



Estimation statistique pour les dynamiques collectives

Laetitia Della Maestra

► To cite this version:

Laetitia Della Maestra. Estimation statistique pour les dynamiques collectives. Statistiques [math.ST]. Université Paris sciences et lettres, 2022. Français. NNT : 2022UPSLD053 . tel-04123861

HAL Id: tel-04123861

<https://theses.hal.science/tel-04123861>

Submitted on 9 Jun 2023

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-PSL

Estimation statistique pour les dynamiques en interaction

Soutenue par

Laetitia Della Maestra

Le 30 novembre 2023

École doctorale n°543

École Doctorale SDOSE

Spécialité

Mathématiques

Composition du jury :

Mme Fabienne Comte Professeur, Université Paris-Cité	<u>Présidente du jury</u>
M. Marc Hoffmann Professeur, Université Paris Dauphine-PSL	<u>Directeur de thèse</u>
Mme Eva Löcherbach Professeur, Université Paris 1 Panthéon-Sorbonne	<u>Rapporteur</u>
M. Mark Podolskij Professeur, Université du Luxembourg	<u>Rapporteur</u>
Mme Valentine Genon-Catalot Professeur, Université Paris-Cité	<u>Examineur</u>
M. Benjamin Jourdain Chercheur, CERMICS	<u>Examineur</u>
Mme Béatrice de Tilière Professeur, Université Paris Dauphine-PSL	<u>Examineur</u>

Cette thèse se situe à l'interface de la Statistique, paramétrique et non-paramétrique, des processus stochastiques, et de l'analyse des équations aux dérivées partielles paraboliques. Le fil conducteur des travaux présentés ici a été d'obtenir des informations sur le flot de mesures de probabilités, solution d'une équation de Fokker-Planck-Kolmogorov non-linéaire, dite équation de McKean-Vlasov, et sur le coefficient de drift de cette équation, à l'aide de l'observation continue de la solution du système linéaire de diffusions en interaction l'approchant en limite de champ moyen, *i.e.* quand le nombre de particules de ce système croît vers l'infini.

Dans une première partie, nous présentons les résultats obtenus quand le coefficient de drift dépend d'un paramètre inconnu. Nous suivons le programme d'estimation de Ibragimov et Hasminskii : nous montrons que la propriété de normalité locale asymptotique est vérifiée par ce modèle et nous en déduisons des propriétés fines de convergence et de normalité asymptotique de l'estimateur du maximum de vraisemblance de ce paramètre. Nous proposons un nouveau critère de vérification de la non-dégénérescence et de l'identifiabilité de ce problème statistique.

Dans une deuxième partie, nous présentons les résultats obtenus quand le coefficient de drift est inconnu, mais appartient à une certaine classe de fonctions caractérisées par leur régularité : les fonctions localement hölderiennes. Cette partie repose sur une nouvelle inégalité de concentration de type Bernstein, et sur des techniques de statistique non-paramétrique : nous construisons des estimateurs à noyaux adaptatifs de la densité solution de l'équation de McKean-Vlasov et de son drift à l'aide de la méthode de Goldenshluger-Lepski, et nous montrons que ces estimateurs sont optimaux au sens minimax pour un risque ponctuel. Enfin, nous construisons un estimateur de la force d'interaction dans le cas d'un drift de type Vlasov à l'aide d'une méthode de déconvolution.

Mots-Clés

Statistique paramétrique/non-paramétrique, Statistique des processus aléatoires, modèle de McKean-Vlasov, Systèmes de particules en interaction, Limite de Champ moyen, Propagation du chaos, Inégalités de concentration, Equation aux dérivées partielles paraboliques non-linéaires, Equation de Fokker-Planck-Kolmogorov non-linéaire, Espaces de Hölder, Estimation minimax adaptative, Normalité locale asymptotique, Programme d'Ibragimov-Hasminski, Estimateur du maximum de vraisemblance, Inégalités oracles, Estimateurs à noyaux, Sélection de fenêtres, Méthode de Goldenshluger-Lepski, Déconvolution.

This thesis lies at the crossroads of non-parametric and parametric Statistics of processes, and parabolic partial differential equations analysis. The aim of the present work was to obtain informations about the flow of probability measures solution of a nonlinear Fokker-Planck-Kolmogorov equation, known as McKean-Vlasov equation, and about the drift coefficient of this equation, thanks to the continuous observation of the solution of a linear system of interacting diffusions as an approximation in a mean-field limit, *i.e.* as the number of particles in this system grows to infinity.

In a first part, we present our results when the drift coefficient depends on an unknown parameter. We follow the estimation program of Ibragimov and Hasminskii : we show that our model verifies the local asymptotic normality property, and we deduce from it precise properties of convergence and asymptotic normality of the maximum likelihood estimator of the parameter. We give a new criterium for verifying that this statistical problem is non-degenerate and identifiable.

In a second part, we present our results when the drift coefficient is unknown, but lies in a certain class of functions characterized by their regularity : the local Hölder functions. This part relies on a new Bernstein type concentration inequality, and on nonparametric statistics tools : we build adaptive kernel estimators of the density solution of the McKean-Vlasov equation and of its drift, and we show that these estimators are optimal in a minimax sense for a pointwise risk. Finally, we construct an estimator of the interaction force in the case of a Vlasov type drift by means of a deconvolution method.

Keywords

Nonparametric statistics, Parametric statistics, Statistics of stochastic processes, McKean-Vlasov model, Interacting particle systems, Mean-field limit, Propagation of chaos, Concentration inequalities, Nonlinear Parabolic Partial differential equation, Nonlinear Fokker-Planck-Kolmogorov equation, Adaptive minimax estimation, Local Asymptotic Normality, Ibragimov-Hasminski's program, Maximum likelihood estimator, Oracle inequality, Kernel estimation, Bandwidth selection, Goldenshluger-Lepski's method, Deconvolution.

REMERCIEMENTS

Tout d'abord, un grand merci évidemment à mon directeur de thèse Marc Hoffmann : je te remercie de m'avoir fait confiance et de m'avoir donné l'opportunité de faire ce doctorat. Grâce à toi, j'ai appris de belles mathématiques, j'ai découvert le monde de la recherche, et j'ai maintenant la chance d'enseigner comme je le désirais.

Je suis très honorée qu'Eva Löcherbach et Mark Podolskij aient assuré la charge de rapporter ma thèse, et ce malgré des délais compliqués. Qu'ils en soient ici chaudement remerciés. Eva, merci beaucoup pour vos conseils et vos remarques qui m'ont beaucoup aidés pour la préparation de la soutenance.

Merci beaucoup à Fabienne Comte qui m'a fait l'honneur de présider mon jury de soutenance de thèse, ainsi qu'à Valentine Genon-Catalot et Benjamin Jourdain d'avoir accepté de prendre part à ce jury. Madame Comte, merci également pour votre cours sur l'estimation non-paramétrique. Madame Genon-Catalot, merci pour vos cours sur la statistique des diffusions et la statistique asymptotique.

Je tiens enfin à remercier chaleureusement Béatrice de Tilière, pour son implication auprès des doctorants en tant que directrice de l'Ecole Doctorale de Mathématiques de l'Université Paris Dauphine-PSL, mais également plus particulièrement pour le rôle qu'elle a joué auprès de moi en tant que *mentor* au cours de ma thèse. Tu as été pour moi un soutien constant et bienveillant, et je suis très heureuse que tu aies accepté de faire partie de mon jury.

Le laboratoire du CEREMADE fut un environnement scientifique propice à cette thèse, je remercie tous ses membres pour leur bienveillance à mon égard. Merci en particulier au directeur du CEREMADE, Vincent Rivoirard, au directeur du département MIDO et ex-directeur de l'Ecole Doctorale de Mathématiques de Dauphine Jacques Féjoz, sans qui je n'aurais sans doute jamais fait de thèse, à Pierre Cardaliaguet pour m'avoir fait découvrir Ladyzhenskaya-Solonnikov-Uraltseva, et à François Simenhaus pour avoir encadré mon mémoire de Master 2 sur les *MAMA*. Merci également à Alexandre Afgoustidis, Julien Poisat et Gabriel Turinici : ce fut un vrai plaisir d'enseigner dans vos matières en MIDO. Un grand merci enfin à Béatrice Baeza, Chantal Charlier et Ana Drumea pour leur gentillesse et pour m'avoir accompagnée dans mes différentes démarches administratives.

Un grand merci aux doctorants dauphinois de Mathématiques, cela a été un honneur d'être parmi vous durant ces quatre années et de vous représenter au Conseil de l'Ecole Doctorale. Merci en particulier aux *anciens*, Grégoire et Jean, aux *petits jeunes*, Charly, Ruihua, Marie-Julie, Grégoire & Victoria, et à mon prédécesseur Paulien pour son aide précieuse dans la dernière ligne droite. Et, *last but not least*, merci infiniment à mon *fratellino* Raphaël pour son soutien dans les moments difficiles !

Enfin, un grand merci aux étudiants dauphinois que j'ai eu le plaisir de voir en cours/TD/TP/stage ces quatre dernières années, ainsi qu'aux élèves des classes préparatoires BCPST du Lycée Janson de Sailly que j'ai eu le plaisir de voir en khôlles et en TD d'informatique un peu plus tôt au cours de ma jeune carrière d'enseignante.

Je vais maintenant remonter un peu le temps. . . Ayant fait (presque) toutes mes études, de la première année de Licence à la dernière année de Doctorat, à l'Université Paris Dauphine(-PSL), je tiens à remercier ici tous les enseignants que j'ai eus au cours de ces années. Merci en particulier à Guillaume Vigeral et Jimmy Lamboley pour avoir pris le temps de me conseiller sur mes études il y a déjà quelques années.

Ma plus profonde reconnaissance et mes chaleureux remerciements vont évidemment à Anne-Marie Boussion, pour ses cours d'une très grande clarté, ses conseils et son soutien constant.

Un grand merci également à mes professeurs de la préparation à l'Agrégation de Mathématiques de l'ENS Cachan qui seront sans doute surpris de voir ce que j'ai fait après l'agreg. . . Merci donc à Sandrine Dallaporta, Arthur Leclaire, Tuong-Huy Nguyen et Frédéric Pascal. Merci aussi à mes collègues agrégatifs qui ont égayé cette année de préparation, je pense en particulier à Chloé, Clément, Jean-Léon, et Julien.

Certains enseignants m'ont beaucoup marquée au cours de mes jeunes années et, chacun à leur manière, ont contribué à ma vocation pour l'enseignement : un grand merci donc à mes professeurs de Mathématiques de Première et Terminale, Mme Magail et M. Paul, ainsi qu'à M. Simon vrai conteur d'Histoire, Mmes Nataf et Wagner qui m'ont fait découvrir la Grèce, antique et actuelle, et Mme Rambaud, qui avait prédit que j'écirais un jour (mais je ne suis pas sûre qu'elle pensait à ce genre de récits!).

J'en viens naturellement au groupe de Canto Spontaneo de l'association Polimnia et *alle nostre care professoressa* Anna Andreotti et Margherita Trefoloni, ainsi qu'à notre présidente Francesca Perugini, et au chœur des Centraux, à notre cheffe Anastazija Marinkovic, à notre président Jean-Pierre Schiltz, ainsi qu'au groupe des Alti sympas : merci infiniment à tous pour votre accueil chaleureux et pour ces vraies bouffées d'oxygènes culturelles et artistiques. Un grand merci également à Loïc Berthevas pour son cours d'anglais accéléré qui m'a bien aidée dans la dernière ligne droite.

Pour finir, je souhaiterais remercier ma famille, les Della Maestra-Alciati... d'ici et d'ailleurs, *dal Piemonte e dal Friuli* (un bjel mandi alla famiglia Tonello!) et les Marthon-Combes-Chastanet-Bouysse-Vergne... de Corrèze, ainsi que mes amis qui m'ont soutenue pendant ces quatre ans et ont pardonné mes retards et autres ajournements : Camille, Cécile et leur famille, Francine et sa maman, Géraldine & Claire, Jennifer, Julien et Louisa, Jill - *Mother Christmas* and family, Marc, Morgane, Margaux, Nicolas & Alice, Enzo, Hélène & François, Karine & Jean, Thibault et Valentin, et à Petronela pour son soutien si important dans la dernière ligne droite.

Un grand merci également à Patrizia & Michel Bisson pour leur gentillesse et pour leur énergie inépuisable au service de l'association France-Frioul.

Mes remerciements les plus chaleureux vont, en tant que bébé jansonien, à Françoise et Laure pour leur soutien et pour l'amitié qu'elles me témoignent depuis maintenant si longtemps, et à Chantal & Henri Py sans qui, par transativité, je ne serais sans doute pas devenue enseignante.

Ainsi qu'à Catherine & Laurent, Danielle & Pierre, qui ont la grande gentillesse de prendre soin de Saphira-l'âne-Marthon pendant que nous sommes loin de la Corrèze.

Un grand merci à ma cousine Marie-Pierre et à ses amis de Créteil, Sylvie E. et Siegried I, à la famille Duron-Sallé, à la *Della connection* Thomas & Laurence, Perrine et Paulin, Olivier & Emilie, Evelyne & Jean-Paul, ainsi qu'à Jean & Jacqueline, Gisèle et sa famille, Ana, Nicolas, Viviana et la famille De Sousa, Alexandrine & Hervé, Victoria, Anne-Sophie et Jean-Louis, Colette & Bertin, qui a toujours encouragé ma vocation de matheuse, et la famille Diarra, ma marraine Danielle, Mamie, et ma petite soeur Flora.

Enfin, je souhaite remercier du fond du coeur Maman & Papa, qui m'ont portée à bout de bras depuis si longtemps et sans qui cette thèse n'aurait jamais été terminée (et qui ne sont pas pour rien non plus dans ma vocation d'enseignante!).

Mes plus affectueuses pensées vont à ceux avec qui j'aurais tant aimé partager ce tournant de ma vie :
Marie-Claire, Philippe, Fulvio et Papi.

