbb{E}_{\bar{\mu}^b}^b \left[\left\{ \left| \hat{\mu}_{n,\boldsymbol{\eta},p}(x) - \mathbb{E}_{\bar{\mu}^b}^b[\hat{\mu}_{n,\boldsymbol{\eta},p}(x)] \right|^2 - V_n^p(\boldsymbol{\eta}) \right\}_+ \right] \\ &= \int_{V_n^p(\boldsymbol{\eta})}^{+\infty} \mathbb{P}_{\bar{\mu}^b} \left(\left| \hat{\mu}_{n,\boldsymbol{\eta},p}(x) - \mathbb{E}_{\bar{\mu}^b}^b[\hat{\mu}_{n,\boldsymbol{\eta},p}(x)] \right| \geq z^{1/2} \right) dz \\ &\leq K \int_{V_n^p(\boldsymbol{\eta})}^{+\infty} \exp \left(-\frac{n_p^2 \beta^2 z}{32n_p v^2(\alpha, n, \boldsymbol{\eta}, p) + \tau \beta \|\mathbb{K}_{\boldsymbol{\eta}}\| n_p \log(n_p) z^{1/2}} \right) dz, \end{aligned}$$

where we used at the last line the concentration inequality of Bernstein's type given by Theorem 4.12.

Then, using Equation (4.68), we obtain

$$\mathbb{E}_{\bar{\mu}^b}^b \left[\left\{ \left| \hat{\mu}_{n,\boldsymbol{\eta},p}(x) - \mathbb{E}_{\bar{\mu}^b}^b[\hat{\mu}_{n,\boldsymbol{\eta},p}(x)] \right|^2 - V_n^p(\boldsymbol{\eta}) \right\}_+ \right] \lesssim \mathbf{I} + \mathbf{II},$$

where

$$\begin{aligned} \mathbf{I} &:= K \int_{V_n^p(\boldsymbol{\eta})}^{+\infty} \exp \left(-\frac{n_p \beta^2 z}{32v^2(\alpha, n, \boldsymbol{\eta}, p)} \right) dz; \\ \mathbf{II} &:= K \int_{V_n^p(\boldsymbol{\eta})}^{+\infty} \exp \left(-\frac{n_p \beta z^{1/2}}{\tau \log(n_p) \|\mathbb{K}_{\boldsymbol{\eta}}\|_{\infty}} \right) dz. \end{aligned}$$

The first term can be computed explicitly

$$\mathbf{I} = \frac{32K v^2(\alpha, n, \boldsymbol{\eta}, p)}{n_p \beta^2} \exp \left(-\frac{n_p \beta^2 V_n^p(\boldsymbol{\eta})}{32v^2(\alpha, n, \boldsymbol{\eta}, p)} \right).$$

Moreover, by definition of $v^2(\alpha, n, \boldsymbol{\eta}, p)$ in Theorem 4.12, and considering the fact that $\eta_1 \leq \dots \leq \eta_d$,

we obtain $v^2(\alpha, n, \boldsymbol{\eta}, p) = n_p v_n^p(\boldsymbol{\eta})$. Then,

$$\mathbf{I} \lesssim v_n^p(\boldsymbol{\eta}) \exp\left(-\frac{\bar{\omega} \log(n_p) \beta^2}{32}\right).$$

Moreover, we know that

$$v_n^p(\boldsymbol{\eta}) = \frac{1}{T_n} \max \left\{ p \Delta_n \prod_{i=1}^d \eta_i^{-1}, \min \left\{ \sum_{i=1}^d |\log(\eta_i)| \prod_{i=3}^d \eta_i^{-1}, (\eta_2 \eta_3)^{-1/2} \prod_{i=4}^d \eta_i^{-1} \right\} \right\}.$$

Using the particular structure of the grid \mathcal{H}_n^p , we obtain $v_n^p(\mathbf{h}) \lesssim 1 + n_p/T_n$, and

$$\mathbf{I} \lesssim n_p^{-\frac{\bar{\omega} \beta^2}{32}} \left(1 + \frac{n_p}{T_n}\right). \quad (4.70)$$

For the second term \mathbf{II} , using Equation (4.67) we obtain

$$\mathbf{II} \lesssim \frac{\log(n_p) \|\mathbb{K}_{\boldsymbol{\eta}}\|_{\infty} \sqrt{V_n^p(\boldsymbol{\eta})}}{n_p} \exp\left(-\frac{\beta n_p V_n^p(\boldsymbol{\eta})^{1/2}}{\tau \log(n_p) \|\mathbb{K}_{\boldsymbol{\eta}}\|_{\infty}}\right).$$

Moreover, we know that $\|\mathbb{K}_{\boldsymbol{\eta}}\|_{\infty} \lesssim \prod_{i=1}^d \eta_i^{-1}$. Then,

$$\frac{\log(n_p) \|\mathbb{K}_{\boldsymbol{\eta}}\|_{\infty} \sqrt{V_n^p(\boldsymbol{\eta})}}{n_p} \lesssim \frac{v_n^p(\mathbf{h})^{1/2} \log(n_p)^2 \bar{\omega}^{1/2}}{n_p \prod_{i=1}^d \eta_i}.$$

Using the definition of $v_n^p(\mathbf{h})$, we obtain

$$\begin{aligned} \prod_{i=1}^d \eta_i^{-1} v_n^p(\mathbf{h})^{1/2} &= \frac{1}{\sqrt{T_n}} \max \left\{ \sqrt{p \Delta_n} \prod_{i=1}^d h_i^{-3/2}, \right. \\ &\quad \left. \min \left\{ (h_1 h_2)^{-1} \sum_{i=1}^d |\log(h_i)| \prod_{i=3}^d h_i^{-3/2}, h_1^{-1} (h_2 h_3)^{-5/4} \prod_{i=4}^d h_i^{-3/2} \right\} \right\} \end{aligned}$$

Using once again the lower bound on the η_i for $i = 1, \dots, d$, we obtain

$$\prod_{i=1}^d \eta_i^{-1} v_n^p(\mathbf{h})^{1/2} \lesssim 1 + \frac{\sqrt{n_p}}{\sqrt{T_n}}.$$

Finally, we can write

$$\frac{n_p V_n^p(\boldsymbol{\eta})^{1/2}}{\log(n_p) \|\mathbb{K}_{\boldsymbol{\eta}}\|_{\infty}} \gtrsim \bar{\omega}^{1/2} \log(n_p) \quad (4.71)$$

In fact,

$$\frac{n_p V_n^p(\boldsymbol{\eta})^{1/2}}{\log(n_p) \|\mathbb{K}_{\boldsymbol{\eta}}\|_{\infty}} \gtrsim \frac{\bar{\omega}^{1/2} n_p \prod_{i=1}^d \eta_i v_n^p(\boldsymbol{\eta})^{1/2}}{\log(n_p)^{1/2}},$$

and

$$\prod_{i=1}^d \eta_i v_n^p(\boldsymbol{\eta})^{1/2} = \left(\prod_{i=1}^d \eta_i \right)^{1/2} \frac{1}{\sqrt{T_n}} \max \left\{ \sqrt{p\Delta_n}, \min \left\{ \left(\sum_{i=1}^d |\log(\eta_i)| \eta_1 \eta_2 \right)^{1/2}, \sqrt{\eta_1 (\eta_2 \eta_3)^{1/4}} \right\} \right\}$$

This ensures that

$$\begin{aligned} \prod_{i=1}^d \eta_i v_n^p(\boldsymbol{\eta})^{1/2} &\geq \frac{1}{\sqrt{n_p}} \left(\prod_{i=1}^d \eta_i \right)^{1/2} \\ &\geq \frac{\log(n_p)^{3/2}}{n_p}, \end{aligned}$$

implying Equation (4.71). Finally, this ensures that for any $\mathbf{h}, \boldsymbol{\eta} \in \mathcal{H}_n^p$, we have

$$\mathbf{I}_{\mathbf{h}, \boldsymbol{\eta}}^{n,p}(x) \lesssim \frac{n_p^{1-\bar{\omega}\beta/32}}{T_n} + \frac{n_p^{1/2-\bar{\omega}^{1/2}\beta/\tau}}{\sqrt{T_n}}. \quad (4.72)$$

For the term $\mathbf{II}_{\mathbf{h}, \boldsymbol{\eta}}^{n,p}(x)$. Observe that such a term can be treated exactly as the previous one noting that

$$\|\mathbf{K}_{\mathbf{h}} * \mathbf{K}_{\boldsymbol{\eta}}\|_{\infty} \leq \|\mathbf{K}_{\boldsymbol{\eta}}\|_{\infty} \|\mathbf{K}_{\mathbf{h}}\|_1 \quad \text{and} \quad \|\mathbf{K}_{\mathbf{h}} * \mathbf{K}_{\boldsymbol{\eta}}\|_1 \leq \|\mathbf{K}_{\mathbf{h}}\|_1^2.$$

In fact, following the proofs of Lemmas 4.26, 4.27, 4.28 and 4.29, we get that for all $\varepsilon > 0$

$$\mathbb{P}_{\bar{\mu}^b}^b \left(\left| \hat{\mu}_{n,(\mathbf{h}, \boldsymbol{\eta}),p}(x) - \mathbb{E}_{\bar{\mu}^b}^b [\hat{\mu}_{n,(\mathbf{h}, \boldsymbol{\eta}),p}(x)] \right| > \varepsilon \right) \leq K \exp \left(- \frac{n_p \varepsilon^2 \beta^2}{32v^2(\alpha, n, \boldsymbol{\eta}, p) + \tau\beta\varepsilon \|\mathbf{K}_{\boldsymbol{\eta}}\|_{\infty} \log n_p} \right).$$

Finally, we obtain the same bound for $\mathbf{II}_{\mathbf{h}, \boldsymbol{\eta}}^{n,p}(x)$,

$$\mathbf{II}_{\mathbf{h}, \boldsymbol{\eta}}^{n,p}(x) \lesssim \frac{n_p^{1-\bar{\omega}\beta/32}}{T_n} + \frac{n_p^{1/2-\bar{\omega}^{1/2}\beta/\tau}}{\sqrt{T_n}}.$$

Finally, we consider the term $\mathbf{III}_{\mathbf{h}, \boldsymbol{\eta}}^{n,p}(x)$. We recall

$$\mathbf{III}_{\mathbf{h}, \boldsymbol{\eta}}^{n,p}(x) = \left| \mathbb{E}_{\bar{\mu}^b}^b [\hat{\mu}_{n,\boldsymbol{\eta},p}(x)] - \mathbb{E}_{\bar{\mu}^b}^b [\hat{\mu}_{n,(\mathbf{h}, \boldsymbol{\eta}),p}(x)] \right|^2.$$

In following, let us denote $\mu_{\boldsymbol{\eta}}(x) = \mathbb{E}_{\bar{\mu}^b}^b [\hat{\mu}_{n,\boldsymbol{\eta},p}(x)]$ and $\mu_{(\mathbf{h}, \boldsymbol{\eta})}(x) = \mathbb{E}_{\bar{\mu}^b}^b [\hat{\mu}_{n,(\mathbf{h}, \boldsymbol{\eta}),p}(x)]$. Moreover, recalling that $\tilde{\tau}_{n,p}$ is defined in Equation (4.15) and denoting by $\bar{\rho}_{n,p}$ the law of the preaveraged process

defined in Equation (4.12) under $\mathbb{P}_{\bar{\mu}^b}^b$, we write

$$\begin{aligned}
 \mu_{(\mathbf{h}, \boldsymbol{\eta})}(x) &= \sum_{\gamma} \mathbf{u}_{\gamma} \mathbb{E}_{\bar{\mu}^b}^b [\hat{\nu}_{n,(\mathbf{h}, \boldsymbol{\eta}),p}(x + \gamma \tilde{\tau}_{n,p})] \\
 &= \sum_{\gamma} \mathbf{u}_{\gamma} \int_{\mathbb{R}^d} \mathbf{K}_{\mathbf{h}} * \mathbf{K}_{\boldsymbol{\eta}}(x + \gamma \tilde{\tau}_{n,p} - y) \bar{\rho}_{n,p}(dy) \\
 &= \sum_{\gamma} \mathbf{u}_{\gamma} ((\mathbf{K}_{\mathbf{h}} * \mathbf{K}_{\boldsymbol{\eta}}) * \bar{\rho}_{n,p})(x + \gamma \tilde{\tau}_{n,p}) \\
 &= \mathbf{K}_{\boldsymbol{\eta}} * \left(\sum_{\gamma} \mathbf{u}_{\gamma} \mathbf{K}_{\mathbf{h}} * \bar{\rho}_{n,p}(\cdot + \gamma \tilde{\tau}_{n,p}) \right) (x) \\
 &= \mathbf{K}_{\boldsymbol{\eta}} * \mu_{\mathbf{h}}(x).
 \end{aligned} \tag{4.73}$$

Now, let us perform the same decomposition as in Section 9 (see more precisely Equations (4.43) and (4.44)), and write

$$\begin{aligned}
 \mu_{\boldsymbol{\eta}}(x) &= \sum_{\gamma} \mathbf{u}_{\gamma} \mathbb{E}_{\bar{\mu}^b}^b [\hat{\nu}_{n,\boldsymbol{\eta},p}(x + \gamma \tilde{\tau}_{n,p})] \\
 &= \sum_{\gamma} \mathbf{u}_{\gamma} B_1(x + \gamma \tilde{\tau}_{n,p}) + \sum_{\gamma} \mathbf{u}_{\gamma} B_2(x + \gamma \tilde{\tau}_{n,p}).
 \end{aligned}$$

For the second term on the right hand side, we obtain thanks to Equation (4.49) that

$$\sum_{\gamma} \mathbf{u}_{\gamma} B_2(x + \gamma \tilde{\tau}_{n,p}) \lesssim \sqrt{p \Delta_n}. \tag{4.74}$$

Moreover,

$$\begin{aligned}
 \sum_{\gamma} \mathbf{u}_{\gamma} B_1(x + \gamma \tilde{\tau}_{n,p}) &= \sum_{\gamma} \mathbf{u}_{\gamma} \mathbf{K}_{\boldsymbol{\eta}} * (\bar{\mu}^b * \varphi_{\tilde{\tau}_{n,p}})(x + \gamma \tilde{\tau}_{n,p}) \\
 &= \mathbf{K}_{\boldsymbol{\eta}} * \left(\sum_{\gamma} \mathbf{u}_{\gamma} (\bar{\mu}^b * \varphi_{\tilde{\tau}_{n,p}})(\cdot + \gamma \tilde{\tau}_{n,p}) \right) (x).
 \end{aligned}$$

Moreover, it is easy to see using the definition of $(\mathbf{u}_{\gamma})_{\gamma}$, that for any $z \in \mathbb{R}^d$,

$$\left| \sum_{\gamma} \mathbf{u}_{\gamma} (\bar{\mu}^b * \varphi_{\tilde{\tau}_{n,p}})(z + \gamma \tilde{\tau}_{n,p}) - \bar{\mu}^b(z) \right| \lesssim \frac{\tau_n^{2\alpha_1}}{p^{\alpha_1}}.$$

Then, we can write

$$\mu_{\boldsymbol{\eta}}(x) = \mathbf{K}_{\boldsymbol{\eta}} * \bar{\mu}^b(x) + \varepsilon_{n,p}(x), \tag{4.75}$$

for some function $\varepsilon_{n,p}$ such that for all $x \in \mathbb{R}^d$, $|\varepsilon_{n,p}(x)| \lesssim \sqrt{p \Delta_n} + \tau_n^{2\alpha_1}/p^{\alpha_1}$. Then, combining Equations (4.73) and (4.75), we obtain that for all $x \in \mathbb{R}^d$,

$$\mathbb{I}_{\mathbf{h}, \boldsymbol{\eta}}^{n,p}(x) \lesssim B_{n,\mathbf{h},p}(x)^2 + p \Delta_n + \frac{\tau_n^{2\alpha_1}}{p^{\alpha_1}}.$$

Finally, using $\#\mathcal{H}_n^p \lesssim T_n$, we obtain that

$$\begin{aligned} \sup_{\eta \in \mathcal{H}_n^p} \left\{ \left| \widehat{\mu}_{n,(\mathbf{h},\eta),p}(x) - \widehat{\mu}_{n,\eta,p}(x) \right|^2 - V_n^p(\boldsymbol{\eta}) \right\}_+ &\lesssim B_{n,\mathbf{h},p}(x)^2 + n_p^{1-\bar{\omega}\beta/32} \\ &+ \sqrt{T_n} n_p^{1/2-\bar{\omega}^{1/2}\beta/\tau} + p\Delta_n + \frac{\tau_n^{2\alpha_1}}{p^{\alpha_1}}. \end{aligned}$$

Finally, using the fact that $T_n \leq n_p$ and taking $\bar{\omega} > 2\tau^2/\beta^2 \vee 64/\beta$, we get the expected result. \square

We are now ready to move on to the proof of Proposition 4.14.

Proof of Proposition 4.14. Let us consider $\mathbf{h} \in \mathcal{H}_n^p$, then one can write

$$\begin{aligned} \mathbb{E}_{\bar{\mu}^b}^b \left[\left| \widehat{\mu}_{n,\mathbf{h}^*,p}(x) - \bar{\mu}^b(x) \right|^2 \right] &\lesssim \mathbb{E}_{\bar{\mu}^b}^b \left[\left| \widehat{\mu}_{n,\mathbf{h},p}(x) - \bar{\mu}^b(x) \right|^2 \right] \\ &+ \mathbb{E}_{\bar{\mu}^b}^b \left[\left| \widehat{\mu}_{n,\mathbf{h}^*,p}(x) - \widehat{\mu}_{n,\mathbf{h},p}(x) \right|^2 \right]. \end{aligned}$$

We know that the first term on the right-hand side can be controlled in the following way

$$\mathbb{E}_{\bar{\mu}^b}^b \left[\left| \widehat{\mu}_{n,\mathbf{h},p}(x) - \bar{\mu}^b(x) \right|^2 \right] \leq B_{n,\mathbf{h},p}(x)^2 + V_n^p(\mathbf{h}).$$

For the second term, we obtain

$$\begin{aligned} \mathbb{E}_{\bar{\mu}^b}^b \left[\left| \widehat{\mu}_{n,\mathbf{h}^*,p}(x) - \widehat{\mu}_{n,\mathbf{h},p}(x) \right|^2 \right] &\lesssim \mathbb{E}_{\bar{\mu}^b}^b \left[\left| \widehat{\mu}_{n,\mathbf{h}^*,p}(x) - \widehat{\mu}_{n,(\mathbf{h}^*,\mathbf{h}),p}(x) \right|^2 \right] \\ &+ \mathbb{E}_{\bar{\mu}^b}^b \left[\left| \widehat{\mu}_{n,\mathbf{h},p}(x) - \widehat{\mu}_{n,(\mathbf{h}^*,\mathbf{h}),p}(x) \right|^2 \right]. \end{aligned}$$

This finally gives

$$\mathbb{E}_{\bar{\mu}^b}^b \left[\left| \widehat{\mu}_{n,\mathbf{h}^*,p}(x) - \bar{\mu}^b(x) \right|^2 \right] \lesssim \mathbb{E}_{\bar{\mu}^b}^b [A_n^p(\mathbf{h})] + V_n^p(\mathbf{h}) + \mathbb{E}_{\bar{\mu}^b}^b \left[\left| \widehat{\mu}_{n,\mathbf{h},p}(x) - \bar{\mu}^b(x) \right|^2 \right].$$

Moreover, using the result of Lemma 4.33, we obtain

$$\mathbb{E}_{\bar{\mu}^b}^b \left[\left| \widehat{\mu}_{n,\mathbf{h}^*,p}(x) - \bar{\mu}^b(x) \right|^2 \right] \lesssim \inf_{\mathbf{h} \in \mathcal{H}_n^p} \{B_{n,\mathbf{h},p}(x)^2 + V_n^p(\mathbf{h})\} + n^{-\gamma} + p\Delta_n + \tau_n^{2\alpha_1}/p^{\alpha_1},$$

for some $\gamma > 1$. \square

14 Proof of the results of Section 7

First recall the following technical lemma which extensively used in the following.

Lemma 4.34. For $a_1, a_2 \in \mathbb{R}^d$, $\nu_1, \nu_2 \geq 0$, such that $\nu_1 + \nu_2 > 0$, we have

$$\int_{\mathbb{R}^d} e^{-\nu_1|x-a_1|^2-\nu_2|x-a_2|^2} dx = \frac{\kappa_1}{(\nu_1 + \nu_2)^{d/2}} e^{-\frac{\nu_1\nu_2}{\nu_1+\nu_2}|a_1-a_2|^2},$$

for some constant $\kappa_1 > 0$ depending only on the dimension d .

14.1 Proof of Lemma 4.17

We first consider the joint density of $(p^{-1} \sum_{k=0}^{p-1} X_{k\Delta_n}, X_{p\Delta_n})$ conditional on $X_0 = x$, denoted $\mathfrak{p}_{p,n}^b(x; \cdot, \cdot)$.

We claim that under the assumptions of Lemma 4.17, there exists constants C_1, λ_1, c_1 and $\eta_1 > 0$ such that if $p\Delta_n \leq \eta_1$, we have for any x, y, z

$$\mathfrak{p}_{p,n}^b(x; y, z) \leq \frac{C_1}{(p\Delta_n)^d} \exp\left(-\lambda_1 \frac{|y-x|^2 + |z-x|^2}{p\Delta_n} + V(x) - V(z)\right) \quad (4.76)$$

Using Lemma 4.34, (4.37) is a mere consequence of (4.76): since $V(z)$ is bounded below, it suffices to integrate (4.76) with respect to z to get (4.37). The rest of this proof is devoted to showing (4.76).

Note first that $\mathfrak{p}_{p,n}^b(x; y, z) = (p\Delta_n)^{-d} \mathfrak{q}_{p,n}^b(x; \sqrt{1/(p\Delta_n)}(y-x), \sqrt{1/(p\Delta_n)}(z-x))$ where $\mathfrak{q}_{p,n}^b(x; \cdot, \cdot)$ is the joint density of $(p^{-1} \sum_{k=0}^{p-1} (X_{k\Delta_n} - x), (p\Delta_n)^{-1/2} (X_{p\Delta_n} - x))$ under the condition that $X_0 = x$. Therefore, it is enough to prove that

$$\mathfrak{q}_{p,n}^b(x; y, z) \leq c_2^{-1} e^{-c_2(y^2+z^2)+V(x)-V(x+(p\Delta_n)^{1/2}z)}, \quad (4.77)$$

for some $c_2 > 0$. The methodology we use to prove (4.77) closely resembles the one of Theorem 4 in [GG08]; we refer to the comments of Lemma 4.17 for more details. We introduce the process $\mathcal{X}_t^{p,n} = (p\Delta_n)^{1/2} (X_{tp\Delta_n} - x)$ which satisfies

$$d\mathcal{X}_t^{p,n} = b_{p,n}(\mathcal{X}_t^{p,n}) dt + dW_t^{p,n},$$

where $W^{p,n}$ is a d -dimensional Brownian motion and $b_{p,n}(w) = (p\Delta_n)^{1/2} b(x + (p\Delta_n)^{1/2} w)$. Moreover,

$$\frac{1}{p} \left(\frac{1}{p\Delta_n}\right)^{1/2} \sum_{k=0}^{p-1} (X_{k\Delta_n} - x) = \frac{1}{p} \sum_{\ell=0}^{p-1} \mathcal{X}_{\ell/p}^{p,n} \quad \text{and} \quad \left(\frac{1}{p\Delta_n}\right)^{1/2} (X_{p\Delta_n} - x) = \mathcal{X}_1^{p,n}.$$

Let us now define the stochastic process

$$(\mathcal{E}_t^{p,n})_{t \geq 0} = \left(\exp\left(-\int_0^t b_{p,n}(\mathcal{X}_s^{p,n}) dW_s^{p,n} - 1/2 \int_0^t |b_{p,n}(\mathcal{X}_s^{p,n})|^2 ds\right) \right)_{t \geq 0}.$$

Under Assumption 7, we get that Novikov criterion holds following the steps of the proof of Lemma 4.21. This ensures that $(\mathcal{E}_t^{p,n})_{t \geq 0}$ is a martingale with constant expectation equal to 1. Then it defines a change of measure and we can consider a probability measure \mathbb{Q}_x^b , under which we get rid of the influence of the drift for the dynamic of $\mathcal{X}^{n,p}$. More precisely, we define the probability measure \mathbb{Q}_x^b on

$\sigma(\{W_t^{p,n}, t \leq 1\})$ by

$$\frac{d\mathbb{P}_x^b}{d\mathbb{P}_x} = \exp\left(-\int_0^1 b_{p,n}(\mathcal{X}_t^{p,n}) dW_t^{p,n} - \frac{1}{2} \int_0^1 |b_{p,n}(\mathcal{X}_t^{p,n})|^2 dt\right).$$

Using the Itô formula, we obtain

$$\begin{aligned} \frac{d\mathbb{P}_x^b}{d\mathbb{Q}_x^b} &= \exp\left(\int_0^1 b_{p,n}(\mathcal{X}_t^{p,n}) d\mathcal{X}_t^{p,n} - \frac{1}{2} \int_0^1 |b_{p,n}(\mathcal{X}_t^{p,n})|^2 dt\right) \\ &= \exp\left(B_{p,n}(\mathcal{X}_t^{p,n}) - \frac{1}{2} \int_0^1 |b_{p,n}(\mathcal{X}_t^{p,n})|^2 + \nabla \cdot b_{p,n}(\mathcal{X}_t^{p,n}) dt\right), \end{aligned}$$

where for all $w \in \mathbb{R}^d$, $B_{p,n}(w) := V(x) - V(x + (p\Delta_n)^{1/2}w)$. Now let f and g be two bounded positive functions. Then we have

$$\begin{aligned} &\mathbb{E}_x\left[f\left(\frac{1}{p} \sum_{\ell=0}^{p-1} \mathcal{X}_{\ell/p}^{p,n}\right)g(\mathcal{X}_1^{p,n})\right] \\ &= \mathbb{E}^{\mathbb{Q}_x^b}\left[f\left(\frac{1}{p} \sum_{\ell=0}^{p-1} \mathcal{X}_{\ell/p}^{p,n}\right)g(\mathcal{X}_1^{p,n}) \exp\left(B_{p,n}(\mathcal{X}_t^{p,n}) - \frac{1}{2} \int_0^1 |b_{p,n}(\mathcal{X}_t^{p,n})|^2 + \nabla \cdot b_{p,n}(\mathcal{X}_t^{p,n}) dt\right)\right]. \end{aligned}$$