*Melius est enim in via claudicare,
quam praeter viam fortiter ambulare.*
Saint Augustin, Sermon CXLI, v.4

TABLE DES MATIÈRES

Notations	1
Introduction	7
1 Présentation du modèle de McKean-Vlasov	7
1.1 Équation aux dérivées partielles de McKean-Vlasov	7
1.2 Équation différentielle stochastique de McKean-Vlasov	8
1.3 Interactions & Limites de champ moyen, Propagation du chaos	11
1.4 Historique & Applications	16
2 Hypothèses & Premières propriétés	19
2.1 Hypothèses générales sur (b, σ, μ_0)	19
2.2 Modèle statistique considéré	25
2.3 Quelques propriétés classiques & D'autres moins classiques	26
3 Objectifs statistiques & Résultats obtenus	35
3.1 Estimation paramétrique	36
3.2 Estimation non-paramétrique de μ^b	48
3.3 Estimation non-paramétrique de $b(\cdot, \cdot, \mu^b)$	58
3.4 Comparaison : paramétrique ou non-paramétrique ?	70
3.5 Estimation du noyau d'interaction dans le cadre non-paramétrique	72
3.6 Inégalité de concentration	74
4 Composition de la thèse	77
1 The LAN property for McKean-Vlasov models in a mean-field regime	79
1 Introduction	79
1.1 Motivation	79
1.2 Setting	80
1.3 Results and organisation of the paper	80
2 Construction and properties of the statistical model	83
2.1 Notation	83
2.2 Model assumptions	83
2.3 The companion McKean-Vlasov product experiment and its <i>linear</i> counterpart	86
2.4 Identifiability and non-degeneracy of the Fisher information	90

3	Main results	93
3.1	The LAN property	93
3.2	Maximum likelihood estimation and properties	94
4	Examples	95
4.1	McKean-like models	95
4.2	Generalised linear like models	97
4.3	A double layer potential model	98
4.4	A genuinely non-linear example	100
5	Proof of the main results	101
5.1	Preliminaries : couplings	101
5.2	Proof of Theorem 1.17	102
5.3	Proof of Theorem 1.19	105
6	Remaining proofs	109
6.1	Proof of Proposition 1.9	109
6.2	Proof of Proposition 1.10	110
6.3	Proof of Proposition 1.11	111
6.4	Proof of Proposition 1.16	111
7	Appendix	112
7.1	Proof of Lemma 1.7	112
7.2	Proof of Lemma 1.8	113
7.3	Proof of Lemma 1.20	113
7.4	Proof of Lemma 1.21	114
2	Nonparametric estimation for interacting particle systems : McKean-Vlasov models	117
1	Introduction	117
1.1	Setting	117
1.2	Motivation	118
1.3	Results and organisation of the paper	119
2	Model assumptions and construction of estimators	120
2.1	The system of interacting particles and its limit	120
2.2	Kernel estimators	123
3	Nonparametric oracle estimation	124
3.1	Oracle estimation of $\mu_t(x)$	124
3.2	Oracle estimation of $b(t, x, \mu_t)$	126
4	Adaptive minimax estimation	127
4.1	Anisotropic Hölder smoothness classes for McKean-Vlasov models	127
4.2	Minimax adaptive estimation of $\mu_t(x)$	128
4.3	Minimax adaptive estimation of $b(t, x, \mu_t)$	129
5	Estimation of the interaction in the Vlasov model	130
5.1	Identification of the interaction F	130
5.2	Consistent estimation of F	131
6	Probabilistic tools : a concentration inequality	132
6.1	A Bernstein inequality	132
6.2	Preparation for the proof of Theorem 2.18	133
6.3	Proof of Theorem 2.18	134
6.4	Proof of Proposition 2.19	136

7	Proof of the nonparametric estimation results	141
7.1	Proof of Theorem 2.7	141
7.2	Proof of Theorem 2.9	144
7.3	Proof of Theorem 2.14	154
7.4	Proof of Theorem 2.15	157
7.5	Proof of Theorem 2.17	159
8	Appendix	161
8.1	Proof of Lemma 2.20	162
8.2	Proof of Lemma 2.22	163
8.3	(Sketch of) proof of Proposition 2.13	166
Appendices		169
1	Existence, Uniqueness & Regularity	169
1.1	Existence, Unicity & Uniformly bounded moments	169
1.2	Regularity in time	171
1.3	Unique probability solution of the Cauchy problem	172
1.4	Regularity of the map $b \mapsto \mu^b$	173
1.5	Existence of a regular density of μ^b	174
1.6	Positivity & Local bounds of μ^b	175
1.7	Injectivity of $b \mapsto \mu^b$	176
1.8	More regularity in the Vlasov case	177
1.9	Continuity in the parameter ϑ	179
2	Solution of the particles system	184
2.1	Proof of Theorem 31(i) : $\mathbb{P}_b^N$ is well-defined	184
2.2	Proof of Theorem 31 (iv)	185
3	Distance in total variation between $\mathbb{P}_b^N$ and $\overline{\mathbb{P}}_b^{\otimes N}$	187
4	Discrete observations	188
4.1	Consistency of the AMLE : Proof of Theorem 59	188
4.2	Asymptotic normality of the AMLE : proof of Theorem 60	192

Throughout our present work, we will make a constant use of the following notations :

- ↪ $T > 0$ is the *time horizon* (fixed once for all) ;
- ↪ $d \geq 1$ is the *dimension* of the space where our particles will live and evolve ;
- ↪ $^\top$ will stand for the *transpose* of a vector or a matrix ;
- ↪ $\mathcal{M}_d(\mathbb{R})$ (resp. $\mathcal{S}_d(\mathbb{R})$, $\mathcal{S}_d^+(\mathbb{R})$, $\mathcal{S}_d^{++}(\mathbb{R})$) is the set of the d -dimensional square matrices (resp. the set of the d -dimensional *symmetric* square matrices, the set of the d -dimensional *symmetric non-negative* square matrices, the set of the d -dimensional *definite positive symmetric* square matrices) ; for $M \in \mathcal{M}_d(\mathbb{R})$, $\text{Tr}(M)$ (resp. $\det(M)$) stands for the *trace* (resp. the *determinant*) of M ; I_d stands for the d -dimensional *identity* matrix ;
- ↪ For some real numbers x, x' ,

$$\{x\}_+ := \max(x, 0) , \quad \{x\}_- := -\min(x, 0) , \quad |x| := \{x\}_+ + \{x\}_- , \quad \lfloor x \rfloor := \max\{n \in \mathbb{Z} ; n < x\} ;$$

$$x \wedge x' := \min(x, x') , \quad x \vee x' := \max(x, x') ;$$
- ↪ For some sets $A, B \subset \mathbb{R}^q, q \geq 1$, $A - B := \{x - y ; (x, y) \in A \times B\}$, and in particular, for $x_0 \in \mathbb{R}^q$, $x_0 - B := \{x_0\} - B$;
- ↪ For some positive integer ℓ , $\mathfrak{S}_\ell$ is the set of the permutations of $\{1, \dots, \ell\}$;
- ↪ For some finite set $S \subset \mathbb{R}^q, q \geq 1$, $\text{Card}(S)$ is the *cardinality* of S , that is to say the number of elements in S ;
- ↪ We say that $s = (s_1, \dots, s_d) \in \mathbb{N}^d$ is a *multi-index* ; we define the *order* of s by

$$|s| := s_1 + \dots + s_d ,$$

the *s-derivative* by

$$D^s := \frac{\partial^{|s|}}{\partial_1^{s_1} \dots \partial_d^{s_d}} , \quad \text{where } \partial_i = \partial_{x_i}, \quad 1 \leq i \leq d ,$$

and the *power-s* by

$$y^s := y_1^{s_1} \dots y_d^{s_d}, \quad y = (y_1, \dots, y_d) \in \mathbb{R}^d;$$

- $\Leftrightarrow$ let $U = [0, T]$ (or $U = (0, T)$), $U = \mathbb{R}^d$, $U = [0, T] \times \mathbb{R}^d$, $U = (0, T) \times \mathbb{R}^d$, for some integer $q \geq 1$, and some function $f : U \rightarrow \mathbb{R}^q$, we will denote by $f^1, \dots, f^q : U \rightarrow \mathbb{R}$ its *coordinate* functions, that is to say $f = (f^1, \dots, f^q)$, and we define $\mathcal{C}(U; \mathbb{R}^q)$ as the space of continuous functions from U to $\mathbb{R}^q$, $\mathcal{C}_b(U; \mathbb{R}^q)$ as the space of continuous and bounded functions from U to $\mathbb{R}^q$, $\mathcal{C}_c(U; \mathbb{R}^q)$ as the space of continuous functions from U to $\mathbb{R}^q$ with compact support in U ;
 - for $1 \leq k \leq \infty$, we define $\mathcal{C}^k(\mathbb{R}^d; \mathbb{R}^q)$ as the space of k -times continuously differentiable functions from $\mathbb{R}^d$ to $\mathbb{R}^q$ (that is to say, for every $s \in \mathbb{N}^d, |s| \leq k$, $D^s f^1, \dots, D^s f^q$ exist and are continuous on $\mathbb{R}^d$), we also define $\mathcal{C}_b^k(\mathbb{R}^d; \mathbb{R}^q)$ (resp. $\mathcal{C}_c^k(\mathbb{R}^d; \mathbb{R}^q)$) as the functions of $\mathcal{C}^k(\mathbb{R}^d; \mathbb{R}^q) \cap \mathcal{C}_b(\mathbb{R}^d; \mathbb{R}^q)$ (resp. $\mathcal{C}_c(\mathbb{R}^d; \mathbb{R}^q)$) such that every s -derivative, for $s \in \mathbb{N}^d, |s| \leq k$, is bounded (resp. the functions of $\mathcal{C}^k(\mathbb{R}^d; \mathbb{R}^q)$ with compact support in $\mathbb{R}^d$);
 - for $q = 1$, and $f \in \mathcal{C}^1(\mathbb{R}^d; \mathbb{R})$ (resp. $f \in \mathcal{C}^2(\mathbb{R}^d; \mathbb{R})$) we will denote by $\nabla f : x \in \mathbb{R}^d \mapsto (\partial_{x_1} f(x), \dots, \partial_{x_d} f(x))^\top$ the *gradient* of f (resp. by $D^2 f : x \in \mathbb{R}^d \mapsto [\partial_{x_i x_j}^2 f(x)]_{1 \leq i, j \leq d}$ the *Hessian* of f) and we will sometimes write ∇_x (resp. D_x^2) to precise that the differentiation is made with respect to the variable x , in the case where f is a function of more than just one variable;
 - for some integers n_1, n_2 , we define $\mathcal{C}^{n_1, n_2}((0, T) \times \mathbb{R}^d; \mathbb{R}^q)$, for some positive integer q , as the space of continuous functions $f \in \mathcal{C}((0, T) \times \mathbb{R}^d; \mathbb{R}^q)$, such that for every $s \in \mathbb{N}, 0 \leq s \leq n_1, s' \in \mathbb{N}^d, 0 \leq |s'| \leq n_2$, $\partial_t^s f, D_x^{s'} f$ exist and are in $\mathcal{C}((0, T) \times \mathbb{R}^d; \mathbb{R}^q)$; for J an interval such that $\bar{J} \subset (0, T)$ and U some ball of $\mathbb{R}^d$, we define $\mathcal{C}^{n_1, n_2}(J \times U; \mathbb{R}^q)$ in the same manner;
- $\Leftrightarrow$ Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space, and let X a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in some Polish space $(E, \mathcal{E})$. We will denote by $\mathcal{L}_{\mathbb{P}}(X)$ (or just $\mathcal{L}(X)$ if there is no possible misunderstanding) the *law* of X under $\mathbb{P}$, and for some probability law P on E , we will denote by $X \sim P$ the fact that X is of law P (under $\mathbb{P}$);
- $\Leftrightarrow$ For some integer $r \geq 1$, $\mathcal{P}_r(\mathbb{R}^d)$ is the set of *probability measures* ν on $\mathbb{R}^d$ with a *finite moment of order r* , that is to say such that

$$\mathfrak{m}_r(\nu) := \int_{\mathbb{R}^d} |y|^r \nu(dy) < \infty;$$

for $d = 1$, we will also use the moments

$$\tilde{\mathfrak{m}}_r(\nu) := \int_{\mathbb{R}} y^r \nu(dy);$$

- $\Leftrightarrow$ The space $\mathcal{P}_1(\mathbb{R}^d)$ is endowed with the *1-Wasserstein distance* $\mathcal{W}_1$ (also called the *Kantorovich-Rubinstein distance*, see the Chapter 6 of the Villani's monograph [185]) defined by : for $\nu, \nu' \in \mathcal{P}_1(\mathbb{R}^d)$,

$$\mathcal{W}_1(\nu, \nu') := \inf_{Z \sim \nu, Z' \sim \nu'} \mathbb{E}_{\mathbb{P}}[|Z - Z'|],$$

where the infimum is taken over all random variables Z, Z' defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathbb{R}^d$ respectively of laws ν and ν' ;

- ⇒ We will also consider the *total variation distance* $\|\cdot\|_{\text{TV}}$ defined on $\mathcal{P}_1(\mathbb{R}^d)$ by : for $\nu, \nu' \in \mathcal{P}_1(\mathbb{R}^d)$,

$$\|\nu' - \nu\|_{\text{TV}} := \sup_{B \in \mathcal{B}(\mathbb{R}^d)} |\nu'(B) - \nu(B)| = \inf_{Z \sim \nu, Z' \sim \nu'} \mathbb{P}(Z \neq Z'),$$

where $\mathcal{B}(\mathbb{R}^d)$ is the *Borel σ -algebra* of $\mathbb{R}^d$, and the infimum is taken over all random variables Z, Z' defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathbb{R}^d$ respectively of laws ν and ν' ;