Moreover, we have $\nabla \cdot b_{p,n}(w) \leq Cp\Delta_n$. Since Girsanov Theorem ensures that $\mathcal{X}^{p,n}$ is a Brownian motion under \mathbb{Q}_x^b , we get the bound

$$\mathbb{E}_x^b\left[f\left(\frac{1}{p} \sum_{\ell=0}^{p-1} \mathcal{X}_{\ell/p}^{p,n}\right)g(\mathcal{X}_1^{p,n})\right] \leq C_1 \mathbb{E}^b\left[f\left(\frac{1}{p} \sum_{\ell=0}^{p-1} W_{\ell/p}\right)g(W_1)e^{B_{p,n}(W_1)}\right], \quad (4.78)$$

for some Brownian motion W and some constant C_1 , which does not depend on p nor n . Therefore, it is enough to prove that

$$\mathbb{E}^b\left[f\left(\frac{1}{p} \sum_{\ell=0}^{p-1} W_{\ell/p}\right)g(W_1)e^{B_{p,n}(W_1)}\right] \leq C'_1 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(u)g(v)e^{-C'_2(|u|^2+|v|^2)+B_{p,n}(v)} du dv \quad (4.79)$$

holds for some positive constants C'_1 and C'_2 . Again, this is closely related to Lemma 4 of [GG08]. Recall that $\mathcal{W}_t^* = W_t - tW_1$ defines a Brownian bridge on $[0, 1]$, independent of W_1 . Thus, if $h(v) = g(v)e^{B_{p,n}(v)}$, we get:

$$\begin{aligned} \mathbb{E}^b\left[f\left(\frac{1}{p} \sum_{\ell=0}^{p-1} W_{\ell/p}\right)h(W_1)\right] &= \mathbb{E}^b\left[f\left(\frac{1}{p} \sum_{\ell=0}^{p-1} \mathcal{W}_{\ell/p}^* + \frac{\ell}{p}W_1\right)h(W_1)\right] \\ &= \mathbb{E}^b\left[\mathbb{E}^b\left[f\left(\frac{1}{p} \sum_{\ell=0}^{p-1} \mathcal{W}_{\ell/p}^* + \frac{\ell}{p}W_1\right)h(W_1)\middle|W_1\right]\right] \\ &= \mathbb{E}^b\left[\psi_{p,n}^b(W_1)\right], \end{aligned} \quad (4.80)$$

where for all $\omega \in \mathbb{R}^d$,

$$\psi_{p,n}^b(\omega) = \mathbb{E}^b \left[f \left(\frac{1}{p} \sum_{\ell=0}^{p-1} \mathcal{W}_{\ell/p}^* + \frac{\ell}{p} \omega \right) h(\omega) \right].$$

We know that the Brownian Bridge \mathcal{W}^* itself admits the following decomposition, see *e.g.* [GG08]:

$$\mathcal{W}_t^* = \xi \eta_t + \mathcal{W}_t^{**},$$

where ξ is a standard random variable, η is the deterministic function

$$\eta_t = \begin{cases} t & \text{if } t \in [0, 1/2], \\ (1-t) & \text{if } t \in [1/2, 1], \end{cases}$$

and \mathcal{W}^{**} is the process on $[0, 1]$ constructed as the concatenation of two independent Brownian bridges, on $[0, 1/2]$ and $[1/2, 1]$ respectively. Moreover in this decomposition the random variable η and the process \mathcal{W}^{**} are independent. Then,

$$\psi_{p,n}^b(\omega) = h(\omega) \mathbb{E} \left[f \left(\frac{\xi}{p} \sum_{\ell=0}^{p-1} \eta_{\ell/p} + \frac{1}{p} \sum_{\ell=0}^{p-1} \mathcal{W}_{\ell/p}^{**} + \frac{p-1}{2p} \omega \right) \right]. \quad (4.81)$$

Let $c_p = p^{-1} \sum_{\ell=0}^{p-1} \eta_{\ell/p}$, which is bounded uniformly in p . Using the independence between ξ and \mathcal{W}^{**} , we have:

$$\begin{aligned} \mathbb{E}^b \left[\left(\frac{\xi}{p} \sum_{\ell=0}^{p-1} \eta_{\ell/p} + \frac{1}{p} \sum_{\ell=0}^{p-1} \mathcal{W}_{\ell/p}^{**} + \frac{p-1}{2p} \omega \right) \right] &= \mathbb{E}^b \left[\int_{\mathbb{R}^d} f \left(c_p v + \frac{1}{p} \sum_{\ell=0}^{p-1} \mathcal{W}_{\ell/p}^{**} + \frac{p-1}{2p} \omega \right) \frac{1}{(2\pi)^{d/2}} e^{-|v|^2/2} dv \right] \\ &= \mathbb{E}^b \left[\int_{\mathbb{R}^d} f(u) (\sqrt{2\pi} c_p)^{-d} e^{-\frac{1}{2c_p^2} |u - \frac{1}{p} \sum_{\ell=0}^{p-1} \mathcal{W}_{\ell/p}^{**} - \frac{p-1}{2p} \omega|^2} du \right]. \end{aligned}$$

We then use the fact that for $\varepsilon \in (0, 1)$, $|x - y|^2 \geq \varepsilon/(1 + \varepsilon)|x|^2 - \varepsilon|y|^2$, to get

$$\mathbb{E}^b \left[\psi_{p,n}^b(W_1) \right] \leq \int_{\mathbb{R}^d} \frac{1}{(\sqrt{2\pi} c_p)^d} f(u) e^{-\frac{|u|^2 \varepsilon}{2c_p^2(1+\varepsilon)} + \frac{c_p^2 \varepsilon (p-1)^2 |\omega|^2}{4c_p^2 p^2}} \mathbb{E} \left[e^{c_p^{-2} \varepsilon \sup_{t \in [0,1]} |\mathcal{W}_t^{**}|^2} \right]. \quad (4.82)$$

Using that \mathcal{W}^{**} is a concatenation of two Brownian bridges, we know that there exists $\varepsilon_+ > 0$ and C_ε such that for all $\varepsilon \leq \varepsilon_+$,

$$\mathbb{E} \left[e^{c_p^2 \varepsilon \sup_{t \in [0,1]} |\mathcal{W}_t^{**}|^2} \right] \leq C_\varepsilon.$$

Moreover, thanks to the boundedness of c_p with respect to p , we get that C_ε does not depend on p .

Then, plugging (4.82) into Equation (4.80), we get that for $\varepsilon \leq \varepsilon_+$,

$$\begin{aligned} \mathbb{E}^b \left[f \left(\frac{1}{p} \sum_{\ell=0}^{p-1} W_{\ell/p} \right) h(W_1) \right] &\leq \frac{C_\varepsilon}{(\sqrt{2\pi}c_p)^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(u)h(v) e^{-\frac{c_p^2 u^2 \varepsilon}{2(1+\varepsilon)} + \frac{c_p^2 \varepsilon (p-1)^2 |v|^2}{4p^2}} e^{-|v|^2/2} du dv \\ &\leq \frac{C_\varepsilon}{(\sqrt{2\pi}c_p)^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(u)g(v) e^{B_{p,n}(v)} e^{-\frac{u^2 \varepsilon}{2c_p^2(1+\varepsilon)} + \frac{\varepsilon(p-1)^2 |v|^2}{4c_p^2 p^2}} e^{-|v|^2/2} du dv. \end{aligned}$$

Then, we get Equation (4.79) as soon as $\varepsilon < \min(2c_p p^2 / (p-1)^2, \varepsilon_+)$, which concludes the proof.

14.2 Proof of Lemma 4.18

We consider $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ non-negative bounded with compact support. From Girsanov Theorem, we can show that for any $x \in \mathbb{R}^d$,

$$\mathbb{E}_x^b[\varphi(X_t)] \leq (2\pi)^{-d/2} e^{b_1 t/2} \frac{e^{V(x)}}{t^{d/2}} \int_{\mathbb{R}^d} \varphi(y) e^{-\frac{|x-y|^2}{2t} - V(y)} dy$$

and we conclude using $t \leq 1$, and the boundedness of $\bar{\mu}^b$.

14.3 Proof of Corollary 4.19

First note that since X is a Markov process, we have

$$\mathfrak{p}_{p,n,t}^b(x; y) = \int_{\mathbb{R}^d} p_t^b(x; z) \mathfrak{p}_{p,n}^b(z; y) dz.$$

Using the bounds given by Lemmas 4.17 and 4.18, combined with Lemma 4.34, we get

$$\mathfrak{p}_{p,n,t}^b(x; y) \lesssim (p\Delta_n + 2t\lambda_1)^{-d/2} \exp\left(\frac{\lambda_1}{p\Delta_n + 2t} |x - y|^2 + V(x)\right).$$

14.4 Proof of Lemma 4.20

Note that since the distribution of X_0 is $\bar{\mu}^b$ which is the invariant measure of X , the distribution of $(p^{-1} \sum_{\ell=0}^{p-1} X_{\ell\Delta_n}, X_{p\Delta_n})$ is $\bar{\pi}^b$. Thus for any non negative function $\varphi : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$, we have

$$\begin{aligned} \mathbb{E}_{\bar{\mu}^b}^b \left[\varphi \left(\frac{1}{p} \sum_{\ell=0}^{p-1} X_{\ell\Delta_n}, X_{p\Delta_n} \right) \right] &= \int_{\mathbb{R}^d} \mathbb{E}_x^b \left[\varphi \left(\frac{1}{p} \sum_{\ell=0}^{p-1} X_{\ell\Delta_n}, X_{p\Delta_n} \right) \right] \bar{\mu}^b(x) dx \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(y, z) \mathfrak{p}_{p,n}^b(x; y, z) \bar{\mu}^b(x) dx dz dy. \end{aligned}$$

From Equation (4.76), we get

$$\begin{aligned}
 & \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(y, z) \mathfrak{p}_{p,n}^b(x; y, z) \bar{\mu}^b(x) dx dz dy \\
 & \leq \frac{C_1}{(p\Delta_n)^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(y, z) \exp\left(-\lambda_1 \frac{|y-z|^2 + |z-x|^2}{p\Delta_n}\right) e^{-V(z)+V(x)} \frac{e^{-2V(x)}}{Z_V} dx dz dy \\
 & \leq \frac{C_1 \|\bar{\mu}^b\|_\infty^{1/2}}{Z_V^{1/2} (p\Delta_n)^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(y, z) e^{-V(z)} \left(\int_{\mathbb{R}^d} \exp\left(-\frac{\lambda_1}{p\Delta_n} (|y-x|^2 + |z-x|^2)\right) dx \right) dy dz.
 \end{aligned}$$

Use Lemma 4.34 to control the integral in the variable x , and get

$$\begin{aligned}
 & \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(y, z) p_{p,n}(x; y, z) \bar{\mu}^b(x) dx dz dy \\
 & \leq \frac{C_1 \kappa_1 \|\bar{\mu}^b\|_\infty^{1/2}}{Z_V^{1/2} (p\Delta_n)^{d/2}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(y, z) e^{-V(z)} \exp\left(-\frac{\lambda_1}{2p\Delta_n} |y-z|^2\right) dy dz.
 \end{aligned}$$

We can conclude that for all $(y, z) \in \mathbb{R}^{2d}$, we have

$$\bar{\pi}^b(y, z) \leq \frac{C_1 \kappa_1 \|\bar{\mu}^b\|_\infty^{1/2}}{Z_V^{1/2} (p\Delta_n)^{d/2}} \exp\left(-\frac{\lambda_1}{2p\Delta_n} |y-z|^2 - V(z)\right).$$

Appendices

Appendix **A**

Reflection coupling and convergence toward equilibrium

1 Probabilistic context

We consider a differentiable function V with Lipschitz-continuous gradient and the diffusion process

$$dX_t = -\nabla V(X_t)dt + \sigma dB_t \quad (\text{A.1})$$

It is well known that when the potential V is uniformly convex, it is possible to obtain convergence rates with synchronized couplings. Indeed, for two solutions X and Y of the equation (A.1), with the same Brownian motion B but different initial conditions μ_0, ν_0 , we can write:

$$\begin{aligned} d|X_t - Y_t|^2 &= -2(X_t - Y_t) \cdot (\nabla V(X_t) - \nabla V(Y_t))dt \\ &\leq -2\beta|X_t - Y_t|^2, \end{aligned} \quad (\text{A.2})$$

where $\beta > 0$ is such that $\nabla^2 V \geq \beta$. Thus, it directly follows that

$$d_2^{\mathbb{R}^d}(\mu_t, \nu_t) \leq \mathbb{E}[|X_t - Y_t|^2]^{1/2} \leq d_2^{\mathbb{R}^d}(\mu_0, \nu_0)e^{-\beta t}. \quad (\text{A.3})$$

When V is not assumed to be uniformly convex, we need somewhat finer arguments since the calculations of equation (A.2) no longer hold. In [Ebe16], the author proves similar results as in Equation (A.3) but in a Wasserstein-1 distance. More precisely, the idea of [Ebe16] is to consider the distance $d_f^{\mathbb{R}^d}$ defined on \mathbb{R}^d by $d_f(x, y) = f(|x - y|)$, and then the associated distance on $\mathcal{P}_1(\mathbb{R}^d)$,

$$d_f^{\mathbb{R}^d}(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d} d_f(x, y)\pi(dx, dy).$$