- ⇒ We will also make use of the *2-Wasserstein distance* $\mathcal{W}_2$ on $\mathcal{P}_2(\mathbb{R}^d)$ (see again [185]) defined by : for $\nu, \nu' \in \mathcal{P}_2(\mathbb{R}^d)$,

$$\mathcal{W}_2(\nu, \nu') := \inf_{Z \sim \nu, Z' \sim \nu'} \mathbb{E}_{\mathbb{P}} \left[|Z - Z'|^2 \right]^{1/2},$$

where the infimum is taken over all random variables Z, Z' defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathbb{R}^d$ respectively of laws ν and ν' ;

- ⇒ We say that $f : [0, T] \rightarrow \mathbb{R}$ belongs to the *Hölder class* $\mathcal{H}^\alpha(t_0)$ for some $t_0 \in (0, T)$ and $\alpha > 0$, if there exists some relatively compact neighbourhood $\mathcal{U}_{t_0}$ of t_0 in $(0, T)$ such that $f \in \mathcal{C}^{[\alpha]}(\mathcal{U}_{t_0}; \mathbb{R})$ and $\partial_t^{[\alpha]} f$ is $(\alpha - [\alpha])$ -Hölder on $\mathcal{U}_{t_0}$, that is to say

$$[\partial_t^{[\alpha]} f]_{\alpha - [\alpha]; \mathcal{U}_{t_0}} := \sup_{t, t' \in \mathcal{U}_{t_0}, t \neq t'} \frac{|\partial_t^{[\alpha]} f(t') - \partial_t^{[\alpha]} f(t)|}{|t' - t|^{\alpha - [\alpha]}} < \infty;$$

without loss of generality we will consider $\mathcal{U}_{t_0}$ as an interval such that $\overline{\mathcal{U}_{t_0}} \subset (0, T)$, and we obtain a norm on $\mathcal{H}^\alpha(t_0)$ by setting

$$|f|_{\mathcal{H}^\alpha(t_0)} = |f|_{\mathcal{H}^\alpha(\mathcal{U}_{t_0})} := \sum_{s=0}^{[\alpha]} \sup_{t \in \mathcal{U}_{t_0}} |\partial_t^s f(t)| + [\partial_t^{[\alpha]} f]_{\alpha - [\alpha]; \mathcal{U}_{t_0}},$$

or by considering the equivalent norm

$$|f|_{\mathcal{H}^\alpha(t_0)} = \sup_{t \in \mathcal{U}_{t_0}} |f(t)| + \max \left(\max_{1 \leq s \leq [\alpha]} \sup_{t \in \mathcal{U}_{t_0}} |\partial_t^s f(t)|; [\partial_t^s f]_{\alpha - [\alpha]; \mathcal{U}_{t_0}} \right);$$

the extension of these definitions to $\mathbb{R}^q$ -valued functions, for some positive integer q , is straightforward by considering coordinate functions : a mapping $f = (f^k)_{1 \leq k \leq q} : [0, T] \rightarrow \mathbb{R}^q$ belongs to $\mathcal{H}^\alpha(t_0)$ if $f^k \in \mathcal{H}^\alpha(t_0)$ for every $k = 1, \dots, q$;

we say that f belongs to the *Hölder class* $\mathcal{H}_{\text{loc}}^\alpha((0, T))$ if for every $t_0 \in (0, T)$, $f \in \mathcal{H}^\alpha(t_0)$, that is to say for every interval J with $\bar{J} \subset (0, T)$, $f \in \mathcal{H}^\alpha(J)$ in the sense that $f \in \mathcal{C}^{[\alpha]}(J; \mathbb{R})$ and $\partial_t^{[\alpha]} f$ is $(\alpha - [\alpha])$ -Hölder on J , and we obtain a norm $|f|_{\mathcal{H}^\alpha(J)}$ on $\mathcal{H}^\alpha(J)$ by replacing $\mathcal{U}_{t_0}$ by J in the definition of $|f|_{\mathcal{H}^\alpha(\mathcal{U}_{t_0})}$;

- ⇒ We say that $f : \mathbb{R}^d \rightarrow \mathbb{R}$ belongs to the *Hölder class* $\mathcal{H}^\beta(x_0)$ for some $x_0 \in \mathbb{R}^d$ and $\beta > 0$, if there exists some relatively compact neighbourhood $\mathcal{U}_{x_0}$ of x_0 in $\mathbb{R}^d$ such that $f \in \mathcal{C}^{[\beta]}(\mathcal{U}_{x_0}; \mathbb{R})$ and for every $s \in \mathbb{N}^d$, $|s| = [\beta]$, $D_x^s f$ is $(\beta - [\beta])$ -Hölder on $\mathcal{U}_{x_0}$, that is to say

$$[D_x^s f]_{\beta - [\beta]; \mathcal{U}_{x_0}} := \sup_{x, x' \in \mathcal{U}_{x_0}, x \neq x'} \frac{|D_x^s f(x') - D_x^s f(x)|}{|x' - x|^{\beta - [\beta]}} < \infty;$$

without loss of generality we will consider $\mathcal{U}_{x_0}$ as a ball, and we obtain a norm on $\mathcal{H}^\beta(x_0)$ by setting

$$|f|_{\mathcal{H}^\beta(x_0)} = |f|_{\mathcal{H}^\beta(\mathcal{U}_{x_0})} := \sum_{s \in \mathbb{N}^d, 0 \leq |s| \leq \lfloor \beta \rfloor} \sup_{x \in \mathcal{U}_{x_0}} |D_x^s f(x)| + \sum_{s \in \mathbb{N}^d, |s| = \lfloor \beta \rfloor} [D_x^s f]_{\beta - \lfloor \beta \rfloor; \mathcal{U}_{x_0}},$$

or by considering the equivalent norm

$$|f|_{\mathcal{H}^\beta(x_0)} = \sup_{x \in \mathcal{U}_{x_0}} |f(x)| + \max \left(\max_{s \in \mathbb{N}^d, 1 \leq |s| \leq \lfloor \beta \rfloor} \sup_{x \in \mathcal{U}_{x_0}} |D_x^s f(x)|; \max_{s \in \mathbb{N}^d, |s| = \lfloor \beta \rfloor} [D_x^s f]_{\beta - \lfloor \beta \rfloor; \mathcal{U}_{x_0}} \right);$$

the extension of these definitions to $\mathbb{R}^q$ -valued functions, for some positive integer q , is straightforward by considering coordinate functions : a mapping $f = (f^k)_{1 \leq k \leq q} : \mathbb{R}^d \rightarrow \mathbb{R}^q$ belongs to $\mathcal{H}^\beta(x_0)$ if $f^k \in \mathcal{H}^\beta(x_0)$ for every $k = 1, \dots, q$;

we say that f belongs to the *Hölder class* $\mathcal{H}_{\text{loc}}^\beta(\mathbb{R}^d)$ if for every $x_0 \in \mathbb{R}^d$, $f \in \mathcal{H}^\beta(x_0)$, that is to say for every ball U of $\mathbb{R}^d$, $f \in \mathcal{H}^\beta(U)$ in the sense that $f \in \mathcal{C}^{\lfloor \beta \rfloor}(U; \mathbb{R})$ and, for every $s \in \mathbb{N}^d$, $|s| = \lfloor \beta \rfloor$, $D_x^s f$ is $(\beta - \lfloor \beta \rfloor)$ -Hölder on U , and we obtain a norm $|f|_{\mathcal{H}^\beta(U)}$ on $\mathcal{H}^\beta(U)$ by replacing $\mathcal{U}_{x_0}$ by U in the definition of $|f|_{\mathcal{H}^\beta(\mathcal{U}_{x_0})}$;

- ⇒ We say that $f : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ belongs to the *anisotropic Hölder class* $\mathcal{H}^{\alpha, \beta}(t_0, x_0)$ for some $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$ and $\alpha, \beta > 0$, if there exists some relatively compact neighbourhood $\mathcal{U}'_{x_0}$ of x_0 in $\mathbb{R}^d$ such that $f(\cdot, x) \in \mathcal{H}^\alpha(t_0)$ uniformly in $x \in \mathcal{U}'_{x_0}$, and there exists some relatively compact neighbourhood $\mathcal{U}'_{t_0}$ of t_0 in $(0, T)$ such that $f(t, \cdot) \in \mathcal{H}^\beta(x_0)$ uniformly in $t \in \mathcal{U}'_{t_0}$, and we obtain a norm on $\mathcal{H}^{\alpha, \beta}(t_0, x_0)$ by setting

$$|f|_{\mathcal{H}^{\alpha, \beta}(t_0, x_0)} := \sup_{x \in \mathcal{U}'_{x_0}} |f(\cdot, x)|_{\mathcal{H}^\alpha(t_0)} + \sup_{t \in \mathcal{U}'_{t_0}} |f(t, \cdot)|_{\mathcal{H}^\beta(x_0)};$$

the extension to $\mathbb{R}^q$ -valued functions, for some positive integer q , is straightforward : a mapping $f = (f^k)_{1 \leq k \leq q} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^q$ belongs to $\mathcal{H}^{\alpha, \beta}(t_0, x_0)$ if $f^k \in \mathcal{H}^{\alpha, \beta}(t_0, x_0)$ for every $k = 1, \dots, q$;

we say that f belongs to the *Hölder class* $\mathcal{H}_{\text{loc}}^{\alpha, \beta}((0, T) \times \mathbb{R}^d)$ if for every $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$, $f \in \mathcal{H}^{\alpha, \beta}(t_0, x_0)$, that is to say for every interval J such that $\bar{J} \subset (0, T)$, and every ball U of $\mathbb{R}^d$, $f \in \mathcal{H}^{\alpha, \beta}(J \times U)$ in the sense that $f \in \mathcal{C}^{\lfloor \alpha \rfloor, \lfloor \beta \rfloor}(J \times U; \mathbb{R})$, $\partial_t^{\lfloor \alpha \rfloor} f$ is $(\alpha - \lfloor \alpha \rfloor)$ -Hölder in t on J , uniformly with respect to $x \in U$, that is to say

$$[\partial_t^{\lfloor \alpha \rfloor} f]_{\alpha - \lfloor \alpha \rfloor; J}^{\infty; U} := \sup_{x \in U} [\partial_t^{\lfloor \alpha \rfloor} f(\cdot, x)]_{\alpha - \lfloor \alpha \rfloor; J} = \sup_{x \in U, t, t' \in J, t \neq t'} \frac{|\partial_t^{\lfloor \alpha \rfloor} f(t', x) - \partial_t^{\lfloor \alpha \rfloor} f(t, x)|}{|t' - t|^{\alpha - \lfloor \alpha \rfloor}} < \infty,$$

for every $s \in \mathbb{N}^d$, $|s| = \lfloor \beta \rfloor$, $D_x^s f$ is $(\beta - \lfloor \beta \rfloor)$ -Hölder in x on U , uniformly with respect to $t \in J$, that is to say

$$[D_x^s f]_{\beta - \lfloor \beta \rfloor; U}^{\infty; J} := \sup_{t \in J} [D_x^s f(t, \cdot)]_{\beta - \lfloor \beta \rfloor; U} = \sup_{t \in J, x, x' \in U, x \neq x'} \frac{|D_x^s f(t, x') - D_x^s f(t, x)|}{|x' - x|^{\beta - \lfloor \beta \rfloor}} < \infty,$$

and so, we obtain a norm $|\cdot|_{\mathcal{H}^{\alpha, \beta}(J \times U)}$ on $\mathcal{H}^{\alpha, \beta}(J \times U)$ by setting

$$\begin{aligned} |f|_{\mathcal{H}^{\alpha, \beta}(J \times U)} &:= \sum_{s=0}^{\lfloor \alpha \rfloor} \sup_{(t, x) \in J \times U} |\partial_t^s f(t, x)| + \sum_{s \in \mathbb{N}^d, 0 \leq |s| \leq \lfloor \beta \rfloor} \sup_{(t, x) \in J \times U} |D_x^s f(t, x)| \\ &\quad + [\partial_t^{\lfloor \alpha \rfloor} f]_{\alpha - \lfloor \alpha \rfloor; J}^{\infty; U} + \sum_{s \in \mathbb{N}^d, |s| = \lfloor \beta \rfloor} [D_x^s f]_{\beta - \lfloor \beta \rfloor; U}^{\infty; J}; \end{aligned}$$

- $\Leftrightarrow$ We say that $f : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ belongs to the *parabolic Hölder space* $\mathcal{C}_{\text{loc}}^{1+\frac{\delta}{2}, 2+\delta}((0, T) \times \mathbb{R}^d)$ (see *e.g.* [113] and [118]) if $f \in \mathcal{C}^{1,2}((0, T) \times \mathbb{R}^d)$ and $\partial_t f, \partial_{x_i x_j}^2 f \in \mathcal{C}_{\text{loc}}^{\frac{\delta}{2}, \delta}((0, T) \times \mathbb{R}^d)$.
- $\Leftrightarrow$ We finally introduce the following *parabolic Sobolev spaces* (see sections 1.1. and 6.2. of [17]) : for an interval J of $(0, T)$ and an open ball U of $\mathbb{R}^d$, for $\delta \in [1, \infty)$, we say that $f : J \times U \rightarrow \mathbb{R}$ is in $\mathbb{H}^{1,\delta}(J \times U)$ if for almost every $t \in J$, $f(t, \cdot)$ is in the usual *Sobolev space* $W^{1,\delta}(U)$ (that is to say $f(t, \cdot) \in L^\delta(U)$ and $f(t, \cdot)$ has *generalized* derivatives $\partial_{x_1} f(t, \cdot), \dots, \partial_{x_d} f(t, \cdot)$ which are themselves in $L^\delta(U)$), and

$$\|f\|_{\mathbb{H}^{1,\delta}(J \times U)} := \left(\int_J (\|f(t, \cdot)\|_{L^\delta(U)} + \sum_{i=1}^d \|\partial_{x_i} f(t, \cdot)\|_{L^\delta(U)})^\delta dt \right)^{1/\delta} < \infty ; \quad (1)$$

we denote by $\mathbb{H}^{-1,\delta}(J \times U)$ the *dual space* of $\mathbb{H}_0^{1, \frac{\delta}{\delta-1}}(J \times U)$ where for $\delta' \in [1, \infty)$, $\mathbb{H}_0^{1,\delta'}(J \times U)$ is the set of functions in $\mathbb{H}^{1,\delta'}(J \times U)$ with $W_0^{1,\delta'}(U)$ (the closure of $\mathcal{C}_c^\infty(U; \mathbb{R})$ in $W^{1,\delta'}(U)$) instead of $W^{1,\delta'}(U)$ in the definition ;
 we say that f is in $\mathfrak{H}^{1,\delta}(J \times U)$ if $f \in \mathbb{H}^{1,\delta}(J \times U)$, $\partial_t f$ exists (in the generalized sense) and is in $\mathbb{H}^{-1,\delta}(J \times U)$ (see also *e.g.* [70]) ;

- $\Leftrightarrow$ For $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\nu \in \mathcal{P}_1(\mathbb{R}^d)$, the *convolution between f and ν* is defined by

$$(f \star \nu)(x) := \left(\int_{\mathbb{R}^d} f^k(x-y) \nu(y) dy \right)_{1 \leq k \leq d} ;$$

- $\Leftrightarrow$ For $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ with $f^k \in L^1(\mathbb{R}^d)$, $1 \leq k \leq d$, the *Fourier transform of f* is defined by

$$\mathcal{F}(f)(\xi) := \left(\int_{\mathbb{R}^d} e^{-2i\pi \xi^\top x} f^k(x) dx \right)_{1 \leq k \leq d}, \quad \xi \in \mathbb{R}^d ;$$

1 Présentation du modèle de McKean-Vlasov

Nous présentons dans cette section notre cadre de travail dans cette thèse : le *modèle de McKean-Vlasov*. Nous pourrions même parler d'écosystème, puisqu'il peut s'entendre au sens d'une *équation différentielle stochastique*, ou bien d'une *équation aux dérivées partielles* qui lui est d'une certaine manière équivalente, et que l'on peut également lui adjoindre un *système d'équations différentielles stochastiques* composé de N *processus de diffusions en interaction* qui en constitue une approximation en *limite de champ moyen*, c'est-à-dire quand le nombre de diffusions N devient très grand. Dans la Section 2, nous donnerons les hypothèses précises sous lesquelles nous nous placerons tout au long de ce manuscrit, puis nous montrerons que sous ces dernières les objets mathématiques manipulés ci-dessous existent bel et bien, et nous énoncerons les propriétés probabilistes et analytiques préalables à notre étude statistique du modèle de McKean-Vlasov.

1.1 Équation aux dérivées partielles de McKean-Vlasov

Cette thèse a pour objet l'étude statistique du *problème de Cauchy* suivant constitué d'une *équation aux dérivées partielles* (EDP) *parabolique*, dite de *Fokker-Planck-Kolmogorov*, et d'une condition initiale :

$$\begin{cases} \partial_t \mu_t = \frac{1}{2} \sum_{i,j=1}^d \partial_{x_i x_j}^2 ((\sigma \sigma^\top)(t, \cdot)_{ij} \mu_t) - \sum_{i=1}^d \partial_{x_i} (b^i(t, \cdot, \mu_t) \mu_t), & (t, x) \in (0, T) \times \mathbb{R}^d, \\ \mu_{|t=0} = \mu_0, \end{cases} \quad (2)$$

où l'horizon de temps $T > 0$ est fixé, ainsi que la dimension $d \geq 1$ de l'espace, la fonction de *diffusion* est $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathcal{M}_d(\mathbb{R})$, la fonction de *drift* est $b : [0, T] \times \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}^d$, et la condition initiale est $\mu_0 \in \mathcal{P}_1(\mathbb{R}^d)$. L'EDP du problème (2) est dite *non-linéaire* car le drift b dépend de μ_t .

Nous nous placerons dans notre étude sous des hypothèses classiques pour le résolution de ces problèmes : intégrabilité forte de μ_0 , stricte uniforme ellipticité de $c := \sigma \sigma^\top$, lipschitziannité de b et σ (les conditions précises seront détaillées dans la Section 2.1).

Nous nous inspirerons fortement de la monographie de Bogachev, Krylov, Röckner et Shaposhnikov [17], en particulier de ses Chapitres 6 à 9, pour notre étude analytique de cette EDP.

Nous considérerons dans un premier temps ce problème dans un *sens faible, sur les mesures* (voir Section 2.3) : nous montrerons que ce problème admet une unique solution μ , dite *solution de probabilité* selon la terminologie de [17], qui est une mesure borélienne sur $[0, T] \times \mathbb{R}^d$ de la forme $\mu = (\mu_t)_{t \in [0, T]}$ i.e. que $\mu(dt, dx) = \mu_t(dx)dt$, et telle que pour tout $t \in [0, T]$, μ_t est une mesure de probabilité sur $\mathbb{R}^d$. Nous noterons $\mathcal{S}$ l'application qui au triplet (b, c, μ_0) associe la solution μ . Nous verrons ensuite que, sous nos hypothèses, pour tout $t \in (0, T)$, μ_t admet une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$, que nous noterons par simplicité également μ_t , et que celle-ci vérifie bien sûr l'EDP (2), au *sens faible, sur les fonctions*. Nous démontrerons plusieurs propriétés d'intégrabilité et de régularité vérifiées par cette solution (certaines de ces propriétés sont élémentaires et bien connues, et nous ne les démontrerons que par souci d'exhaustivité, d'autres sont, à notre connaissance, originales dans le cadre choisi), et nous énoncerons également des conditions suffisantes pour que μ soit solution au *sens classique, sur les fonctions* du problème (2).

Le premier exemple d'étude d'une telle équation est donné par l'article fondateur de McKean de 1966 [146], dans lequel l'auteur se place dans le cadre $d = 1$, $\sigma(t, x) \equiv 1$, $b(t, x, \nu) = \int_{\mathbb{R}} (y - x) \nu(dy)$, autrement dit dans lequel le problème (2) se réécrit

$$\begin{cases} \partial_t \mu_t = \frac{1}{2} \Delta \mu_t - \partial_x \left(\left(\int_{\mathbb{R}^d} (y - \cdot) \mu_t(dy) \right) \mu_t \right), & (t, x) \in (0, T) \times \mathbb{R}^d, \\ \mu_{|t=0} = \mu_0, \end{cases} \quad (3)$$

Nous utiliserons tout au long de cette introduction comme exemple simple le modèle suivant, généralisation du modèle de McKean :

$$d \geq 1, \sigma(t, x) \equiv \sigma I_d, \sigma > 0, b(t, x, \nu) = (\vartheta_1 x + \vartheta_2) + \vartheta_3 \int_{\mathbb{R}^d} (y - x) \nu(dy). \quad (4)$$

La fonction de drift b dépend donc ici d'un paramètre $\vartheta = (\vartheta_1, \vartheta_2, \vartheta_3) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}$, et l'expression de μ peut alors être obtenue explicitement (voir (9)).

En effet, cette thèse traite à la fois du cadre *paramétrique*, dans lequel nous supposons que $b = b_0(\vartheta; \cdot)$, avec $\vartheta \in \Theta$ où Θ est un sous-ensemble connu compact de $\mathbb{R}^p$, $p \geq 1$ quelconque, et b_0 est une fonction supposée connue à ϑ près, et du cadre *non-paramétrique*, dans lequel b est supposée complètement inconnue. Nous noterons $\mu = \mu^b$ quand nous voudrions mettre en évidence la dépendance de μ en la fonction de drift b . En particulier, dans le cadre paramétrique, nous noterons $\mu = \mu^\vartheta$ pour mettre en évidence la dépendance en le paramètre ϑ .

Dans la sous-section suivante, nous allons donner, suivant en cela les idées séminales de l'article [146], une interprétation probabiliste du problème (2) sous forme d'*équation différentielle stochastique* (EDS) dite de *McKean-Vlasov* : par métonymie, nous nommerons donc également EDP de McKean-Vlasov l'équation du problème (2).

1.2 Équation différentielle stochastique de McKean-Vlasov

Le problème (2) a une contrepartie stochastique, que nous allons définir dans un cadre canonique qui sera utile tout au long de ce manuscrit. Soit $\mathcal{C} = \mathcal{C}([0, T]; \mathbb{R}^d)$ l'espace des fonctions continues

sur $[0, T]$ à valeurs dans $\mathbb{R}^d$, muni de la filtration $(\mathcal{F}_t)_{0 \leq t \leq T}$ induite par les applications canoniques $X_t : \omega \in \mathcal{C} \mapsto \omega(t) \in \mathbb{R}^d$. Nous appellerons processus canonique sur $(\mathcal{C}, \mathcal{F}_T)$

$$X = (X_t)_{t \in [0, T]} : \omega \in \mathcal{C} \mapsto \omega = (\omega(t))_{t \in [0, T]} \in \mathcal{C}.$$

Pour toute condition initiale μ_0 et fonction de diffusion σ fixées, pour une fonction de drift b , nous noterons $\bar{\mathbb{P}}_b$ l'unique mesure de probabilité sur $(\mathcal{C}, \mathcal{F}_T)$ sous laquelle $X_0 \sim \mu_0$ et le processus

$$(\bar{B}_t^b)_{t \in [0, T]} := \left(\int_0^t c^{-1/2}(s, X_s) (dX_s - b(s, X_s, \mathcal{L}_{\bar{\mathbb{P}}_b}(X_s)) ds) \right)_{t \in [0, T]}$$

est un mouvement brownien standard sur $\mathbb{R}^d$ ($c^{-1/2}$ désignant une des racines carrées de c^{-1} , son choix ne modifiant pas le problème, voir *e.g.* la partie 5.3 de [171]). Il s'agit donc d'un problème de type *point fixe* : nous admettons pour le moment que la famille $(\bar{\mathbb{P}}_b)_b$ est bien définie sous nos hypothèses (voir le Théorème 25 (i)), et nous le prouverons rigoureusement dans l'Annexe 1.1. En particulier, le processus canonique X sur $(\mathcal{C}, \mathcal{F}_T)$ est solution, sous $\bar{\mathbb{P}}_b$, de l'équation de McKean-Vlasov

$$\begin{cases} dX_t = b(t, X_t, \mathcal{L}_{\bar{\mathbb{P}}_b}(X_t))dt + \sigma(t, X_t)d\bar{B}_t^b, & t \in [0, T], \\ \mathcal{L}(X_0) = \mu_0. \end{cases} \quad (5)$$

Par définition, $\mathcal{L}_{\bar{\mathbb{P}}_b}(X) = \bar{\mathbb{P}}_b$, et nous noterons, pour tout $t \in [0, T]$, $\mu_t^b = \mathcal{L}_{\bar{\mathbb{P}}_b}(X_t)$.

Dans le cadre paramétrique pour lequel $b = b_0(\vartheta; \cdot)$, nous noterons également $\bar{\mathbb{P}}_\vartheta$ à la place de $\bar{\mathbb{P}}_b$, et μ_t^ϑ à la place de μ_t^b .

La particularité de l'équation (5) par rapport à une EDS usuelle est que le drift ne dépend plus seulement de X_t mais également de la loi de X_t : c'est ce que l'on appelle, en hommage à [146], une *non-linéarité au sens de McKean*. La forme de cette fonction de drift

$$b : (t, x, \nu) \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \mapsto b(t, x, \nu) \in \mathbb{R}^d$$

permet donc de modéliser une *interaction* entre une particule, via sa position $x \in \mathbb{R}^d$, et une loi de probabilité, $\nu \in \mathcal{P}_1(\mathbb{R}^d)$: dans l'équation (5), la particule X_t interagit ainsi avec sa propre loi, $\mathcal{L}_{\bar{\mathbb{P}}_b}(X_t)$.

En définissant $\beta^b : (t, x) \in [0, T] \times \mathbb{R}^d \mapsto b(t, x, \mu_t^b)$, nous pouvons également dire que, sous $\bar{\mathbb{P}}_b$, X est solution de l'EDS *usuelle* (autrement dit *linéaire au sens de McKean*)

$$\begin{cases} dX_t = \beta^b(t, X_t)dt + \sigma(t, X_t)d\bar{B}_t^b, & t \in [0, T], \\ \mathcal{L}(X_0) = \mu_0. \end{cases} \quad (6)$$

Le lien avec l'équation (2) s'établit de la manière habituelle suivante pour les équations de Fokker-Planck-Kolmogorov. Prenons une fonction test régulière $\Phi \in \mathcal{C}_c^\infty(\mathbb{R}^d; \mathbb{R})$. D'après la formule d'Itô (voir *e.g.* [171]), nous avons, pour tout $t \in [0, T]$,

$$\Phi(X_t) = \Phi(X_0) + \int_0^t \left(\frac{1}{2} \sum_{i,j=1}^d \int_0^s \partial_{ij}^2 \Phi(X_s) c_{ij}(s, X_s) + \sum_{i=1}^d \partial_i \Phi(X_s) b^i(s, X_s, \mu_s^b) \right) ds$$

$$+ \sum_{i=1}^d \int_0^t \partial_i \Phi(X_s) \sigma_{i.}(s, X_s) d\bar{B}_s^b ,$$

d'où, en passant à l'espérance $\mathbb{E}_{\tilde{\mathbb{P}}_b}$,

$$\int_{\mathbb{R}^d} \Phi d\mu_t^b = \int_{\mathbb{R}^d} \Phi d\mu_0 + \int_0^t \int_{\mathbb{R}^d} \left(\frac{1}{2} \sum_{i,j=1}^d \int_0^t \partial_{ij}^2 \Phi(x) c_{ij}(s, x) + \sum_{i=1}^d \partial_i \Phi(x) b^i(s, x, \mu_s^b) \right) \mu_s^b(dx) ds ,$$

ce qui signifie exactement (voir la Définition 24) que μ^b est solution de probabilité du problème (2).