2 The coupling method

The idea of the coupling is the following: A coupling by reflection of two solutions of (A.1) with initial distributions μ and ν is a diffusion process (X_t, Y_t) with values in \mathbb{R}^{2d} defined by $L(X_0, Y_0) = \eta$ where η is a coupling of μ and ν , and with dynamic

$$\begin{aligned} dX_t &= -\nabla V(X_t) dt + \sigma dB_t \quad \text{for } t \geq 0, \\ dY_t &= -\nabla V(Y_t) dt + \sigma (\text{Id} - 2e_t e_t^\top) dB_t \quad \text{for } t < T, \quad Y_t = X_t \text{ for } t \geq T. \end{aligned} \quad (\text{A.4})$$

where

$$e_t := (X_t - Y_t) / |X_t - Y_t|,$$

and $T := \inf \{t \geq 0 : X_t = Y_t\}$ is the coupling time. As previously mentioned, this type of coupling was introduced by [LR86]. The reflection coupling can be realized as a diffusion process in \mathbb{R}^{2d} , and the marginal processes $(X_t)_{t \geq 0}$ and $(Y_t)_{t \geq 0}$ are solutions of with Brownian motions B_t and

$$\tilde{B}_t = \int_0^t (I_d - 2\mathbb{I}_{\{s < T\}} e_s e_s^\top) dB_s,$$

respectively. By Lévy's characterization, \tilde{B} is indeed a Brownian motion. The difference vector

$$E_t := X_t - Y_t$$

solves the one-dimensional SDE.

$$\begin{aligned} dE_t &= -(\nabla V(X_t) - \nabla V(Y_t)) dt + 2\sigma dW_t \quad \text{for } t < T \\ E_t &= 0 \quad \text{for } t \geq T \end{aligned}$$

with

$$W_t = \int_0^t e_s^\top dB_s.$$

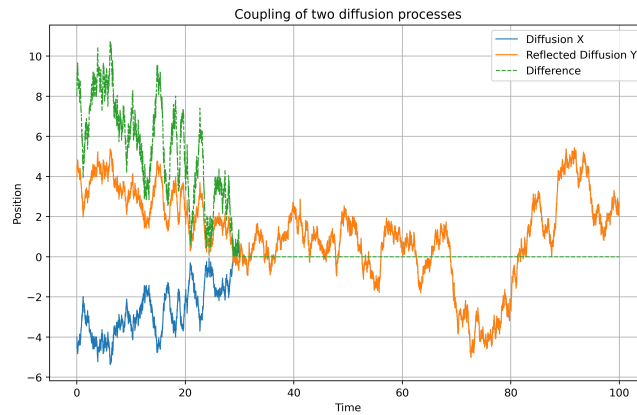


Figure A.1: Realisation of a coupling by reflection in dimension $d = 1$.

The main point of this coupling method is to preserve the noise whenever the particles are *not close*.

Thinking once again of a double-well potential $V : x \mapsto |x|^4/4 - |x|^2/2$, we typically want to avoid the situation where X_t is around -1 and Y_t is around 1, for the same time $t > 0$. To prevent an occurrence of this situation we keep track of the noise whenever this happens. Once the particles touch each other, they stay together. The first observation here is that considering the fact that particles does not touch before the coupling time T , we can prove (see [Ebe16]) that for all $t \geq T$,

$$d|X_t - Y_t| = 2\sigma dW_t + e_t \cdot (b(X_t) - b(Y_t))dt$$

This coupling allows for the presence of noise when the particles are far apart, enabling contraction properties to be achieved. Unlike the convex case, these properties cannot be obtained in $d_2^{\mathbb{R}^d}(\cdot, \cdot)$. Specifically, the idea developed in [Ebe16] is to consider a slight deformation of the Wasserstein-1 distance that offers contraction properties. In order to do so, we look for a function f such that there exists $c > 0$ which makes $(e^{ct}f(|E_t|))_{t \geq 0}$ is a super-martingale. In [Ebe16], the author construct a function f satisfying

$$f''(r) - \frac{1}{4}r\kappa(r)f'(r) \leq -\frac{c}{2\sigma^2}f(r) \quad \text{for a.e. } r > 0. \quad (\text{A.5})$$

From this, we obtain that

$$\frac{d}{dt}\mathbb{E}[f(|X_t - Y_t|)] \leq -c\mathbb{E}[f(|X_t - Y_t|)] \quad (\text{A.6})$$

and conclude using Gronwall Lemma.

Remark A.1. *On the choice of the function f :*

1. Choosing $f(x) = x$ yields the standard L^1 -Wasserstein distance $d_f^{\mathbb{R}^d} = d_1^{\mathbb{R}^d}$. In this case it is well known that the transition kernels $p_t(x, dy)$ of the diffusion process (X_t) satisfies

$$d_1^{\mathbb{R}^d}(\mu p_t, \nu p_t) \leq e^{-\beta t/2} d_1^{\mathbb{R}^d}(\mu, \nu) \quad \text{for any } \mu, \nu \quad \text{and } t \geq 0,$$

provided $\nabla^2 V \geq \beta I_d$ holds globally.

2. On the other hand, choosing $f(x) = \mathbb{I}_{(0, \infty)}(x)$ yields the total variation distance $d_f^{\mathbb{R}^d} = d_{TV}^{\mathbb{R}^d}$. In this case,

$$d_f^{\mathbb{R}^d}(\mu p_t, \nu p_t) \leq \mathbb{P}[T > t] \quad \text{for any } \mu, \nu \quad \text{and } t \geq 0,$$

where T stands for the coupling time between two realisations of X with initial conditions μ and ν respectively. There is no strict contractivity of p_t w.r.t. $d_{TV}^{\mathbb{R}^d}$ in general. Indeed, in many applications $d_{TV}^{\mathbb{R}^d}(\mu p_t, \nu p_t)$ only decreases after a certain amount of time.

3. The idea of [Ebe16] is then to find a modification of the previous cases which allows to get contractivity.

Appendix **B**

Goldenshulger-Lepski adaptive procedure

1 Framework and Kernel estimator

We consider X^1, \dots, X^n random variables in \mathbb{R}^d admitting a common density with respect to the Lebesgue measure, denoted by f . We wish to estimate $f(x)$ for $x \in \mathbb{R}^d$ based on the observations.

A natural initial intuition can be formulated as follows: the more irregular the function to be estimated, the more observations are needed to estimate it accurately. A single observation is sufficient to estimate a constant function, whereas a function with discontinuities requires many more observations. The appropriate way to measure the regularity of a function in this context is as follows (see also Section ...):

Definition B.1. Let $\alpha = (\alpha_1, \dots, \alpha_d) \in (0, \infty)^d$ and $\mathcal{L} = (\mathcal{L}_1, \dots, \mathcal{L}_d) \in (0, \infty)^d$. A function $g : \mathbb{R}^d \rightarrow \mathbb{R}$ is said to belong to the anisotropic Hölder class $\mathcal{H}_d(\alpha, \mathcal{L})$ of functions if, for all $1 \leq i \leq d$, g is $\lfloor \alpha_i \rfloor$ -differentiable in the i -th variable and the partial derivatives satisfy for $0 \leq k \leq \lfloor \alpha_i \rfloor$

$$\|\partial_i^k g\|_\infty \leq \mathcal{L}_i \quad \text{and} \quad \forall t \in \mathbb{R}, \|\partial_i^{\lfloor \alpha_i \rfloor} g(\cdot + te_i) - \partial_i^{\lfloor \alpha_i \rfloor} g(\cdot)\|_\infty \leq \mathcal{L}_i |t|^{\alpha_i - \lfloor \alpha_i \rfloor}$$

where (e_1, \dots, e_d) is the canonical basis of \mathbb{R}^d .

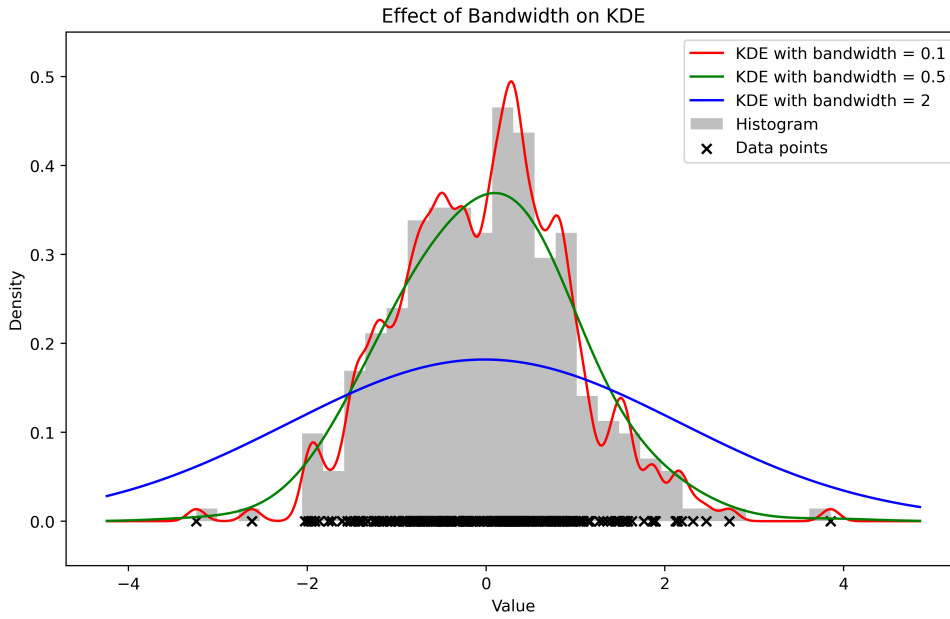
A natural estimator in this kind of situation is called a kernel estimator [?, ?, ?]. We consider a function $K : \mathbb{R} \rightarrow \mathbb{R}$ with certain properties that will be detailed later. We consider a bandwidth $\mathbf{h} = (h_1, \dots, h_d) \in (0, +\infty)^d$ and define

$$\hat{f}_h(x) = \frac{1}{n} \sum_{k=1}^n \mathbf{K}_h(x - X^k), \quad (\text{B.1})$$

where for any $y \in \mathbb{R}^d$,

$$\mathbf{K}_h(y) = \prod_{i=1}^d h_i^{-1} K(y_i/h_i). \quad (\text{B.2})$$

For any $i \in \{1, \dots, d\}$, h_i gives the precision with which we look in the i -th direction. The following graphic represents the impact of the choice of the bandwidth in dimension 1.


 Figure B.1: Role of the bandwidth in dimension $d = 1$.

The choice of bandwidth, h , is crucial in non-parametric estimation because it directly affects the bias-variance trade-off in the estimator \hat{f} . If the bandwidth is too small, the estimator captures too much noise, leading to high variance and an overfitting problem where the estimate is too sensitive to the random fluctuations in the data. On the other hand, if the bandwidth is too large, the estimator smooths out important features of the underlying function, resulting in high bias and an underfitting problem where the estimate fails to capture the true structure of the data. Therefore, selecting an appropriate bandwidth is essential to achieve a good balance between bias and variance, ensuring that the estimator is both accurate and reliable.

In this section, we provide an explicit procedure to control the pointwise quadratic loss for the kernel estimator:

$$\mathcal{R}_{n,h} = \mathbb{E}[|\hat{f}_h(x) - f(x)|^2]. \quad (\text{B.3})$$

Of course, this control will depend on h and α . In a first time, let us assume that α is known to make things clear.

1.1 Case of a known regularity

Let us assume in this Section that $f \in \mathcal{H}(\alpha, \mathcal{L})$ and that $\alpha = (\alpha_1, \dots, \alpha_d)$ is known. Then, we perform the following decomposition

$$\hat{f}_h(x) - f(x) = \underbrace{\hat{f}_h(x) - \mathbb{E}[\hat{f}_h(x)]}_{\text{centred stochastic term}} + \underbrace{\mathbb{E}[\hat{f}_h(x)] - f(x)}_{\text{bias}}.$$

Then, we mainly have to control the bias and variance of the estimator. This is called bias-variance decomposition of the loss. Let us begin with the variance

$$\begin{aligned}\mathbb{E}[|\hat{f}_h(x) - \mathbb{E}[\hat{f}_h(x)]|^2] &= \frac{1}{n} \text{Var}[\mathbf{K}_h(x - X^1)] \\ &\leq \frac{1}{n} \mathbb{E}[|\mathbf{K}_h(x - X^1)|^2] \\ &= \frac{1}{n} \int_{\mathbb{R}^d} |\mathbf{K}_h(x - y)|^2 f(y) dy \\ &\leq \frac{\|f\|_\infty \|\mathbf{K}\|_2^{2d}}{n \prod_{i=1}^d h_i}.\end{aligned}$$

Now, for the bias term, let us begin by noting that

$$\begin{aligned}\mathbb{E}[\hat{f}_h(x)] &= \int_{\mathbb{R}^d} \mathbf{K}_h(x - y) f(y) dy \\ &= \int_{\mathbb{R}^d} \mathbf{K}(x) f(x + \mathbf{h}y) dy \\ &\lesssim \sum_{i=1}^d h_i^{\alpha_i}.\end{aligned}$$

Finally, we obtain

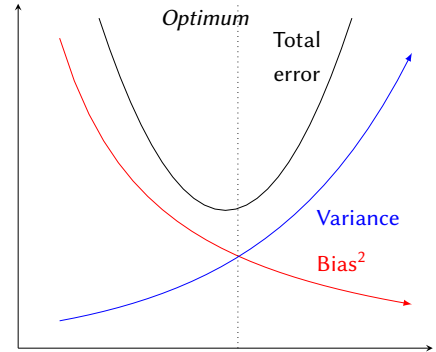
$$\mathbb{E}[|\hat{f}_h(x) - f(x)|^2] \lesssim \sum_{i=1}^d h_i^{2\alpha_i} + \frac{1}{n \prod_{i=1}^d h_i}. \quad (\text{B.4})$$

To find the optimal bandwidth, we need to balance the trade-off between bias and variance.

The optimal bandwidth is given by $h_i^* = n^{-\frac{\alpha_i}{\bar{\alpha}(2\bar{\alpha}+d)}}$, where the effective smoothness $\bar{\alpha}$ is defined by

$$\bar{\alpha}^{-1} := \frac{1}{d} \sum_{i=1}^d \frac{1}{\alpha_i}.$$

This formula makes sense because it matches our intuition: the more irregular the function f is in direction i , the smaller h_i^* needs to be to estimate f accurately.



Finally, plugging the optimal bandwidth into (B.4), we obtain

$$\mathbb{E}[|\hat{f}_{h^*}(x) - f(x)|^2] \lesssim n^{-\frac{2\bar{\alpha}}{\bar{\alpha}+d}} \quad (\text{B.5})$$

Remark B.2. Some remarks are in order: (1) The constant hidden under the sign ' \lesssim ' is uniform over all the Hölder class $\mathcal{H}(\alpha, \mathcal{L})$. (2) The rate of estimation is very common in non-parametric estimation [?]. We also see what is often called the curse of dimensionality because the presence of d makes the estimation

procedure very slow in high dimensions. Some new techniques use the geometric distribution of the data to reduce the effective dimension and improve the speed [?, ?] to name just a few. **(3)** The main difficulty here is that the optimal bandwidth depends on the regularity of the function f that is unknown a priori. The next section is dedicated to an effective procedure to choose a good bandwidth.

1.2 Case of an unknown regularity

Before delving into the section, let us make a quick recap of concentration inequalities

Proposition B.3 (Bernstein inequality). *For any sequence Z_1, \dots, Z_N of real-valued independent random variables bounded by some constant $M > 0$ and such that $\mathbb{E}[Z_i] = 0$, we have*

$$\mathbb{P}\left(\sum_{i=1}^N Z_i \geq y\right) \leq \exp\left(-\frac{t^2}{2\left(\sum_{i=1}^N \mathbb{E}[Z_i^2] + \frac{Mt}{3}\right)}\right),$$

for any $t > 0$.

Our goal is to define a choice $\hat{\mathbf{h}}$ of parameter such that the estimator $\hat{f}_{\hat{\mathbf{h}}}$ satisfies an oracle inequality:

$$\mathbb{E}\left[|\hat{f}_{\hat{\mathbf{h}}}(x) - f(x)|^2\right] \leq C \inf_{\mathbf{h} \in (0,1)^d} \mathbb{E}\left[|\hat{f}_{\mathbf{h}}(x) - f(x)|^2\right] + R_n. \quad (\text{B.6})$$

There are several ways ..., let us present the one developed by GL [?] Let us now move on to explain the Glodenshulger-Lepski procedure. (Un petit point historique)

1.3 The Goldenshulger-Lepski procedure

Let us introduce some notation:

- $\mathcal{H}_n \subset (0, 1]^d$ is a mesh of possible bandwidth;
- For any $\mathbf{h}, \boldsymbol{\eta} \in \mathcal{H}_n$, we define $B_{\mathbf{h}}(\boldsymbol{\eta}) := \left\{ \left| \hat{f}_{\boldsymbol{\eta}}(x) - \hat{f}_{\mathbf{h}, \boldsymbol{\eta}}(x) \right| - V(\boldsymbol{\eta}) \right\}_+$;
- For any $\mathbf{h} \in \mathcal{H}_n$, we define $B_{\mathbf{h}} = \max_{\boldsymbol{\eta} \in \mathcal{H}_n} B_{\mathbf{h}}(\boldsymbol{\eta})$;
- For any $\mathbf{h} \in \mathcal{H}_n$, we set $V_{\mathbf{h}} = \bar{\omega} \log(n) n^{-1} \prod_{i=1}^d h_i^{-1}$, for some constant $\bar{\omega} > 0$.

What do we have to do to make this procedure work?

1. Find a mesh \mathcal{H}_n with some theoretical guarantees.
2. Find $\bar{\omega}$ in the definition of the penalization $V_{\mathbf{h}}$.

1.3.1 Heuristic idea

The difficulty here is that we cannot access the bias of the estimator: $|f(x) - \mathbf{K}_{\mathbf{h}} * f(x)|$. This definition is as follows: to approximate the bias $|f(x) - f * K_{\mathbf{h}}(x)|$, we replace f with an estimator with a fixed window, $\hat{f}_{\boldsymbol{\eta}}$. This leads to $\left| \hat{f}_{\boldsymbol{\eta}}(x) - \mathbf{K}_{\mathbf{h}} * \hat{f}_{\boldsymbol{\eta}}(x) \right| = \left| \hat{f}_{\boldsymbol{\eta}}(x) - \hat{f}_{\mathbf{h}, \boldsymbol{\eta}}(x) \right|$. But, unlike the bias, this quantity

"contains" randomness and thus variability: to correct this, it is necessary to subtract the corresponding "part of variance" V_η . Finally, since there is no reason to choose one window η over another, we sweep through the entire collection. Of course, this is only a "hand-waving" justification, and it can be rigorously demonstrated that B_h has the same order of magnitude as the bias,

$$B_h \leq C |f(x) - f \star K_h(x)|.$$

1.4 The procedure

Now that we have defined an estimator of the squared bias, and a proxy of the variance, we consider the following bandwidth choice procedure:

$$\hat{h} := \arg \min_{h \in \mathcal{H}_n} \{B_h + V_h\},$$

and we define the Goldenshulger-Lepski estimator $\hat{f}_{GL}(x) := \hat{f}_{\hat{h}}(x)$. This procedure offers the following technical guarantee:

Proposition B.4. *There exists $\underline{\omega} > 0$ such that for all $\bar{\omega} > \underline{\omega}$,*

$$\mathbb{E} \left[|\hat{f}_{GL}(x) - f(x)|^2 \right] \leq C \inf_{h \in (0,1]^d} \mathbb{E} \left[|\hat{f}_h(x) - f(x)|^2 \right] + n^{-1}, \quad (\text{B.7})$$

for some $C \geq 1$.

Proof of Proposition B.4. Step 1. Let us begin this proof with the following decomposition which stands for any $h \in (0, 1]^d$,

$$\begin{aligned} |\hat{f}_{GL}(x) - f(x)|^2 &\lesssim \left\{ \hat{f}_{\hat{h}}(x) - \hat{f}_{h, \hat{h}}(x) - V_{\hat{h}} \right\}_+ + V_{\hat{h}} \\ &\quad + \left\{ \hat{f}_{h, \hat{h}}(x) - \hat{f}_h(x) - V_h \right\}_+ + V_h \\ &\quad + |\hat{f}_h(x) - f(x)|^2 \\ &\lesssim B_h + V_{\hat{h}} + B_{\hat{h}} + V_h + |\hat{f}_h(x) - f(x)|^2 \\ &\lesssim B_h + V_h + |\hat{f}_h(x) - f(x)|^2, \end{aligned} \quad (\text{B.8})$$

where the last line is straightforward as \hat{h} minimizes $h \mapsto B_h + V_h$. Now, we are really happy the the last term because as done in the case with known regularity, we obtain

$$\mathbb{E}[|\hat{f}_h(x) - f(x)|^2] \lesssim \sum_{i=1}^d h_i^{2\alpha_i} + \frac{1}{n \prod_{i=1}^d h_i}.$$

The term V_h is deterministic, so we essentially need to control $\mathbb{E}[B_h]$. Recalling that $B_h = \max_{\eta \in \mathcal{H}_n} B_h(\eta)$,

we consider $\boldsymbol{\eta} \in \mathcal{H}_n$ and write

$$\begin{aligned} B_{\mathbf{h}}(\boldsymbol{\eta}) &\lesssim \left\{ |\widehat{f}_{\boldsymbol{\eta}}(x) - f_{\boldsymbol{\eta}}(x)|^2 - V_{\boldsymbol{\eta}} \right\}_+ \\ &\quad + \left\{ |\widehat{f}_{\boldsymbol{\eta},\mathbf{h}}(x) - f_{\boldsymbol{\eta},\mathbf{h}}(x)|^2 - V_{\boldsymbol{\eta}} \right\}_+ \\ &\quad + |f_{\boldsymbol{\eta},\mathbf{h}}(x) - f_{\boldsymbol{\eta}}(x)|^2, \end{aligned}$$

where $f_{\boldsymbol{\eta},\mathbf{h}}(x) = \mathbb{E}[\widehat{f}_{\boldsymbol{\eta},\mathbf{h}}(x)]$ and $f_{\boldsymbol{\eta}}(x) = \mathbb{E}[\widehat{f}_{\boldsymbol{\eta}}(x)]$.

Step 2. Difference of the bias. Let us notice that

$$\begin{aligned} |f_{\boldsymbol{\eta},\mathbf{h}}(x) - f_{\boldsymbol{\eta}}(x)|^2 &= |\mathbf{K}_{\boldsymbol{\eta}} * f(x) - \mathbf{K}_{\boldsymbol{\eta}} * \mathbf{K}_{\mathbf{h}} * f(x)|^2 \\ &\leq \|\mathbf{K}_{\mathbf{h}}\|_1^2 \|f - \mathbf{K}_{\mathbf{h}} * f\|_{\infty}. \end{aligned}$$

Exactly as in the first section where the intensity of the noise was assumed to be known, we get

$$\|f - \mathbf{K}_{\mathbf{h}} * f\|_{\infty} \lesssim \sum_{i=1}^d h_i^{2\alpha_i}.$$

Finally we obtain

$$\max_{\boldsymbol{\eta} \in \mathcal{H}_n} |f_{\boldsymbol{\eta}}(x) - f_{\boldsymbol{\eta},\mathbf{h}}(x)|^2 \lesssim \sum_{i=1}^n h_i^{2\alpha_i}. \quad (\text{B.9})$$

Step 3. Now, we will use the concentration in order to control the remaining terms. We write:

$$\mathbb{E} \left[\max_{\boldsymbol{\eta} \in \mathcal{H}_n} \left\{ |\widehat{f}_{\boldsymbol{\eta},\mathbf{h}}(x) - f_{\boldsymbol{\eta},\mathbf{h}}(x)|^2 - V_{\boldsymbol{\eta}} \right\}_+ \right] \leq \sum_{\boldsymbol{\eta} \in \mathcal{H}_n} \mathbb{E} \left[\left\{ |\widehat{f}_{\boldsymbol{\eta},\mathbf{h}}(x) - f_{\boldsymbol{\eta},\mathbf{h}}(x)|^2 - V_{\boldsymbol{\eta}} \right\}_+ \right].$$

Now, for any $\boldsymbol{\eta}$ and $i \in \{1, \dots, n\}$, let us define

$$S_{\boldsymbol{\eta},\mathbf{h}}^i := \mathbf{K}_{\boldsymbol{\eta}} * \mathbf{K}_{\mathbf{h}}(x - X_i) - \mathbb{E}[\mathbf{K}_{\boldsymbol{\eta}} * \mathbf{K}_{\mathbf{h}}(x - X_i)]. \quad (\text{B.10})$$

First, let us remark that

$$\begin{aligned} |S_{\boldsymbol{\eta},\mathbf{h}}| &\leq \|\mathbf{K}_{\mathbf{h}} * \mathbf{K}_{\boldsymbol{\eta}}\|_{\infty} \\ &\leq 3\kappa_1 \prod_{i=1}^d \eta_i^{-1}. \end{aligned}$$

where the previous inequalities hold in the almost sure sense for some $\kappa_1 > 0$. Then, thanks to Bernstein inequality, we know that

$$\mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n S_{\boldsymbol{\eta},\mathbf{h}}^i \geq t \right) \leq \exp \left(- \frac{n^2 t^2}{2 \left(\sum_{i=1}^n \mathbb{E}[|S_{\boldsymbol{\eta},\mathbf{h}}^i|^2] + \kappa_1 \prod_{i=1}^d \eta_i^{-1} t n \right)} \right). \quad (\text{B.11})$$

For $i \in \{1, \dots, n\}$,

$$\begin{aligned} \mathbb{E}[|S_{\boldsymbol{\eta}, \mathbf{h}}^i|^2] &\leq \int_{\mathbb{R}^d} |\mathbf{K}_{\mathbf{h}} * \mathbf{K}_{\boldsymbol{\eta}}(x-y)|^2 f(y) dy \\ &\leq \|f\|_{\infty} \|\mathbf{K}_{\mathbf{h}} * \mathbf{K}_{\boldsymbol{\eta}}\|_2^2, \end{aligned}$$

Finally, using Young's inequality we obtain

$$\mathbb{E}[|S_{\boldsymbol{\eta}, \mathbf{h}}^i|^2] \leq \kappa_2 \prod_{i=1}^d \eta_i^{-1}, \quad (\text{B.12})$$

for some $\kappa_2 > 0$. Now for any $\boldsymbol{\eta} \in \mathcal{H}_n$ and using Bernstein inequality given in Proposition ...,

$$\begin{aligned} \mathbb{E} \left[\left\{ |\widehat{f}_{\boldsymbol{\eta}, \mathbf{h}}(x) - f_{\boldsymbol{\eta}, \mathbf{h}}(x)|^2 - V_{\boldsymbol{\eta}} \right\}_+ \right] &= \int_0^{+\infty} \mathbb{P} \left(|\widehat{f}_{\boldsymbol{\eta}, \mathbf{h}}(x) - f_{\boldsymbol{\eta}, \mathbf{h}}(x)| > t + V_{\boldsymbol{\eta}} \right) dt \\ &\leq \int_{V_{\boldsymbol{\eta}}}^{+\infty} \exp \left(-\frac{nt^2 \prod_{i=1}^d \eta_i}{2(\kappa_2 + \kappa_1 \sqrt{t})} \right) dt. \end{aligned}$$