Cette interprétation probabiliste du problème (2) était déjà présente dans l'article de McKean [146] : il en déduisait que la solution μ de (3) est la densité de l'unique solution forte, sur un espace de probabilité filtré quelconque $(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{t \in [0, T]}, \tilde{\mathbb{P}})$ donné, muni d'un mouvement brownien d -dimensionnel $\tilde{B}$ et d'une variable aléatoire $\tilde{X}_0$ de loi μ_0 , $\tilde{\mathcal{F}}_0$ -mesurable, et indépendante de $\tilde{B}$, de

$$\begin{cases} d\tilde{X}_t = (\mathbb{E}_{\tilde{\mathbb{P}}}[\tilde{X}_t] - \tilde{X}_t)dt + d\tilde{B}_t, & t \in [0, T], \\ \tilde{X}_{|t=0} = \tilde{X}_0. \end{cases} \quad (7)$$

Cette EDS, qui est un cas particulier d'EDS d'*Ornstein-Uhlenbeck* (voir *e.g.* l'Exemple 5.6.8 de [110]), a pour solution explicite

$$\tilde{X}_t = e^{-t} \tilde{X}_0 + (1 - e^{-t}) \tilde{\mathbf{m}}_1(\mu_0) + \int_0^t e^{-(t-s)} d\tilde{B}_s, t \in [0, T],$$

qui est donc un processus gaussien conditionnellement à $\tilde{X}_0$ sous $\tilde{\mathbb{P}}$, ce dont McKean déduit que, pour tout $(t_0, x_0) \in (0, T) \times \mathbb{R}$,

$$\mu_{t_0}(x_0) = \int_{\mathbb{R}} \frac{1}{\sqrt{\pi(1 - e^{-2t_0})}} \exp \left(- \frac{(x_0 - e^{-t_0}y - (1 - e^{-t_0})\tilde{\mathbf{m}}_1(\mu_0))^2}{1 - e^{-2t_0}} \right) \mu_0(dy) .$$

Nous pouvons étendre son raisonnement à notre exemple (4),

$$d \geq 1, \sigma(t, x) \equiv \sigma I_d, \sigma > 0, b(t, x, \nu) = b_0(\vartheta; t, x, \nu) := (\vartheta_1 x + \vartheta_2) + \vartheta_3 \int_{\mathbb{R}^d} (y - x) \nu(dy) .$$

$\mu^b = \mu^\vartheta$ est alors la densité de l'unique solution forte de

$$\begin{cases} d\tilde{X}_t = (\vartheta_1 \tilde{X}_t + \vartheta_2)dt + \vartheta_3 (\mathbb{E}_{\tilde{\mathbb{P}}}[\tilde{X}_t] - \tilde{X}_t)dt + \sigma d\tilde{B}_t, & t \in [0, T], \\ \tilde{X}_{|t=0} = \tilde{X}_0. \end{cases} \quad (8)$$

Sous certaines conditions sur ϑ et $\tilde{\mathbf{m}}_1(\mu_0)$ (voir la Section 4.1 du Chapitre 1), en notant pour $t \in (0, T)$

$$\tilde{m}(t) := \frac{\vartheta_2}{\vartheta_1} (e^{\vartheta_1 t} - 1) + \tilde{\mathbf{m}}_1(\mu_0) e^{\vartheta_1 t} (1 - e^{-\vartheta_3 t}), \tilde{\sigma}(t) := \frac{\sigma^2}{2(\vartheta_1 - \vartheta_3)} (e^{2(\vartheta_1 - \vartheta_3)t} - 1) ,$$

nous avons, sous $\tilde{\mathbb{P}}$, pour tout $t_0 \in (0, T)$, sachant $\tilde{X}_0$,

$$\tilde{X}_{t_0} \sim \mathcal{N}_d \left(e^{(\vartheta_1 - \vartheta_3)t_0} \tilde{X}_0 + \tilde{m}(t_0), \tilde{\sigma}(t_0) I_d \right),$$

ce qui implique que, pour tout $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$,

$$\mu_{t_0}^\vartheta(x_0) = \int_{\mathbb{R}^d} \frac{1}{(2\pi\tilde{\sigma}(t_0))^{d/2}} \exp \left(- \frac{|x_0 - e^{(\vartheta_1 - \vartheta_3)t_0} y - \tilde{m}(t_0)|^2}{2\tilde{\sigma}(t_0)} \right) \mu_0(dy). \quad (9)$$

Nous disposons en fait ici d'un des seuls cas pour lesquels l'application $\mathcal{S}$ est explicite (voir [9] pour quelques autres cas particuliers en champ moyen). Elle est de plus injective en ϑ , à σ, μ_0 fixés (voir également le Théorème 25 (ix) pour ces questions dans le cadre général).

Il convient cependant de préciser que, dans le cas sans interaction, *i.e.* lorsque $b(t, x, \nu) = b(t, x)$ ne dépend pas de la mesure, si c est de plus assez régulière, nous disposons de formules explicites pour μ^b et donc pour $\mathcal{S}$ (voir *e.g.* Genon-Catalot et Jacod [80]), formules qui deviennent même assez simples lorsque $d = 1$, $b(t, x) = b(x)$, $c(t, x) = c(x)$ et μ_0 est la distribution invariante du problème (sous réserve qu'elle existe). Cependant, nous nous concentrerons dans cette thèse sur le cadre de champ moyen, et donc avec interaction, ce qui entraîne qu'en toute généralité l'application $\mathcal{S}$ ne peut pas s'exprimer explicitement.

Dans la sous-section suivante, nous allons introduire les *observations* que nous avons utilisées pour procéder à l'estimation statistique des fonctions μ^b et $b(\cdot, \cdot, \mu^b)$. Pour cela, nous allons inverser notre point de vue sur l'équation de McKean-Vlasov : nous n'allons plus temporairement la considérer *per se*, mais au contraire comme l'équation limite en un certain sens d'un système constitué d'un grand nombre d'équations.

1.3 Interactions & Limites de champ moyen, Propagation du chaos

Genèse de ces concepts

Le processus stochastique solution de (5) peut être approché en *limite de champ moyen* (ou *mean-field limit*) par un *système de particules en interaction* quand le nombre N de particules tend vers l'infini. C'est d'ailleurs cet aspect simplificateur de (5), qui revient à étudier l'évolution d'une unique particule plutôt que celle d'un très grand nombre de particules interagissant, qui est à l'origine de son élaboration. Précisons avant toute chose que le terme *particule* sera compris dans cette thèse au sens générique d'agent, d'individu, d'unité.

Les grands systèmes de particules en interaction sont en effet présents dans de nombreux domaines : en Physique de l'infiniment petit (particules élémentaires, ions, atomes, molécules, ...), en Physique de l'infiniment grand (corps célestes, galaxies, ...), en Biologie (dynamiques collectives d'animaux) et, à l'échelle humaine, en Sciences Economiques et Sociales (mouvements de foules, interactions d'agents financiers, propagation d'idées ou de comportements au sein d'une population, ...). Des modèles *microscopiques*, c'est-à-dire à l'échelle des particules considérées, permettent de représenter les interactions au sein de ces systèmes : nous pouvons par exemple penser aux équations de Newton ou de Liouville. Malheureusement, ceux-ci sont d'autant plus compliqués à traiter analytiquement et computationnellement que le nombre de particules N augmente... et celui-ci peut être gigantesque ! Par exemple, la constante d'Avogadro représentant le nombre de

particules dans une mole de matière est de l'ordre de 10^{23} . Nous faisons alors face à ce que l'on peut appeler le *fléau de la dimensionnalité*.

Le problème crucial est donc de trouver un moyen de réduire la complexité du modèle, en réduisant sa dimensionnalité, tout en gardant une représentation aussi proche que possible du modèle initial et de la réalité. La stratégie classique pour atteindre un tel objectif est de déduire du modèle *microscopique* un modèle *macroscopique*, c'est-à-dire une description continue des dynamiques où l'information est contenue dans des densités, typiquement solutions d'EDP non-linéaires, comme l'EDP de McKean-Vlasov.

L'idée d'une telle description de ces systèmes remonte aux premiers travaux de Physique Statistique et de Cinétique des gaz dilués au XIXème siècle, à l'hypothèse dite du *chaos moléculaire* (ou du *Stosszahlansatz* selon la terminologie de P. et T. Ehrenfest [69]) introduite par Boltzmann en 1872 dans [22], selon laquelle les vitesses de deux particules de gaz qui entrent en collision sont a priori non corrélées, hypothèse développée dans la continuité des travaux de Carnot, Thomson, Clausius et Maxwell. Cependant, cette hypothèse est restée pendant longtemps (et reste encore dans de nombreux cas ...) difficile à formuler rigoureusement et à démontrer.

En 1907, Weiss introduisit quant à lui la notion de *champ moléculaire* et les *limites de champ moyen* pour le modèle dit de Curie-Weiss des matériaux ferromagnétiques (voir [188]), dans des régimes *sans collisions* (*i.e.* dans lequel les collisions surviennent rarement et les particules interagissent les unes avec les autres à longue distance). De 1915 à 1919, Jeans réinterpréta cette notion de champ moyen dans le cadre de l'étude des dynamiques galactiques ([103, 104]). De manière indépendante, Vlasov parvint en 1938 dans son article [187] à la même équation que Jeans, mais dans le cadre de l'étude des plasmas, équation qui porte depuis le nom d'*équation de Vlasov* (voir aussi [186]).

En 1956, dans son article *Foundations of Kinetic Theory* [109], Kac se pencha sur le problème de la justification des équations de la Physique Statistique, en particulier de l'équation de Boltzmann, à partir des équations microscopiques. Il développa dans ce but un programme global de la Physique Statistique : il proposa d'étudier des modèles microscopiques plus simples (en particulier les *équations maîtresses*), où l'hypothèse de *chaos moléculaire* est non seulement plus facile à formuler rigoureusement, mais susceptible d'être démontrée mathématiquement, et montra que, pour certains de ces modèles simples, *la propriété de Boltzmann se propage dans le temps* : c'est ce que l'on appellera la *propagation du chaos*.

Enfin, McKean dans son article [146] développa un analogue stochastique à l'équation de Vlasov, créant ainsi ce qu'il sera ensuite convenu d'appeler le *modèle de McKean-Vlasov*, et démontra la propriété de propagation du chaos pour son modèle dans [147].

Présentation de notre système de particules en interaction

Introduisons maintenant rigoureusement notre système de diffusions en interaction. Soit $\mathcal{C}^N := \mathcal{C}([0, T]; (\mathbb{R}^d)^N)$ l'espace des fonctions continues sur $[0, T]$ à valeurs dans $(\mathbb{R}^d)^N$, muni de la filtration $(\mathcal{F}_t^N)_{0 \leq t \leq T}$ induite par les applications canoniques $X_t^i : \omega \in \mathcal{C}^N \mapsto \omega^i(t)$. Nous appellerons processus canonique sur $\mathcal{C}^N$, l'application

$$\mathbb{X}^N := (X^1, \dots, X^N) : \omega \in \mathcal{C}^N \mapsto (\omega^i(t))_{1 \leq i \leq N, t \in [0, T]} \in \mathcal{C}^N.$$

Pour toute condition initiale $\mu_0^{\otimes N}$ et toute fonction de diffusion σ fixées, pour une fonction de drift b , nous noterons $\mathbb{P}_b^N$ l'unique probabilité sur $(\mathcal{C}^N, \mathcal{F}_T^N)$ sous laquelle $\mathbb{X}_0^N \sim \mu_0^{\otimes N}$ et le processus

$$\mathbb{B}^{N,b} = (B_t^{i,b})_{1 \leq i \leq N, t \in [0, T]} := \left(\int_0^t c^{-1/2}(s, X_s^i) (dX_s^i - b(s, X_s^i, \mu_s^N) ds) \right)_{1 \leq i \leq N, t \in [0, T]}$$

est un mouvement brownien standard sur $(\mathbb{R}^d)^N$; $(\mu_t^N)_{t \in [0, T]}$ désigne ici la *mesure empirique* du système de particules $(\mathbb{X}_t^N)_{t \in [0, T]}$, *i.e.* que $(\mu_t^N)_{t \in [0, T]}$ est la famille de mesures aléatoires définies sur $(\mathcal{C}^N, \mathcal{F}_T^N)$ et à valeurs dans $\mathcal{P}_1(\mathbb{R}^d)$, telle que pour tout $t \in [0, T]$

$$\mu_t^N := \mu^N(\mathbb{X}_t^N) = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^i}, \quad t \in [0, T],$$

où $\mu^N(\mathbb{Y}^N) := \frac{1}{N} \sum_{i=1}^N \delta_{Y^i}$ pour tout vecteur $\mathbb{Y}^N = (Y^1, \dots, Y^N) \in (\mathbb{R}^d)^N$: autrement dit, nous avons, pour tout $\omega \in \mathcal{C}^N$, $\mu_t^N(\omega) = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^i(\omega)} = \frac{1}{N} \sum_{i=1}^N \delta_{\omega_t^i}$.

Nous noterons $\mathbb{P}_\vartheta^N$ à la place de $\mathbb{P}_b^N$ dans le cadre paramétrique, pour mettre en évidence la dépendance en le paramètre $\vartheta \in \Theta$. Nous admettons pour le moment que la famille $(\mathbb{P}_b^N)_b$ est bien définie sous nos hypothèses (voir le Théorème 31), et nous le prouverons rigoureusement dans l'Annexe 1. En particulier, le processus canonique $\mathbb{X}^N$ est solution, sous $\mathbb{P}_b^N$, du *système d'équations de diffusions en interaction*

$$\begin{cases} dX_t^i = b(t, X_t^i, \mu_t^N) dt + \sigma(t, X_t^i) dB_t^{i,b}, & t \in [0, T], 1 \leq i \leq N, \\ \mathcal{L}(\mathbb{X}_0^N) = \mu_0^{\otimes N}. \end{cases} \quad (10)$$

Nous avons donc, par définition, que $\mathcal{L}_{\mathbb{P}_b^N}(\mathbb{X}^N) = \mathbb{P}_b^N$.