Moreover, we know from Proposition ..., that this implies

$$\mathbb{E} \left[\left\{ |\widehat{f}_{\boldsymbol{\eta}, \mathbf{h}}(x) - f_{\boldsymbol{\eta}, \mathbf{h}}(x)|^2 - V_{\boldsymbol{\eta}} \right\}_+ \right] = \text{I} + \text{II},$$

with

$$\text{I} := \int_{V_{\boldsymbol{\eta}}}^{+\infty} \exp \left(-\frac{nt \prod_{i=1}^d \eta_i}{4\kappa_2} \right) dt; \quad \text{II} := \int_{V_{\boldsymbol{\eta}}}^{+\infty} \exp \left(-\frac{n\sqrt{t} \prod_{i=1}^d \eta_i}{4\kappa_1} \right) dt.$$

The term I is easy to compute and we obtain

$$\text{I} = \frac{4\kappa_2}{n \prod_{i=1}^d \eta_i} \exp \left(-\frac{nV_{\boldsymbol{\eta}}}{4\kappa_2} \prod_{i=1}^d \eta_i \right). \quad (\text{B.13})$$

To compute the term II, we use Proposition ... and get

$$\text{II} \leq \frac{4\kappa_1}{n \prod_{i=1}^d \eta_i} V_{\boldsymbol{\eta}}^{1/2} \exp \left(-\frac{nV_{\boldsymbol{\eta}}^{1/2}}{2\kappa_1} \prod_{i=1}^d \eta_i \right). \quad (\text{B.14})$$

Moreover, as previously mentioned,

$$V_{\boldsymbol{\eta}} = \bar{\omega} \log(n) n^{-1} \prod_{i=1}^n \eta_i^{-1},$$

for some $\bar{\omega} > 0$. Then, we obtain

$$\begin{aligned} \text{I} &\lesssim \frac{1}{n \prod_{i=1}^d \eta_i} \exp\left(-\frac{\bar{\omega}}{2\kappa_2} \log(n)\right); \\ \text{II} &\lesssim \frac{\log(n)^{1/2}}{n^{3/2} \prod_{i=1}^d \eta_i^{3/2}} \exp\left(-\frac{n^{1/2} \bar{\omega}^{1/2} \log(n)^{1/2}}{2\kappa_1} \prod_{i=1}^d \eta_i\right) \end{aligned}$$

It is now time to define properly the mesh \mathcal{H}_n . We want to define it in a way that allows to control both I and II. Let us consider

$$\mathcal{H}_n \subset \left\{ \mathbf{h} = (h_1, \dots, h_d) \in (0, 1]^d, h_i \geq n^{-1/d} \log(n)^{2/d} \right\}.$$

Then, we obtain that for $\bar{\omega}$ large enough,

$$\text{I} + \text{II} \lesssim n^{-2}.$$

With the exact same reasoning, we obtain that for $\bar{\omega}$ large enough,

$$\mathbb{E} \left[\left\{ |\hat{f}_\eta(x) - f_\eta(x)|^2 - V_\eta \right\}_+ \right] \lesssim n^{-2}.$$

Now imposing that $\#\mathcal{H}_n \lesssim n$, we get

$$\mathbb{E}[B_{\mathbf{h}}] \lesssim \sum_{i=1}^d h_i^{2\alpha_i} + \frac{1}{n}.$$

Finally, plugging this into (B.8), we conclude the proof of Proposition B.4. □

References

- [ABC⁺00] Cécile ANÉ, Sébastien BLACHÈRE, Djalil CHAFAÏ, Pierre FOUGÈRES, Ivan GENTIL, Florent MALRIEU, Cyril ROBERTO et Grégory SCHEFFER : Sur les inégalités de Sobolev logarithmiques, volume 10 de Panoramas et Synthèses [Panoramas and Syntheses]. Société Mathématique de France, Paris, 2000. With a preface by Dominique Bakry and Michel Ledoux.
- [ABKO23] L. ANGELI, J. BARRÉ, M. KOŁODZIEJCZYK et M. OTTOBRE : Well-posedness and stationary solutions of McKean-Vlasov (S)PDEs. J. Math. Anal. Appl., 526(2):Paper No. 127301, 49, 2023.
- [ACKO24] Letizia ANGELI, Dan CRISAN, Martin KOŁODZIEJCZYK et Michela OTTOBRE : Approximation of non-linear spdes with additive noise via weighted interacting particles systems: the stochastic mckean-vlasov equation. arXiv preprint arXiv:2404.07488, 2024.
- [AG21] Chiara AMORINO et Arnaud GLOTER : Minimax rate of estimation for invariant densities associated to continuous stochastic differential equations over anisotropic Holder classes. arXiv:2110.02774, 2021.
- [AG22] Chiara AMORINO et Arnaud GLOTER : Malliavin calculus for the optimal estimation of the invariant density of discretely observed diffusions in intermediate regime. arXiv:2110.02774, 2022.
- [AG23] Chiara AMORINO et Arnaud GLOTER : Estimation of the invariant density for discretely observed diffusion processes: impact of the sampling and of the asynchronicity. Statistics, 57(1):213–259, 2023.
- [AGS08] Luigi AMBROSIO, Nicola GIGLI et Giuseppe SAVARÉ : Gradient flows in metric spaces and in the space of probability measures. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second édition, 2008.
- [AL14] R. ANDREEV et A. LANG : Kolmogorov-Chentsov theorem and differentiability of random fields on manifolds. Potential Anal., 41(3):761–769, 2014.
- [AN22] Chiara AMORINO et Eulalia NUALART : Optimal convergence rates for the invariant density estimation of jump-diffusion processes. ESAIM Probab. Stat., 26:126–151, 2022.

References

- [ASJ10] Yacine AÏT-SAHALIA et Jean JACOD : Is Brownian motion necessary to model high-frequency data? Ann. Statist., 38(5):3093–3128, 2010.
- [ASJ14] Yacine AÏT-SAHALIA et Jean JACOD : High-frequency financial econometrics. Princeton University Press, 2014.
- [Bai57] Norman T. J. BAILEY : The mathematical theory of epidemics. Hafner Publishing Co., New York, 1957.
- [Bas20] K. BASHIRI : On the long-time behaviour of McKean-Vlasov paths. Electron. Commun. Probab., 25:Paper No. 52, 14, 2020.
- [BB92] James BERGIN et Dan BERNHARDT : Anonymous sequential games with aggregate uncertainty. J. Math. Econom., 21(6):543–562, 1992.
- [BBCG08] Dominique BAKRY, Franck BARTHE, Patrick CATTIAUX et Arnaud GUILLIN : A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case. Electron. Commun. Probab., 13:60–66, 2008.
- [BCCdRH20] Philippe BRIAND, Pierre CARDALIAGUET, Paul-Éric Chaudru de RAYNAL et Ying HU : Forward and backward stochastic differential equations with normal constraints in law. Stochastic Process. Appl., 130(12):7021–7097, 2020.
- [BCG08] Dominique BAKRY, Patrick CATTIAUX et Arnaud GUILLIN : Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. J. Funct. Anal., 254(3):727–759, 2008.
- [BE85] D. BAKRY et Michel ÉMERY : Diffusions hypercontractives. In Séminaire de probabilités, XIX, 1983/84, volume 1123 de Lecture Notes in Math., pages 177–206. Springer, Berlin, 1985.
- [BEH18] Philippe BRIAND, Romuald ELIE et Ying HU : BSDEs with mean reflection. Ann. Appl. Probab., 28(1):482–510, 2018.
- [Ber93] Albert R. BERGSTROM : The history of continuous-time econometric models. Springer, 1993.
- [BGG13] François BOLLEY, Ivan GENTIL et Arnaud GUILLIN : Uniform convergence to equilibrium for granular media. Arch. Ration. Mech. Anal., 208(2):429–445, 2013.
- [BGL14] Dominique BAKRY, Ivan GENTIL et Michel LEDOUX : Analysis and geometry of Markov diffusion operators, volume 348 de Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer, Cham, 2014.
- [BH21] Philippe BRIAND et Hélène HIBON : Particles systems for mean reflected BSDEs. Stochastic Process. Appl., 131:253–275, 2021.

- [Bha78] R. N. BHATTACHARYA : Criteria for recurrence and existence of invariant measures for multidimensional diffusions. Ann. Probab., 6(4):541–553, 1978.
- [Bog07] V. I. BOGACHEV : Measure theory. Vol. I, II. Springer-Verlag, Berlin, 2007.
- [Bol70] Ludwig BOLTZMANN : Weitere studien über das wärmeleichgewicht unter gasmolekülen. Kinetische Theorie II, pages 115–225, 1970.
- [Bos97] Denis BOSQ : Parametric rates of nonparametric estimators and predictors for continuous time processes. Ann. Statist., 25(3):982–1000, 1997.
- [Bos98] D. BOSQ : Nonparametric statistics for stochastic processes, volume 110 de Lecture Notes in Statistics. Springer-Verlag, New York, second édition, 1998. Estimation and prediction.
- [BP03] Federico M. BANDI et Peter C. B. PHILLIPS : Fully nonparametric estimation of scalar diffusion models. Econometrica, 71(1):241–283, 2003.
- [BRV98] S. BENACHOUR, B. ROYNETTE et P. VALLOIS : Nonlinear self-stabilizing processes. II. Convergence to invariant probability. Stochastic Process. Appl., 75(2):203–224, 1998.
- [But14] O. A. BUTKOVSKY : On ergodic properties of nonlinear Markov chains and stochastic McKean-Vlasov equations. Theory Probab. Appl., 58(4):661–674, 2014.
- [Car10] Pierre CARDALIAGUET : Notes on mean field games. Rapport technique, Technical report, 2010.
- [CD18a] René CARMONA et François DELARUE : Probabilistic theory of mean field games with applications. I, volume 83 de Probability Theory and Stochastic Modelling. Springer, Cham, 2018. Mean field FBSDEs, control, and games.
- [CD18b] René CARMONA et François DELARUE : Probabilistic theory of mean field games with applications. II, volume 84 de Probability Theory and Stochastic Modelling. Springer, Cham, 2018. Mean field games with common noise and master equations.
- [CD⁺18c] René CARMONA, François DELARUE et al. : Probabilistic Theory of Mean Field Games with Applications I–II. Springer, 2018.
- [CD22a] Louis-Pierre CHAINTRON et Antoine DIEZ : Propagation of chaos: a review of models, methods and applications. I. Models and methods. Kinet. Relat. Models, 15(6):895–1015, 2022.
- [CD22b] Louis-Pierre CHAINTRON et Antoine DIEZ : Propagation of chaos: a review of models, methods and applications. II. Applications. Kinet. Relat. Models, 15(6):1017–1173, 2022.
- [CDL16] René CARMONA, François DELARUE et Daniel LACKER : Mean field games with common noise. Ann. Probab., 44(6):3740–3803, 2016.

References

- [CDLL19] Pierre CARDALIAGUET, François DELARUE, Jean-Michel LASRY et Pierre-Louis LIONS : The master equation and the convergence problem in mean field games, volume 201 de Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, 2019.
- [CFS15] René CARMONA, Jean-Pierre FOUQUE et Li-Hsien SUN : Mean field games and systemic risk. Commun. Math. Sci., 13(4):911–933, 2015.
- [CG08] Patrick CATTIAUX et Arnaud GUILLIN : Deviation bounds for additive functionals of Markov processes. ESAIM Probab. Stat., 12:12–29, 2008.
- [CG19] Michele COGHI et Benjamin GESS : Stochastic nonlinear Fokker-Planck equations. Nonlinear Anal., 187:259–278, 2019.
- [CGC20] Fabienne COMTE et Valentine GENON-CATALOT : Nonparametric drift estimation for i.i.d. paths of stochastic differential equations. Ann. Statist., 48(6):3336–3365, 2020.
- [CGCR07] Fabienne COMTE, Valentine GENON-CATALOT et Yves ROZENHOLC : Penalized nonparametric mean square estimation of the coefficients of diffusion processes. Bernoulli, 13(2):514–543, 2007.
- [CGM08] P. CATTIAUX, A. GUILLIN et F. MALRIEU : Probabilistic approach for granular media equations in the non-uniformly convex case. Probab. Theory Related Fields, 140(1-2):19–40, 2008.
- [CH88] Raymond J. CARROLL et Peter HALL : Optimal rates of convergence for deconvolving a density. J. Amer. Statist. Assoc., 83(404):1184–1186, 1988.
- [Cha13] Gaëlle CHAGNY : Estimation adaptative avec des données transformées ou incomplètes. Application à des modèles de survie. Thèse de doctorat, Université René Descartes-Paris V, 2013.
- [CL13] F. COMTE et C. LACOUR : Anisotropic adaptive kernel deconvolution. Ann. Inst. Henri Poincaré Probab. Stat., 49(2):569–609, 2013.
- [CL17] Djalil CHAFAÏ et Joseph LEHEC : Logarithmic Sobolev Inequalities Essentials, 2017.
- [CM05] F. COMTE et F. MERLEVÈDE : Super optimal rates for nonparametric density estimation via projection estimators. Stochastic Process. Appl., 115(5):797–826, 2005.
- [CMV03] José A. CARRILLO, Robert J. McCANN et Cédric VILLANI : Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. Rev. Mat. Iberoamericana, 19(3):971–1018, 2003.
- [CMV06] José A. CARRILLO, Robert J. McCANN et Cédric VILLANI : Contractions in the 2-Wasserstein length space and thermalization of granular media. Arch. Ration. Mech. Anal., 179(2):217–263, 2006.