Notre objectif statistique général sera d'estimer les fonctions μ^b et $b(\cdot, \cdot, \mu^b)$ à l'aide des observations $\mathbb{X}^N$. Par exemple, pour notre modèle de référence (4), le problème (10) se réécrit de la manière suivante : sur un espace de probabilité filtré quelconque $(\tilde{\Omega}^N, \tilde{\mathcal{F}}^N, (\tilde{\mathcal{F}}_t^N)_{t \in [0, T]}, \tilde{\mathbb{P}}^N)$ donné, muni de N mouvements browniens d -dimensionnels $\tilde{B}^1, \dots, \tilde{B}^N$ indépendants, et d'un vecteur aléatoire $(\tilde{X}_0^1, \dots, \tilde{X}_0^N)$ de loi $\mu_0^{\otimes N}$, $\tilde{\mathcal{F}}_0^N$ -mesurable, et indépendant de $\tilde{B}^1, \dots, \tilde{B}^N$, nous avons

$$\begin{cases} d\tilde{X}_t^i = (\vartheta_1 \tilde{X}_t^i + \vartheta_2) dt + \vartheta_3 \left(\frac{1}{N} \sum_{j=1}^N \tilde{X}_t^j - \tilde{X}_t^i \right) dt + \sigma d\tilde{B}_t^i, & t \in [0, T], 1 \leq i \leq N, \\ (\tilde{X}_t^1, \dots, \tilde{X}_t^N)_{|t=0} = (\tilde{X}_0^1, \dots, \tilde{X}_0^N), \end{cases} \quad (11)$$

et nous chercherons alors, à l'aide des observations $(\tilde{X}^1, \dots, \tilde{X}^N)$, à estimer $\vartheta = (\vartheta_1, \vartheta_2, \vartheta_3)$, $\mu^b = \mu^\vartheta$, et $b(\cdot, \cdot, \mu^\vartheta) = b_0(\vartheta; \cdot, \cdot, \mu^\vartheta)$.

Revenons au cas général. La présentation qui suit est fortement inspirée de [115] et [184]. Le système d'équation (10) peut se réécrire

$$\begin{cases} d\mathbb{X}_t^N = b^{(N)}(t, \mathbb{X}_t^N) dt + \sigma^{(N)}(t, \mathbb{X}_t^N) d\mathbb{B}_t^{N,b}, & t \in [0, T], \\ \mathcal{L}(\mathbb{X}_0^N) = \mu_0^{\otimes N}, \end{cases} \quad (12)$$

où nous avons posé

$$b^{(N)} : (t, (x^1, \dots, x^N)) \in [0, T] \times (\mathbb{R}^d)^N \mapsto \left(b(t, x^1, \mu^N(x^1, \dots, x^N)), \dots, b(t, x^N, \mu^N(x^1, \dots, x^N)) \right)^\top,$$

$$\sigma^{(N)} : (t, (x^1, \dots, x^N)) \in [0, T] \times (\mathbb{R}^d)^N \mapsto \begin{pmatrix} \sigma(t, x^1) & & (0) \\ & \ddots & \\ (0) & & \sigma(t, x^N) \end{pmatrix}.$$

Notons, pour tout $t \in [0, T]$, $\mu_t^{N,N}$ la loi de $\mathbb{X}_t^N$ sous $\mathbb{P}_b^N$: $\mu_t^{N,N}$ est donc une loi de probabilité sur $(\mathbb{R}^d)^N$ que nous voudrions essayer de décrire plus précisément. Pour cela, nous pouvons tout d'abord utiliser la méthode usuelle consistant à déterminer quelle équation elle vérifie. Pour toute fonction test $\Phi \in \mathcal{C}_c^\infty((\mathbb{R}^d)^N; \mathbb{R})$, nous avons d'après la formule d'Itô (voir *e.g.* [171]), pour tout $t \in [0, T]$,

$$\begin{aligned} \Phi(\mathbb{X}_t^N) &= \Phi(\mathbb{X}_0^N) + \int_0^t \left(\frac{1}{2} \text{Tr}(c^{(N)}(s, \mathbb{X}_s^N) D^2 \Phi(\mathbb{X}_s^N)) + \nabla \Phi(\mathbb{X}_s^N)^\top b^{(N)}(s, \mathbb{X}_s^N) \right) ds \\ &\quad + \int_0^t \nabla \Phi(\mathbb{X}_s^N)^\top \sigma^{(N)}(s, \mathbb{X}_s^N) d\mathbb{B}_s^{N,b}. \end{aligned}$$

En prenant l'espérance $\mathbb{E}_{\mathbb{P}_b^N}$, nous trouvons donc, pour tout $t \in [0, T]$,

$$\int_{(\mathbb{R}^d)^N} \Phi d\mu_t^{N,N} = \int_{(\mathbb{R}^d)^N} \Phi d\mu_0^{\otimes N} + \int_0^t \int_{(\mathbb{R}^d)^N} \left(\frac{1}{2} \text{Tr}(c^{(N)}(s, \cdot) D^2 \Phi) + \nabla \Phi^\top b^{(N)}(s, \cdot) \right) d\mu_s^{N,N} ds,$$

ce qui signifie que $\mu^{N,N}$ est solution de probabilité du problème de Cauchy

$$\begin{cases} \partial_t \mu_t^{N,N} = \frac{1}{2} \sum_{i,j=1}^{dN} \partial_{ij}^2 (c_{ij}^{(N)}(t, \cdot) \mu_t^{N,N}) - \text{div}(b^{(N)}(t, \cdot) \mu_t^{N,N}), \\ \mu_{|t=0}^{N,N} = \mu_0^{\otimes N}. \end{cases} \quad \text{pour } (t, (x^1, \dots, x^N)) \in (0, T) \times (\mathbb{R}^d)^N, \quad (13)$$

$\mu^{N,N}$ est donc solution d'une équation *linéaire*, mais en *grande dimension*. Une telle équation (dite *équation maîtresse* ou *équation de Liouville*) est d'autant plus difficile à traiter que le nombre N devient grand. Or nous avons vu que ce dernier est amené à être gigantesque! Aussi, il convient d'envisager une autre manière d'étudier la loi de $\mathbb{X}^N$.

Heureusement, notre système de particules $\mathbb{X}^N$ est ici *symétrique* sous $\mathbb{P}_b^N$, ce qui signifie que la loi de $(X^1, \dots, X^N)$ sous $\mathbb{P}_b^N$ (c'est-à-dire $\mathbb{P}_b^N$ elle-même) n'est pas modifiée par une permutation des particules. En effet, étant donné que $\mathcal{L}_{\mathbb{P}_b^N}(X_0^1, \dots, X_0^N) = \mu_0^{\otimes N}$, que cette loi est symétrique, et que la dynamique gouvernant l'évolution du système est elle-même symétrique, nous aurons également que $\mathcal{L}_{\mathbb{P}_b^N}(X_t^1, \dots, X_t^N)$ est une loi symétrique pour tout $t \in [0, T]$. En particulier toutes les particules sont de même loi sous $\mathbb{P}_b^N$, loi que nous noterons $\mathbb{P}_b^{1,N}$. De manière générale, pour $1 \leq K \leq N$, nous noterons $\mathbb{P}_b^{K,N}$ la loi sous $\mathbb{P}_b^N$ d'une sous-famille $\mathbb{X}^{K,N}$ de K particules de $\mathbb{X}^N$ (quelconques parmi les N par symétrie), et pour tout $t \in [0, T]$, nous noterons $\mu_t^{K,N}$ la loi de $\mathbb{X}_t^{K,N}$.

Nous pouvons alors nous poser la question suivante : connaissant la loi des N particules au temps initial, pouvons-nous obtenir des informations sur la loi des positions de ces particules à un temps t ultérieur, lorsque $N \rightarrow \infty$? En particulier, quel est le comportement de la densité limite d'une particule du système, *i.e.* de $\mu_t^{1,N}$ lorsque $N \rightarrow \infty$? Etablir une *limite de champ moyen* dans notre cadre de travail consiste précisément à obtenir un modèle sur la densité statistique limite des particules, en partant de la famille $\mathbb{X}^N$, et en passant à la limite lorsque N tend vers l'infini.

De la même manière que ci-dessus pour $\mu^{N,N}$, pour tous $\Phi \in \mathcal{C}_c^\infty(\mathbb{R}^d; \mathbb{R})$ et $t \in [0, T]$,

$$\begin{aligned} \int_{\mathbb{R}^d} \Phi(x^1) d\mu_t^{1,N}(x^1) &= \int_{\mathbb{R}^d} \Phi(x^1) d\mu_0(x^1) + \int_0^t \int_{\mathbb{R}^d} \frac{1}{2} \text{Tr}(c(s, x^1) D^2 \Phi(x^1)) d\mu_s^{1,N}(x^1) ds \\ &\quad + \int_0^t \int_{(\mathbb{R}^d)^N} \langle \nabla \Phi(x^1), b(s, x^1, \mu^N(x^1, \dots, x^N)) \rangle d\mu_s^{N,N}(x^1, \dots, x^N) ds, \end{aligned}$$

et il est facile de voir qu'hormis dans le cas trivial où les trajectoires des particules sont indépendantes, ce qui ne peut arriver que lorsqu'il n'y a pas d'interaction, l'équation vérifiée par $\mu^{1,N}$ n'est pas fermée, puisqu'elle dépend elle-même de $\mu^{N,N}$.

Pour contourner ce fléau de la dimension, la technique usuelle est de travailler avec la *mesure empirique* μ^N du système de particules plutôt qu'avec sa configuration microscopique. En effet, d'une part, nous pouvons considérer que, tous les μ_t^N , $N \geq 1$, $t \in [0, T]$, vivent dans le même espace, l'espace des mesures aléatoires définies sur $(\prod_{i=1}^\infty \mathcal{C}, \otimes_{i=1}^\infty \mathcal{F}_T)$ et à valeurs dans $\mathcal{P}_1(\mathbb{R}^d)$:

$$\mu_t^N : \omega = (\omega^i)_{i \in \mathbb{N}^*} \mapsto \frac{1}{N} \sum_{i=1}^N \delta_{\omega_t^i} \in \mathcal{P}_1(\mathbb{R}^d),$$

et il devient donc possible de se demander si μ^N converge vers quelque chose lorsque $N \rightarrow \infty$. D'autre part, grâce au caractère symétrique de notre système de particules, μ^N contient toutes les informations nécessaires (cf. [173]) pour établir le comportement asymptotique du système quand le nombre de particules N tend vers l'infini.

La limite de champ moyen de notre système de particules solution de (10) est en fait la solution μ^b de l'équation de McKean-Vlasov (5) : pour tous $t \in [0, T]$, $\Phi \in \mathcal{C}_c(\mathbb{R}^d; \mathbb{R})$ et $\epsilon > 0$,

$$\mathbb{P}_b^N \left(\left| \int_{\mathbb{R}^d} \Phi d(\mu_t^N - \mu_t^b) \right| > \epsilon \right) \xrightarrow{N \rightarrow \infty} 0.$$

Nous avons établi dans cette thèse un résultat de *concentration* qui quantifie précisément cette limite de champ moyen (voir le Théorème 109).

L'autre question usuelle lors de l'étude d'un tel système de particules est la question de la *propagation du chaos*. Nous avons supposé que les états des particules au temps initial $t = 0$ sont *i.i.d.* sous $\mathbb{P}_b^N : \mu_0^{N,N} = \mu_0^{\otimes N}$. Est-ce que la propriété d'indépendance reste vraie aux temps ultérieurs, ou au moins approximativement vraie, lorsque $N \rightarrow \infty$? Autrement dit, le chaos est-il préservé par la dynamique? Il se trouve que dans notre cadre la propagation du chaos est bien vérifiée : pour tout $t \in [0, T]$, $(\mu_t^{N,N})_N$ est μ_t^b -chaotique, *i.e.* $\mu_t^{N,N}$ est proche de $(\mu_t^b)^{\otimes N}$ quand $N \rightarrow \infty$, au sens où, pour tout entier $K \geq 1$, pour toute fonction test $\Phi(x_1, \dots, x_N) = \Phi_1(x_1) \times \dots \times \Phi_K(x_K)$, avec $\Phi_i \in \mathcal{C}_b(\mathbb{R}^d; \mathbb{R})$, $1 \leq i \leq K$,

$$\int_{(\mathbb{R}^d)^N} \Phi d\mu_t^{N,N} \xrightarrow{N \rightarrow \infty} \prod_{i=1}^K \int_{\mathbb{R}^d} \Phi_i(x_i) d\mu_t^b(x_i).$$

Les particules sont donc proches d'être indépendantes et identiquement distribuées, en tout temps $t \in [0, T]$, dans la limite $N \rightarrow \infty$.

En réalité, la propriété du *chaos* est même vraie à l'échelle du processus entier : la loi $\mathbb{P}_b^N$ sur $\mathcal{C}([0, T]; (\mathbb{R}^d)^N)$, assimilé à $\mathcal{C}([0, T]; \mathbb{R}^d)^N$, est $\bar{\mathbb{P}}_b$ -chaotique au sens où, pour tout entier $K \geq 1$, pour toute fonction test $\Phi(x^1, \dots, x^K) = \Phi_1(x^1) \times \dots \times \Phi_K(x^K)$, avec $\Phi_i \in \mathcal{C}_b(\mathcal{C}([0, T]; \mathbb{R}^d); \mathbb{R})$ pour tout $1 \leq i \leq K$, nous avons

$$\int_{\mathcal{C}([0, T]; \mathbb{R}^d)^N} \Phi d\mathbb{P}_b^N \xrightarrow{N \rightarrow \infty} \prod_{i=1}^K \int_{\mathcal{C}([0, T]; \mathbb{R}^d)} \Phi_i(x^i) d\bar{\mathbb{P}}_b(x^i) .$$

Et cette propriété implique la propagation du chaos, voir en particulier à ce sujet notre Proposition 32.