- [Con23] Giovanni CONFORTI : Coupling by reflection for controlled diffusion processes: turnpike property and large time behavior of Hamilton-Jacobi-Bellman equations. Ann. Appl. Probab., 33(6A):4608–4644, 2023.
- [CT16] Jakub CHOROWSKI et Mathias TRABS : Spectral estimation for diffusions with random sampling times. Stochastic Process. Appl., 126(10):2976–3008, 2016.
- [CW97] Mu-Fa CHEN et Feng-Yu WANG : Estimation of spectral gap for elliptic operators. Trans. Amer. Math. Soc., 349(3):1239–1267, 1997.
- [Dar70] Gaston DARBOUX : Sur les équations aux dérivées partielles du second ordre, volume 7. 1870.
- [Daw83] Donald A. DAWSON : Critical dynamics and fluctuations for a mean-field model of cooperative behavior. J. Statist. Phys., 31(1):29–85, 1983.
- [DEGZ20] Alain DURMUS, Andreas EBERLE, Arnaud GUILLIN et Raphael ZIMMER : An elementary approach to uniform in time propagation of chaos. Proc. Amer. Math. Soc., 148(12):5387–5398, 2020.
- [Del80] M. DELECROIX : Sur l'estimation des densités d'un processus stationnaire à temps continu. Publ. Inst. Statist. Univ. Paris, 25(1-2):17–39, 1980.
- [Del19] François DELARUE : Restoring uniqueness to mean-field games by randomizing the equilibria. Stoch. Partial Differ. Equ. Anal. Comput., 7(4):598–678, 2019.
- [Dev89] Luc DEVROYE : Consistent deconvolution in density estimation. Canad. J. Statist., 17(2):235–239, 1989.
- [DFT20] François DELARUE et Rinel FOGUEN TCHUENDOM : Selection of equilibria in a linear quadratic mean-field game. Stochastic Process. Appl., 130(2):1000–1040, 2020.
- [DH24] François DELARUE et William R. P. HAMMERSLEY : Rearranged stochastic heat equation: Ergodicity and related gradient descent on the space of probability measures. arXiv 2403.16140, 2024.
- [DHB⁺74] R. D. DESLATTES, A. HENINS, H. A. BOWMAN, R. M. SCHOONOVER, C. L. CARROLL, I. L. BARNES, L. A. MACHLAN, L. J. MOORE et W. R. SHIELDS : Determination of the avogadro constant. Phys. Rev. Lett., 33:463–466, Aug 1974.
- [DMT19] Pierre DEL MORAL et Julian TUGAUT : Uniform propagation of chaos and creation of chaos for a class of nonlinear diffusions. Stoch. Anal. Appl., 37(6):909–935, 2019.
- [DMT24] François DELARUE, Raphaël MAILLET et Etienne TANRÉ : Ergodicity of some stochastic fokker-planck equations with additive common noise. arXiv e-prints, pages arXiv–2405, 2024.

References

- [DR06] Arnak DALALYAN et Markus REISS : Asymptotic statistical equivalence for scalar ergodic diffusions. Probab. Theory Related Fields, 134(2):248–282, 2006.
- [DR07] Arnak DALALYAN et Markus REISS : Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case. Probab. Theory Related Fields, 137(1-2):25–47, 2007.
- [Dud02] R. M. DUDLEY : Real analysis and probability, volume 74 de Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.
- [DV95] Donald DAWSON et Jean VAILLANCOURT : Stochastic McKean-Vlasov equations. NoDEA Nonlinear Differential Equations Appl., 2(2):199–229, 1995.
- [DV21] François DELARUE et Athanasios VASILEIADIS : Exploration noise for learning linear-quadratic mean field games. arXiv preprint arXiv:2107.00839, 2021.
- [Ebe16] Andreas EBERLE : Reflection couplings and contraction rates for diffusions. Probab. Theory Related Fields, 166(3-4):851–886, 2016.
- [EZ19] Andreas EBERLE et Raphael ZIMMER : Sticky couplings of multidimensional diffusions with different drifts. Ann. Inst. Henri Poincaré Probab. Stat., 55(4):2370–2394, 2019.
- [Fan91] Jianqing FAN : On the optimal rates of convergence for nonparametric deconvolution problems. Ann. Statist., 19(3):1257–1272, 1991.
- [Fan93] Jianqing FAN : Adaptively local one-dimensional subproblems with application to a deconvolution problem. Ann. Statist., 21(2):600–610, 1993.
- [FG16] Peter K. FRIZ et Benjamin GESS : Stochastic scalar conservation laws driven by rough paths. Ann. Inst. H. Poincaré C Anal. Non Linéaire, 33(4):933–963, 2016.
- [FG19] Benjamin FEHRMAN et Benjamin GESS : Well-posedness of nonlinear diffusion equations with nonlinear, conservative noise. Arch. Ration. Mech. Anal., 233(1):249–322, 2019.
- [Fra24] Céline FRANCONY : Intégration du risque climatique dans les scénarios risque-neutre des trajectoires actions. Mémoire d'actuariat - Institut des Actuaires, 2024.
- [Fun84] Tadahisa FUNAKI : A certain class of diffusion processes associated with nonlinear parabolic equations. Z. Wahrsch. Verw. Gebiete, 67(3):331–348, 1984.
- [GC16] Valentine GENON-CATALOT : Cours de statistique des diffusions. Rapport technique, Master course, 2016.
- [GCJ94] V. GENON-CATALOT et J. JACOD : Estimation of the diffusion coefficient for diffusion processes: random sampling. Scand. J. Statist., 21(3):193–221, 1994.
- [GCLP92] V. GENON-CATALOT, C. LAREDO et D. PICARD : Nonparametric estimation of the diffusion coefficient by wavelets methods. Scand. J. Statist., 19(4):317–335, 1992.

- [GG08] Arnaud GLOTTER et Emmanuel GOBET : LAMN property for hidden processes: the case of integrated diffusions. Ann. Inst. Henri Poincaré Probab. Stat., 44(1):104–128, 2008.
- [GHR04] Emmanuel GOBET, Marc HOFFMANN et Markus REISS : Nonparametric estimation of scalar diffusions based on low frequency data. Ann. Statist., 32(5):2223–2253, 2004.
- [GKM⁺96] Carl GRAHAM, Thomas G KURTZ, Sylvie MÉLÉARD, Philip E PROTTER, Mario PULVIRENTI et Denis TALAY : Asymptotic behaviour of some interacting particle systems; McKean–Vlasov and Boltzmann models. Probabilistic Models for Nonlinear Partial Differential Equations: Lectures given at the 1st Session of the Centro Internazionale Matematico Estivo (CIME) held in Montecatini Terme, Italy, May 22–30, 1995, pages 42–95, 1996.
- [GL08] Alexander GOLDENSHLUGER et Oleg LEPSKI : Universal pointwise selection rule in multivariate function estimation. Bernoulli, 14(4):1150–1190, 2008.
- [GL09] Alexander GOLDENSHLUGER et Oleg LEPSKI : Structural adaptation via \mathbb{L}_p -norm oracle inequalities. Probab. Theory Related Fields, 143(1-2):41–71, 2009.
- [GL11] Alexander GOLDENSHLUGER et Oleg LEPSKI : Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. Ann. Statist., 39(3):1608–1632, 2011.
- [GS15] Benjamin GESS et Panagiotis E. SOUGANIDIS : Scalar conservation laws with multiple rough fluxes. Commun. Math. Sci., 13(6):1569–1597, 2015.
- [GS17] Benjamin GESS et Panagiotis E. SOUGANIDIS : Stochastic non-isotropic degenerate parabolic-hyperbolic equations. Stochastic Process. Appl., 127(9):2961–3004, 2017.
- [GSSS15] Kay GIESECKE, Konstantinos SPILIOPOULOS, Richard B. SOWERS et Justin A. SIRIGNANO : Large portfolio asymptotics for loss from default. Math. Finance, 25(1):77–114, 2015.
- [GT19] Wilfrid GANGBO et Adrian TUDORASCU : On differentiability in the Wasserstein space and well-posedness for Hamilton-Jacobi equations. J. Math. Pures Appl. (9), 125:119–174, 2019.
- [HCM07] Minyi HUANG, Peter E. CAINES et Roland P. MALHAMÉ : Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and decentralized ϵ -Nash equilibria. IEEE Trans. Automat. Control, 52(9):1560–1571, 2007.
- [HIP08] Samuel HERRMANN, Peter IMKELLER et Dierk PEITHMANN : Large deviations and a Kramers’ type law for self-stabilizing diffusions. Ann. Appl. Probab., 18(4):1379–1423, 2008.
- [HJ15] Maxime HAURAY et Pierre-Emmanuel JABIN : Particle approximation of Vlasov equations with singular forces: propagation of chaos. Ann. Sci. Éc. Norm. Supér. (4), 48(4):891–940, 2015.

References

- [HMC06] Minyi HUANG, Roland P. MALHAMÉ et Peter E. CAINES : Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. Commun. Inf. Syst., 6(3):221–251, 2006.
- [Hof97] Marc HOFFMANN : Estimation non-paramétrique du coefficient de diffusion pour une perte L_p . C. R. Acad. Sci. Paris Sér. I Math., 324(4):475–480, 1997.
- [Hof99] Marc HOFFMANN : Adaptive estimation in diffusion processes. Stochastic Process. Appl., 79(1):135–163, 1999.
- [HP13] Nikolaus HAUTSCH et Mark PODOLSKIJ : Preaveraging-based estimation of quadratic variation in the presence of noise and jumps: theory, implementation, and empirical evidence. J. Bus. Econom. Statist., 31(2):165–183, 2013.
- [HT10] S. HERRMANN et J. TUGAUT : Non-uniqueness of stationary measures for self-stabilizing processes. Stochastic Process. Appl., 120(7):1215–1246, 2010.
- [Hul03] John C. HULL : Options futures and other derivatives. Pearson Education India, 2003.
- [HvS21] William R. P. HAMMERSLEY, David ŠIŠKA et Ł ukasz SZPRUCH : Weak existence and uniqueness for McKean-Vlasov SDEs with common noise. Ann. Probab., 49(2):527–555, 2021.
- [Jea15] JH JEANS : On the theory of star-streaming and the structure of the universe. Monthly Notices of the Royal Astronomical Society, 76(2):70–84, 1915.
- [Jea19] James JEANS : Problems of cosmogony and stellar dynamics. University press, 1919.
- [JJ24] Bao JIANHAI et Wang JIAN : Long time behavior of one-dimensional mckean-vlasov sdes with common noise. arXiv preprint arXiv:2401.07665, 2024.
- [JKO98] Richard JORDAN, David KINDERLEHRER et Felix OTTO : The variational formulation of the Fokker-Planck equation. SIAM J. Math. Anal., 29(1):1–17, 1998.
- [JLM⁺09] Jean JACOD, Yingying LI, Per A. MYKLAND, Mark PODOLSKIJ et Mathias VETTER : Microstructure noise in the continuous case: the pre-averaging approach. Stochastic Process. Appl., 119(7):2249–2276, 2009.
- [JM15] Jean JACOD et Per A. MYKLAND : Microstructure noise in the continuous case: approximate efficiency of the adaptive pre-averaging method. Stochastic Process. Appl., 125(8):2910–2936, 2015.
- [JPV10] Jean JACOD, Mark PODOLSKIJ et Mathias VETTER : Limit theorems for moving averages of discretized processes plus noise. Ann. Statist., 38(3):1478–1545, 2010.
- [JT21] Benjamin JOURDAIN et Alvin TSE : Central limit theorem over non-linear functionals of empirical measures with applications to the mean-field fluctuation of interacting diffusions. Electron. J. Probab., 26:Paper No. 154, 34, 2021.

- [Kac56] M. KAC : Foundations of kinetic theory. In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III, pages 171–197. Univ. California Press, Berkeley-Los Angeles, Calif., 1956.
- [Kin78] J. F. C. KINGMAN : Uses of exchangeability. Ann. Probability, 6(2):183–197, 1978.
- [KM12] I. KONTOYIANNIS et S. P. MEYN : Geometric ergodicity and the spectral gap of non-reversible Markov chains. Probab. Theory Related Fields, 154(1-2):327–339, 2012.
- [KNRS22] Chaman KUMAR, NEELIMA, Christoph REISINGER et Wolfgang STOCKINGER : Well-posedness and tamed schemes for McKean-Vlasov equations with common noise. Ann. Appl. Probab., 32(5):3283–3330, 2022.
- [KS91] Ioannis KARATZAS et Steven E. SHREVE : Brownian motion and stochastic calculus, volume 113 de Graduate Texts in Mathematics. Springer-Verlag, New York, second édition, 1991.
- [KX99] Thomas G. KURTZ et Jie XIONG : Particle representations for a class of nonlinear SPDEs. Stochastic Process. Appl., 83(1):103–126, 1999.
- [Lac18] Daniel LACKER : Mean field games and interacting particle systems. Rapport technique, Technical report, 2018.
- [Leb97] Frédérique LEBLANC : Density estimation for a class of continuous time processes. Math. Methods Statist., 6(2):171–199, 1997.
- [Lem21] Michał LEMAŃCZYK : General Bernstein-like inequality for additive functionals of Markov chains. J. Theoret. Probab., 34(3):1426–1454, 2021.
- [Lez01] Pascal LEZAUD : Chernoff and Berry-Esséen inequalities for Markov processes. ESAIM Probab. Statist., 5:183–201, 2001.
- [LL63] Elliott H. LIEB et Werner LINIGER : Exact analysis of an interacting Bose gas. I–II. Phys. Rev. (2), 130:1605–1616, 1963.
- [LL06a] Jean-Michel LASRY et Pierre-Louis LIONS : Jeux à champ moyen. I. Le cas stationnaire. C. R. Math. Acad. Sci. Paris, 343(9):619–625, 2006.
- [LL06b] Jean-Michel LASRY et Pierre-Louis LIONS : Jeux à champ moyen. II. Horizon fini et contrôle optimal. C. R. Math. Acad. Sci. Paris, 343(10):679–684, 2006.
- [LL07] Jean-Michel LASRY et Pierre-Louis LIONS : Mean field games. Jpn. J. Math., 2(1):229–260, 2007.
- [LP02] Damien LAMBERTON et Gilles PAGÈS : Recursive computation of the invariant distribution of a diffusion. Bernoulli, 8(3):367–405, 2002.

References

- [LPS13] Pierre-Louis LIONS, Benoît PERTHAME et Panagiotis E. SOUGANIDIS : Scalar conservation laws with rough (stochastic) fluxes. Stoch. Partial Differ. Equ. Anal. Comput., 1(4):664–686, 2013.
- [LPS14] Pierre-Louis LIONS, Benoît PERTHAME et Panagiotis E. SOUGANIDIS : Scalar conservation laws with rough (stochastic) fluxes: the spatially dependent case. Stoch. Partial Differ. Equ. Anal. Comput., 2(4):517–538, 2014.
- [LR86] Torgny LINDVALL et L. C. G. ROGERS : Coupling of multidimensional diffusions by reflection. Ann. Probab., 14(3):860–872, 1986.
- [LSZ20] Daniel LACKER, Mykhaylo SHKOLNIKOV et Jiacheng ZHANG : Superposition and mimicking theorems for conditional McKean–Vlasov equations, 2020.
- [LSZ23] Daniel LACKER, Mykhaylo SHKOLNIKOV et Jiacheng ZHANG : Superposition and mimicking theorems for conditional McKean-Vlasov equations. J. Eur. Math. Soc. (JEMS), 25(8):3229–3288, 2023.
- [LT89] Ming Chung LIU et Robert L. TAYLOR : A consistent nonparametric density estimator for the deconvolution problem. Canad. J. Statist., 17(4):427–438, 1989.
- [Mai23] Raphaël MAILLET : A note on the Long-Time behaviour of Stochastic McKean-Vlasov Equations with common noise. arXiv preprint arXiv:2306.16130, 2023.
- [Mal01] F. MALRIEU : Logarithmic Sobolev inequalities for some nonlinear PDE's. Stochastic Process. Appl., 95(1):109–132, 2001.
- [Mal03] Florent MALRIEU : Convergence to equilibrium for granular media equations and their Euler schemes. Ann. Appl. Probab., 13(2):540–560, 2003.
- [Mar21] Victor MARX : Infinite-dimensional regularization of McKean-Vlasov equation with a Wasserstein diffusion. Ann. Inst. Henri Poincaré Probab. Stat., 57(4):2315–2353, 2021.
- [McK66] H. P. MCKEAN, Jr. : A class of Markov processes associated with nonlinear parabolic equations. Proc. Nat. Acad. Sci. U.S.A., 56:1907–1911, 1966.
- [Mél96] Sylvie MÉLÉARD : Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995), volume 1627 de Lecture Notes in Math., pages 42–95. Springer, Berlin, 1996.
- [MS24] Raphaël MAILLET et Grégoire SZYMANSKI : Estimation of the invariant measure of a multidimensional diffusion from noisy observations. arXiv preprint arXiv:2404.12181, 2024.
- [MT93] S. P. MEYN et R. L. TWEEDIE : Markov chains and stochastic stability. Communications and Control Engineering Series. Springer-Verlag London, Ltd., London, 1993.

- [NZ79] Xuan-Xanh NGUYEN et Hans ZESSIN : Ergodic theorems for spatial processes. Z. Wahrsch. Verw. Gebiete, 48(2):133–158, 1979.
- [Ott99] Felix OTTO : Evolution of microstructure in unstable porous media flow: a relaxational approach. Comm. Pure Appl. Math., 52(7):873–915, 1999.
- [Ott01] Felix OTTO : The geometry of dissipative evolution equations: the porous medium equation. Comm. Partial Differential Equations, 26(1-2):101–174, 2001.
- [Pan08] Fabien PANLOUP : Recursive computation of the invariant measure of a stochastic differential equation driven by a Lévy process. Ann. Appl. Probab., 18(2):379–426, 2008.
- [Pap95] George C. PAPANICOLAOU : Diffusion in random media. In Surveys in applied mathematics, Vol. 1, volume 1 de Surveys Appl. Math., pages 205–253. Plenum, New York, 1995.
- [Par62] Emanuel PARZEN : On estimation of a probability density function and mode. Ann. Math. Statist., 33:1065–1076, 1962.
- [Pau15] Daniel PAULIN : Concentration inequalities for Markov chains by Marton couplings and spectral methods. Electron. J. Probab., 20:no. 79, 32, 2015.
- [QRZ03] Zhongmin QIAN, Francesco RUSSO et Weian ZHENG : Comparison theorem and estimates for transition probability densities of diffusion processes. Probab. Theory Related Fields, 127(3):388–406, 2003.
- [QZ04] Zhongmin QIAN et Weian ZHENG : A representation formula for transition probability densities of diffusions and applications. Stochastic Process. Appl., 111(1):57–76, 2004.
- [Ric77] Luigi M. RICCIARDI : Diffusion processes and related topics in biology, volume Vol. 14 de Lecture Notes in Biomathematics. Springer-Verlag, Berlin-New York, 1977. Notes taken by Charles E. Smith.
- [Ris89] H. RISKEN : The Fokker-Planck equation, volume 18 de Springer Series in Synergetics. Springer-Verlag, Berlin, second édition, 1989. Methods of solution and applications.
- [Ros56] Murray ROSENBLATT : Remarks on some nonparametric estimates of a density function. Ann. Math. Statist., 27:832–837, 1956.
- [RY99] Daniel REVUZ et Marc YOR : Continuous martingales and Brownian motion, volume 293 de Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, third édition, 1999.
- [SC90] Leonard STEFANSKI et Raymond J. CARROLL : Deconvoluting kernel density estimators. Statistics, 21(2):169–184, 1990.
- [Sch85a] Michael SCHEUTZOW : Noise can create periodic behavior and stabilize nonlinear diffusions. Stochastic Process. Appl., 20(2):323–331, 1985.

References

- [Sch85b] Michael SCHEUTZOW : Some examples of nonlinear diffusion processes having a time-periodic law. Ann. Probab., 13(2):379–384, 1985.
- [Sch93] M. SCHEUTZOW : Stabilization and destabilization by noise in the plane. Stochastic Anal. Appl., 11(1):97–113, 1993.
- [Sch11] Emeline SCHMISSER : Non-parametric drift estimation for diffusions from noisy data. Statist. Decisions, 28(2):119–150, 2011.
- [Sch12] Emeline SCHMISSER : Non-parametric estimation of the diffusion coefficient from noisy data. Stat. Inference Stoch. Process., 15(3):193–223, 2012.
- [Sch13] Emeline SCHMISSER : Penalized nonparametric drift estimation for a multidimensional diffusion process. Statistics, 47(1):61–84, 2013.
- [Spo80] Herbert SPOHN : Kinetic equations from Hamiltonian dynamics: Markovian limits. Rev. Modern Phys., 52(3):569–615, 1980.
- [Str18] Claudia STRAUCH : Adaptive invariant density estimation for ergodic diffusions over anisotropic classes. Ann. Statist., 46(6B):3451–3480, 2018.
- [SV79] Daniel W. STROOCK et S. R. Srinivasa VARADHAN : Multidimensional diffusion processes, volume 233 de Grundlehren der Mathematischen Wissenschaften. Springer-Verlag, Berlin-New York, 1979.
- [Szn91] Alain-Sol SZNITMAN : Topics in propagation of chaos. In École d’Été de Probabilités de Saint-Flour XIX—1989, volume 1464 de Lecture Notes in Math., pages 165–251. Springer, Berlin, 1991.
- [Tch18a] Rinel Foguen TCHUENDOM : Instances of uniqueness restoration and equilibria selection in mean field games. Thèse de doctorat, Université Côte d’Azur, 2018.
- [Tch18b] Rinel Foguen TCHUENDOM : Uniqueness for linear-quadratic mean field games with common noise. Dyn. Games Appl., 8(1):199–210, 2018.
- [Tsy09] Alexandre B. TSYBAKOV : Introduction to nonparametric estimation. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.
- [Tug13] Julian TUGAUT : Convergence to the equilibria for self-stabilizing processes in double-well landscape. Ann. Probab., 41(3A):1427–1460, 2013.
- [Ver12] Mark VERAAR : The stochastic Fubini theorem revisited. Stochastics, 84(4):543–551, 2012.
- [Ves42] E. VESSIOT : Sur les équations aux dérivées partielles du second ordre, $F(x, y, z, p, q, r, s, t) = 0$, intégrables par la méthode de Darboux. J. Math. Pures Appl. (9), 21:1–66, 1942.

-
- [Vil09] Cédric VILLANI : Optimal transport, volume 338 de Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 2009. Old and new.
- [Vla45] A. VLASOV : On the kinetic theory of an assembly of particles with collective interaction. Acad. Sci. USSR. J. Phys., 9:25–40, 1945.
- [vRS05] Max-K. von RENESSE et Karl-Theodor STURM : Transport inequalities, gradient estimates, entropy, and Ricci curvature. Comm. Pure Appl. Math., 58(7):923–940, 2005.
- [Wei07] Pierre WEISS : L’hypothèse du champ moléculaire et la propriété ferromagnétique. J. Phys. Theor. Appl., 6(1):661–690, 1907.
- [Yag20] Nima YAGHMAEI : Human cost of disasters: An overview of the last 20 years, 2000-2019. UN Office for Disaster Risk Reduction, 2020.
- [Zol76] V. M. ZOLOTAREV : Metric distances in spaces of random variables and of their distributions. Mat. Sb. (N.S.), 101(143)(3):416–454, 456, 1976.

RÉSUMÉ

Cette thèse traite du comportement en temps long des équations stochastiques de Fokker-Planck en présence d'un bruit commun additif et présente des méthodes statistiques pour estimer la mesure invariante des processus de diffusion ergodiques multidimensionnels à partir de données bruitées. Dans la première partie, nous analysons les équations différentielles partielles stochastiques de type Fokker-Planck non linéaires, obtenues comme la limite du champ moyen de systèmes de particules en interaction dirigés par des bruits browniens idiosyncrasiques et en présence de bruit commun. Nous établissons des conditions sous lesquelles l'ajout d'un bruit commun permet de restaurer l'unicité de la mesure invariante. La principale difficulté provient de la dimension finie du bruit commun, alors que la variable d'état - *interprétée comme la loi marginale conditionnelle du système compte tenu du bruit commun* - opère dans un espace de dimension infinie. Nous démontrons que l'unicité est rétablie dès lors que le terme d'interaction du champ moyen attire le système vers sa moyenne conditionnelle (par rapport au bruit commun), en particulier lorsque l'intensité du bruit idiosyncrasique est faible, qui sont des cas typiques de perte d'unicité en l'absence de bruit commun.

Dans la deuxième partie, nous développons une méthodologie statistique afin d'approximer la mesure invariante d'un processus de diffusion à partir d'observations bruitées et discrètes de ce même processus. Cette méthode implique une technique de pré-moyennage des données qui réduit l'intensité du bruit tout en conservant les caractéristiques analytiques et les propriétés asymptotiques du signal sous-jacent. Nous étudions le taux de convergence de cet estimateur, qui dépend de la régularité anisotrope de la densité et de l'intensité du bruit. Nous établissons ensuite des conditions sur l'intensité du bruit qui permettent d'obtenir des taux de convergence comparables à ceux des cas sans bruit. Enfin, nous démontrons une inégalité de concentration de type Bernstein pour notre estimateur, ce qui permet de mettre en place une procédure adaptative pour la sélection de la fenêtre du noyau.

MOTS CLÉS

Équations de Fokker-Planck Stochastiques, Bruit Commun, Mesure Invariante, Estimation de Densité par Noyau, Processus de Diffusion Ergodique, Réduction du Bruit, Estimation Non-paramétrique

ABSTRACT

This thesis deals with the long-time behavior of stochastic Fokker-Planck equations with additive common noise and presents statistical methods for estimating the invariant measure of multidimensional ergodic diffusion processes from noisy data. In the first part, we analyze stochastic Fokker-Planck Partial Differential Equations (SPDEs), obtained as the mean-field limit of interacting particle systems influenced by both idiosyncratic and common Brownian noises. We establish conditions under which the addition of common noise restores uniqueness of the invariant measure. The main challenge arises from the finite-dimensional nature of the common noise, while the state variable — *interpreted as the conditional marginal law of the system given the common noise* — operates within an infinite-dimensional space. We demonstrate that uniqueness is restored if the mean field interaction term attracts the system towards its conditional mean given the common noise, particularly when the intensity of the idiosyncratic noise is small. In the second part, we develop a new statistical methodology using kernel density estimation to effectively approximate the invariant measure from noisy observations, highlighting the crucial role of the underlying Markov structure in the denoising process. This method involves a pre-averaging technique that proficiently reduces the intensity of the noise while maintaining the analytical characteristics and asymptotic properties of the underlying signal. We investigate the convergence rate of our estimator, which depends on the anisotropic regularity of the density and the intensity of the noise. We establish noise intensity conditions that allow for convergence rates comparable to those in noise-free environments. Additionally, we demonstrate a Bernstein concentration inequality for our estimator, leading to an adaptive procedure for selecting the kernel bandwidth.

KEYWORDS

Stochastic Fokker-Planck, Common Noise, Invariant Measure, Kernel Density Estimation, Ergodic Diffusion Process, Noise Reduction, Non-parametric Estimation