Nous allons à présent donner un aperçu (non-exhaustif) des articles de ces dernières décennies consacrés à l'étude des propriétés de l'équation de McKean-Vlasov, et présenter quelques articles traitant des applications possibles.

1.4 Historique & Applications

Chronologie des résultats probabilistes & analytiques sur les modèles (5) et (10)

Ces travaux pionniers de Kac et McKean ont ouvert la voie à un domaine des Probabilités très prolifique. Citons tout d'abord les travaux fondateurs des années 1980-1990 qui apportèrent sur le modèle de McKean-Vlasov des résultats de propagation du chaos, d'étude des fluctuations dans différents espaces fonctionnels, et de grandes déviations. Citons les travaux de l'école japonaise à la suite de Tanaka (Hitsuda, Funaki, Mitoma, Nagasawa, Shiga, Tamura, voir [179, 178, 76, 114, 170, 93, 158, 177, 153]) ; de Dawson et Gärtner [55, 56, 57, 78], Sznitman [172], Oelschläger [161], Léonard [124, 125, 126], Scheutzw [168], Finnoff [72], Méléard, Jourdain, Fernandez et Roelly-Coppoletta [149, 150, 71, 107], Ben Arous, Brunaud et Zeitouni [10, 11].

Puis, les années 2000-2010 virent un raffinement de ces résultats de propagation du chaos et de limite de champ moyen avec l'établissement d'inégalités géométriques et de concentration, en particulier par Bolley, Cattiaux, Gentil, Guillin, Malrieu et Villani (voir *e.g.* [142, 20, 38, 21]), ou plus récemment des résultats de Gobet et Pagliarani [86], Jourdain et Tse [108], Chassagneux, Szpruch et Tse [41], Jabir [100].

En parallèle, ces résultats de type champ moyen ont justifié l'élaboration de méthodes particulières pour l'approximation numérique des solutions des EDP et EDS non-linéaires de type McKean-Vlasov, en particulier par Talay, Bossy et leurs co-auteurs [25, 26, 24, 176], Kohatsu-Higa et Ogawa [112] (voir également le livre de Gobet [85]). Et plus récemment dans les travaux de Jourdain, Bencheikh, Delarue, Chassagneux, Szpruch, Tse [13, 174, 175], Belomestny, Schoenmakers, Pilipauskaitė et Podolskij [9, 8], Liu et Pagès [133] (avec également un papier en préparation de Hoffmann et Liu), des méthodes de type Multi-level Monte-Carlo ou de Cubature [52] Souvent motivées par les travaux précédents, des propriétés de régularité de l'équation de McKean-Vlasov ont également été mises en évidence : voir les travaux de Jourdain [107], Morien [156], Ganz-Bustos [77], Crisan et McMurray [51], Chaudru de Raynal et Frikha [62, 59, 61, 58], Tse [180] et Han [91].

Applications des modèles de champ moyen

Cependant, jusqu'au début des années 2000, la formulation d'un programme d'estimation statistique dans ce contexte était globalement hors de portée (avec cependant de notables exceptions comme les travaux de Löcherbach [134] en temps grand, et Kasonga [111]), au moins pour deux raisons : d'abord, les outils probabilistes requis pour l'estimation adaptative non-paramétrique étaient en cours de développement ; deuxièmement, les systèmes de particules microscopiques issus de la Physique Statistique ne sont pas naturellement observables et la motivation pour une étude statistique de ces systèmes n'était donc pas évidente dans ce contexte.

Mais la situation évolua progressivement au cours des deux décennies suivantes, qui virent une sorte de renouveau des modèles de type dynamiques collectives en interaction de champ moyen et McKean-Vlasov dans différents domaines d'applications pour modéliser des mouvements collectifs cette fois-ci observables : allant des biomathématiques (neurosciences, Baladron *et al.* [6], des modèles structurés en dynamique des populations, Mogilner *et al.* [154], Burger *et al.* [32], Brillinger *et al.* [28, 29, 30]) aux Sciences Sociales (dynamiques d'opinions, Chazelle *et al.* [43], comportements coopératifs, Canuto *et al.* [33]) et à la Finance (risque systémique, Fouque et Sun [73]). En particulier, les *jeux à champ moyen*, développés aux alentours de l'année 2006 par Lasry et Lions (voir [121, 122], Cardaliaguet *et al.* [35], les notes de cours de Cardaliaguet [34] et le livre de Carmona et Delarue [37]), et indépendamment par Caines, Huang et Malhamé [96], apparaissent ainsi comme un nouvel objectif pour les développements statistiques (la récente contribution de Giesecke *et al.* [84] ouvre d'ailleurs la voie dans cette direction).

Ce domaine probabiliste a donc atteint ces dernières années la maturité suffisante pour faire l'objet d'études statistiques, et celles-ci se sont en effet développées à partir de 2015 environ (voir les références ci-dessous). Cette thèse s'inscrit dans ce mouvement de fond ainsi que, de manière plus large, dans plusieurs axes de recherche des Statistiques actuelles que nous développons ci-dessous.

Cadre statistique de cette thèse

Cette thèse s'inscrit tout d'abord dans le cadre des études statistiques de processus aléatoires, et plus particulièrement de processus de diffusions. Nous pouvons faire remonter ce domaine des Statistiques aux années 1960-1970 avec les travaux de *l'école russe* (Kolmogorov, Sinai, Arató, Novikov, Shiryaev, Lipster, Ibragimov, Hasminskii, Kutoyants, Spokoiny ...) et de l'école indo-américaine (Prakasa Rao, Rubin, Basawa, Striebel ...), complétés à partir des années 1980-1990, par les travaux de *l'école française* (Jacod, Bosq, Dacunha-Castelle, Florens-Zmirou, Ait-Sahalia, Banon, Genon-Catalot, Picard, Laredo, Comte, Gassiat, Hoffmann ...), de Höpfner, Löcherbach, Reiss, Belomestny ..., puis depuis les années 2000 par beaucoup d'autres (souvent élèves des précédents!).

Mais cette thèse s'inscrit également dans le cadre des modèles statistiques relatifs à des EDP, dont l'intérêt a progressivement émergé au sein de la communauté statisticienne au cours de la dernière décennie : nous pouvons citer les travaux Doumic *et al.* [67, 68], Hoffmann et Olivier [95], Boumezoued *et al.* [27], Ngoc *et al.* [94], Maïda *et al.* [141]. L'analyse d'EDP elliptiques, paraboliques, ou de transport-fragmentation apporte une lumière nouvelle sur la structure des expériences statistiques associées, et enrichit la théorie classique de Statistique des Processus.

Nous faisons donc dans cette thèse un pas en direction d'une certaine classe d'EDP non-linéaires paraboliques, les EDP de McKean-Vlasov, en utilisant des méthodes relativement récentes dans le cadre non-paramétrique (méthode de Goldenshluger-Lepski [87, 88, 89] dans sa version développée dans [27, 105]), et relativement classiques dans le cadre paramétrique (théorie de Ibragimov et

Hasminski [98], et théorie LAN de Le Cam [123]).

Etat de l'art des méthodes d'estimation statistique du drift dans le cadre des systèmes de diffusion en interaction de champ moyen & et de l'équation de McKean-Vlasov

Nous allons enfin passer en revue, en essayant cette fois-ci d'être le plus exhaustif possible, les articles traitant de l'estimation statistique du drift dans le cadre de systèmes de diffusions indépendantes, de systèmes de diffusions en interaction, et enfin dans le cadre de l'équation de McKean-Vlasov elle-même.

Tout d'abord, concernant l'étude statistique de N diffusions *i.i.d.* sur $[0, T]$, à $T > 0$ fixé, et dans l'asymptotique $N \rightarrow \infty$, nous pouvons citer les articles d'estimation non-paramétrique suivants : pour des observations continues, l'article de Comte et Genon-Catalot [48] (par une méthode de projection, le coefficient de diffusion pouvant être non borné) et l'article de Marie et Rosier [144] (par un estimateur de type Nadaraja-Watson, voir [157]) ; pour des observations discrètes, l'article de Denis, Dion-Blanc et Martinez [66] (par un estimateur *ridge* obtenu par minimisation d'un contraste des moindres carrés contraint).

Concernant ensuite les systèmes de diffusions en interaction, nous pouvons faire remonter leur programme d'estimation statistique à Kasonga [111], dans lequel l'auteur estime paramétriquement le drift pour des modèles *linéaires* en la mesure et en le paramètre. Un des objectifs de cette thèse (voir [64] et le Chapitre 1) a été d'étendre les résultats de Kasonga à des modèles plus généraux.

En parallèle de notre travail, dans des articles récents, Giesecke, Schwenkler et Sirignano [84], Sharrock, Kantas, Parpas et Grigorios [169], et Amorino, Heidari, Pilipauskaitė et Podolskij [3], ont également poursuivi dans la voie ouverte par Kasonga, et considéré un cadre paramétrique très général pour lequel des estimateurs de minimum de contraste, convergents et asymptotiquement normaux, sont construits. Chen a quant à lui utilisé les propriétés des processus d'Ornstein-Uhlenbeck pour en déduire des inégalités fines de grandes déviation pour le maximum de vraisemblance [44] tandis que, dans un cadre non-paramétrique, Comte et Marie [49] ont étudié statistiquement des diffusions avec des mouvements browniens corrélés.

Récemment, une série de travaux se sont intéressés à l'estimation non-paramétrique de l'interaction dans un système de particules : tout d'abord dans un cadre déterministe (système du premier ordre homogène [23], [138], système du premier ordre hétérogène [136], système sur une variété riemannienne [140], système hétérogène du second ordre [151], [190]), puis dans un cadre stochastique (à partir de multiples trajectoires dans [137], puis [130] et [129] se sont concentrés sur la condition de coercivité qui permet l'identifiabilité de la fonction d'interaction).

Enfin, concernant de manière spécifique l'estimation du drift dans l'équation de McKean-Vlasov, nous pouvons citer les articles suivants. Tout d'abord dans le cadre paramétrique : [189], [132], [169] (pour des observations continues en temps long), [82] (pour des observations continues en variance petite), [164] (pour des observations discrètes par minimisation d'un contraste basé sur le schéma d'Euler-Maruyama *tamed*, voir [97]) et enfin [81]. Ensuite, dans le cadre non-paramétrique, [119] (pour des observations discrètes), [120] (pour des observations continues), [65] (c.f. Chapitre 2 de cette thèse), et [119, 120] qui se concentrent quant à eux sur la condition de coercivité pour obtenir l'identifiabilité de la fonction d'interaction, et sont la contrepartie stochastique de [137, 130, 129]. Enfin, récemment, Belomestny *et al.* [8] ont travaillé sur l'estimation semi-paramétrique

de l'équation de McKean-Vlasov.

Bibliographie généraliste en lien avec notre modèle & Prolongements possibles

Plusieurs articles, livres ou notes de cours donnent une vue d'ensemble des équations de McKean-Vlasov, des systèmes de particules en interaction, des limites de champ moyen et de la propagation du chaos. Citons, sans objectif d'exhaustivité, Sznitman [173], Méléard [148], Villani [184], Mischler [152], Jabin et Wang [99], Chevallier [45], Carmona et Delarue [37], Lackner [115], Chaintron et Diez [39, 40]. Ils traitent également pour certains de cas que nous n'étudierons pas dans cette thèse : diffusion dépendant de la mesure, drift irrégulier, drift polynomiaux dérivant de potentiels avec des contraintes de convexité pour des questions de temps grand et d'ergodicité, évolution sur un ensemble borné avec des conditions de réflexion aux bords, modèles en interactions modérées, modèles cinétiques, mouvement brownien remplacé par un processus de type Lévy, ajout d'un bruit commun, etc. Certains de ces modèles ont déjà fait l'objet d'études statistiques approfondies (voir les auteurs et références donnés ci-dessus), d'autres pourraient constituer un prolongement naturel des travaux entrepris au cours de cette thèse.

2 Hypothèses & Premières propriétés

2.1 Hypothèses générales sur (b, σ, μ_0)

Hypothèse sur la condition initiale μ_0

Nous ferons une hypothèse d'*intégrabilité forte* sur μ_0 . Plus précisément, dans le cadre paramétrique, nous supposerons que

Hypothèse 1. μ_0 admet des moments finis de tout ordre, i.e. $\mu_0 \in (\bigcap_{r \geq 1} \mathcal{P}_r(\mathbb{R}^d))$.

Et, dans le cadre non-paramétrique, nous ferons l'hypothèse suivante.

Hypothèse 2. Il existe $\gamma_0 > 0, \gamma_1 \geq 1$, tel que $\int_{\mathbb{R}^d} \exp(\gamma_0 |x|^2) \mu_0(dx) \leq \gamma_1$.

Nous noterons $\gamma_0(\mu_0), \gamma_1(\mu_0)$ des réels vérifiant cette hypothèse.

Remarque 3.

- (i) Des exemples de lois de probabilité vérifiant l'Hypothèse 2 sont donnés par les lois à support borné, les lois gaussiennes, ou encore les lois sous-gaussiennes au sens de [31].
- (ii) L'Hypothèse 2 implique l'Hypothèse 1, avec, pour tout $r \geq 1$, $\mathbf{m}_r(\mu_0) \leq \left(\frac{r! \gamma_1(\mu_0)}{\gamma_0(\mu_0)^r} \right)^{1/2}$.
- (iii) Notons, pour tout entier $r \geq 1$, $\mathbf{q}_r(\mu_0) := \left(\frac{\mathbf{m}_{2r}(\mu_0)}{r!} \right)^{1/r}$. L'Hypothèse 1 implique l'Hypothèse 2 si l'on suppose de plus que la suite $(\mathbf{q}_r(\mu_0))_{r \geq 1}$ est bornée. Nous prendrons alors

$$0 < \gamma_0(\mu_0) < \left(\limsup_{r \rightarrow \infty} \mathbf{q}_r(\mu_0) \right)^{-1}, \quad \gamma_1(\mu_0) \geq \sum_{r=0}^{\infty} \frac{\gamma_0(\mu_0)^r \mathbf{m}_{2r}(\mu_0)}{r!}.$$

- (iv) Des exemples de lois de probabilité vérifiant l'Hypothèse 1 mais pas l'Hypothèse 2 sont donnés par les lois μ_0 pour lesquelles il existe une fonction $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ tendant vers $+\infty$ en $+\infty$, tel qu'à une constante près, pour tout entier $r \geq 1$, $\mathbf{m}_{2r}(\mu_0) \geq (rf(r))^r$; ceci est par exemple vérifié pour les lois exponentielles.

Hypothèses sur la fonction de diffusion σ

Dans le cadre paramétrique, comme dans le cadre non-paramétrique, nous supposons tout d'abord que c est *strictement uniformément elliptique*, autrement dit :

Hypothèse 4. *Il existe $\sigma_{\pm} > 0$ tel que, pour tout $(t, x) \in [0, T] \times \mathbb{R}^d$,*

$$\sigma_-^2 |y|^2 \leq (c(t, x)y)^\top y \leq \sigma_+^2 |y|^2 .$$

Remarque 5.

- (i) Des exemples de telles fonctions σ sont donnés par $\sigma(t, x) := f(x)I_d$, où $f : \mathbb{R}^d \rightarrow \mathbb{R}$ bornée et telle que $\inf_{\mathbb{R}^d} |f| > 0$: nous avons alors l'Hypothèse 4 avec $\sigma_+ = |f|_\infty$, $\sigma_- = \inf_{\mathbb{R}^d} |f|$;
- (ii) Si c vérifie l'Hypothèse 4 alors σ et c sont bornées : en effet en notant $|\cdot|$ la norme euclidienne usuelle sur $\mathcal{M}_d(\mathbb{R})$, qui est donc la norme de Frobenius-Hilbert-Schmidt sur cet espace de matrices, il existe une constante $C_d > 0$ tel que pour tout $(t, x) \in [0, T] \times \mathbb{R}^d$

$$|c(t, x)| \leq C_d |\sigma(t, x)|^2 = C_d \operatorname{Tr}(c(t, x)) \leq C_d d \sigma_+^2 .$$

σ sera également supposée *lipschitzienne en $x \in \mathbb{R}^d$ uniformément par rapport à $t \in [0, T]$* :

Hypothèse 6. *Il existe $C > 0$ tel que, pour tout $(t, x, x') \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d$,*

$$|\sigma(t, x) - \sigma(t, x')| \leq C |x - x'| ,$$

et nous noterons $|\sigma|_{Lip}$ la plus petite constante $C \geq 0$ pour laquelle ceci est vrai.

Remarque 7.

- (i) Des exemples de telles fonctions σ sont donnés par $\sigma(t, x) := f(x)I_d$, où $f : \mathbb{R}^d \rightarrow \mathbb{R}$ est lipschitzienne ;
- (ii) Un exemple de fonction σ vérifiant l'Hypothèse 6 mais pas l'Hypothèse 4 est donné par $\sigma(t, x) = f(x)I_d$ où f est lipschitzienne mais non bornée, par exemple $f(x) = |x|$;
- (iii) Un exemple de fonction σ vérifiant l'Hypothèse 4 mais pas l'Hypothèse 6 est donné par $\sigma(t, x) = f(x)I_d$ avec f bornée et $\inf_{\mathbb{R}^d} |f| > 0$, mais f non Lipschitzienne, par exemple $f(x) = 1 + |x|^{1/2} \mathbf{1}_{|x| \leq 1}$.

Hypothèses sur la fonction de drift b

Nous ferons tout d'abord l'hypothèse suivante de *bornitude ponctuelle* de b : dans le cadre paramétrique, nous supposons

Hypothèse 8. $\sup_{(\vartheta, t) \in \Theta \times [0, T]} |b_0(\vartheta; t, 0, \delta_0)| < \infty ,$

et dans le cadre non-paramétrique, nous supposons

Hypothèse 9. $\sup_{t \in [0, T]} |b(t, 0, \delta_0)| < \infty .$

Remarque 10. b vérifie de manière évidente l'Hypothèse 8 (resp. l'Hypothèse 9) lorsqu'il est continu en ϑ et ne dépend pas de t , ou bien lorsqu'il est continu en (ϑ, t) (resp. lorsqu'il ne dépend pas de t ou bien est continu en t).

L'hypothèse suivante traite de la régularité de b en (x, ν) . Dans le cadre paramétrique, nous supposons b_0 *lipschitzienne* en $(x, \nu) \in \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)$ *uniformément par rapport* à $(\vartheta, t) \in \Theta \times [0, T]$:

Hypothèse 11. *Il existe $C > 0$ tel que pour tout $(\vartheta, t) \in \Theta \times [0, T]$*

$$|b_0(\vartheta; t, x', \nu') - b_0(\vartheta; t, x, \nu)| \leq C(|x' - x| + W_1(\nu', \nu)) ,$$

et nous noterons $|b_0|_{\text{Lip}}$ la plus petite constante $C \geq 0$ pour laquelle ceci est vrai.

Exemples 12. (i) Un exemple de telle fonction b_0 est donné par notre modèle de référence (4) :

$$b_0(\vartheta; t, x, \nu) = (\vartheta_1 x + \vartheta_2) + \vartheta_3 \int_{\mathbb{R}^d} (y - x) \nu(dy) , \quad \vartheta = (\vartheta_1, \vartheta_2, \vartheta_3) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R} .$$

(ii) Un autre exemple de telle fonction b_0 est donné par la généralisation de l'exemple précédent :

$$b_0(\vartheta; t, x, \nu) = \vartheta_1 f(x) + \vartheta_2 + \vartheta_3 \int_{\mathbb{R}^d} g(y - x) \nu(dy) , \quad \vartheta = (\vartheta_1, \vartheta_2, \vartheta_3) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R} , \quad (14)$$

où $f, g : \mathbb{R}^d \rightarrow \mathbb{R}^d$ lipschitziennes.

(iii) Le modèle suivant, que nous appellerons *modèle d'attraction-répulsion* ou de *flocking* (voir des modèles d'essaims, de nuées *e.g.* [154], ou dans une version cinétique [19]) et qui permet de modéliser une attraction à longue portée et une répulsion à courte portée, en constitue un autre exemple :

$$b_0(\vartheta; t, x, \nu) = \int_{\mathbb{R}^d} \nabla U_\vartheta(y - x) \nu(dy) , \quad \vartheta = (\vartheta_1, \vartheta_2, \vartheta_3, \vartheta_4) \in (\mathbb{R}_+^*)^4 , \quad (15)$$

où $U_\vartheta : \mathbb{R}^d \rightarrow \mathbb{R}$ est une famille de potentiels de la forme

$$U_\vartheta(z) = -\vartheta_1 \exp(-\vartheta_2 |z|^2) + \vartheta_3 \exp(-\vartheta_4 |z|^2) , \quad z \in \mathbb{R}^d ,$$

ϑ_1, ϑ_3 et ϑ_2, ϑ_4 étant respectivement la force et la longueur de l'attraction et de la répulsion.

(iv) Enfin, un exemple de telle fonction b_0 , cette fois-ci non-linéaire, à la fois en le paramètre ϑ , et *au sens de McKean* en la mesure, est donné par

$$b_0(\vartheta; t, x, \nu) = F\left(\vartheta \int_{\mathbb{R}^d} g(y - x) \nu(dy)\right) , \quad \vartheta \in \mathbb{R} , \quad (16)$$

où, pour un certain entier $q \geq 1$, $g : \mathbb{R}^d \rightarrow \mathbb{R}^q$ et $F : \mathbb{R}^q \rightarrow \mathbb{R}^d$ sont lipschitziennes.

Pour affiner nos résultats (de régularité dans le cadre paramétrique, et de champ moyen de manière générale), nous aurons besoin d'une notion de régularité en la mesure provenant de la Théorie des Jeux à Champ Moyen : la *différentiabilité linéaire fonctionnelle en la mesure* (voir [37], [60], [108]).

Définition 13. Soit un entier $q \geq 1$. Une application $f : \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}^q$ a une différentielle linéaire fonctionnelle s'il existe $\delta_\nu f : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}^q$ tel que

$$f(\nu') - f(\nu) = \int_0^1 \int_{\mathbb{R}^d} \delta_\nu f(y, (1-\lambda)\nu + \lambda\nu') (\nu' - \nu)(dy) d\lambda.$$

vérifiant les propriétés de régularité suivantes

$$\begin{aligned} |\delta_\nu f(y', \nu') - \delta_\nu f(y, \nu)| &\leq C(|y' - y| + \mathcal{W}_1(\nu', \nu)), \\ |\partial_y(\delta_\nu f(y, \nu') - \delta_\nu f(y, \nu))| &\leq C\mathcal{W}_1(\nu', \nu), \end{aligned}$$

pour un certain $C > 0$. Nous noterons $|\delta_\nu f|_{\text{Lip}, \infty}$ la plus petite constante $C \geq 0$ pour laquelle ceci est vrai.

Nous pouvons itérer le processus décrit dans la Définition 13 : nous obtenons alors, pour tout entier $k \geq 1$, une notion de différentielle linéaire fonctionnelle k -ième via l'existence d'applications

$$\delta_\nu^\ell f : (\mathbb{R}^d)^\ell \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}^q \text{ pour } \ell = 1, \dots, k,$$

définies récursivement par $\delta_\nu^\ell f = \delta_\nu \circ \delta_\nu^{\ell-1} f$ et vérifiant les propriétés de régularité associées (et dont nous noterons les constantes minimales $|\delta_\nu^\ell f|_{\text{Lip}, \infty}$).

Remarque 14. Il convient de remarquer que les différentielles linéaires fonctionnelles définies de cette manière le sont en fait à une constante additive près étant donné que $\int_{\mathbb{R}^d} (\nu' - \nu)(dy) = 0$. Une normalisation possible est d'imposer itérativement, pour $1 \leq \ell \leq k$, la condition

$$\delta_\nu^\ell f(y^1, \dots, y^\ell, \nu) = 0 \text{ si un des } y^i, 1 \leq i \leq \ell \text{ est nul.}$$

Les différentielles linéaires fonctionnelles successives sont alors définies de manière unique. Cependant, par abus de notation, nous noterons $\delta_\nu^\ell f(y^1, \dots, y^\ell, \nu)$ pour une des différentielles linéaires fonctionnelles possibles.

Exemples 15. Un exemple d'une classe de fonctions $f : [0, T] \times \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}^q$ admettant une k -ième différentielle linéaire fonctionnelle, uniformément par rapport à $(t, x) \in [0, T] \times \mathbb{R}^d$, mais pas de $(k+1)$ -ième différentielle linéaire fonctionnelle est donné, pour $q' \geq 1$, $\tilde{f} : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{q'}$ régulière, et $F \in \mathcal{C}_b^{k+1}(\mathbb{R}^{q'}; \mathbb{R}^q) \setminus \mathcal{C}_b^{k+2}(\mathbb{R}^{q'}; \mathbb{R}^q)$, par :

$$f(t, x, \nu) = F\left(\int_{\mathbb{R}^d} \tilde{f}(t, x, y) \nu(dy)\right).$$

Nous ferons finalement l'hypothèse suivante sur b dans le cadre non-paramétrique.

Hypothèse 16. Une des trois conditions suivantes est satisfaite pour un certain $C > 0$:

(i) (Lipschitz) $d = 1$ et

$$|b(t, x', \nu') - b(t, x, \nu)| \leq C(|x' - x| + \mathcal{W}_1(\nu', \nu));$$

nous noterons $|b|_{\text{Lip}}$ la plus petite constante $C \geq 0$ pour laquelle ceci est vrai.

(ii) (Existence d'une k -ième différentielle linéaire fonctionnelle en la mesure.) Soit $k \geq 1$. Pour

($d = 1$ et $k \geq 1$) ou ($d = 2$ et $k \geq 2$) ou ($d \geq 3$ et $k \geq d/2$),

nous avons

$$|b(t, x', \nu') - b(t, x, \nu)| \leq C(|x' - x| + \mathcal{W}_1(\nu', \nu)).$$

et l'application

$$\nu \mapsto b(t, x, \nu)$$

admet une k -ième différentielle linéaire fonctionnelle au sens de la Définition 13, uniformément par rapport à (t, x) , avec de plus la représentation suivante :

$$\delta_\nu^k b(t, x, (y_1, \dots, y_k), \nu) = \sum_{\mathcal{J} \subset \{1, \dots, k\}, m \geq 1} \bigotimes_{j \in \mathcal{J}} (\delta_\nu^k b)_{\mathcal{J}, j, m}(t, x, y_j, \nu),$$

où la somme en m est finie avec au maximum m_b termes et les applications $(\delta_\nu^k b)_{\mathcal{J}, j, m} : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ sont telles que

$$|(\delta_\nu^k b)_{\mathcal{J}, j, m}(t, x', y', \nu) - (\delta_\nu^k b)_{\mathcal{J}, j, m}(t, x, y, \nu)| \leq C(|x' - x| + |y' - y|);$$

nous noterons $|\delta_\nu^{k_b} b|_{\text{Lip}}$ la plus petite constante $C \geq 0$ pour laquelle cette hypothèse est vraie, pour l'ordre de différentiabilité k_b^* le plus élevé de b (au sens où b n'est pas $k_b^* + 1$ linéairement différentiable en la mesure).

(iii) (Vlasov.) $d \geq 1$ et pour un certain $\ell \geq 1$

$$b(t, x, \nu) = \int_{(\mathbb{R}^d)^\ell} \tilde{b}(t, x, y^\ell) \nu^{\otimes \ell}(dy^\ell)$$

avec $y^\ell = (y_1, \dots, y_\ell)$, pour une certaine application mesurable

$$\tilde{b} : [0, T] \times \mathbb{R}^d \times (\mathbb{R}^d)^\ell \rightarrow \mathbb{R}^d$$

telle que :

$$|\tilde{b}(t, x', (y^\ell)') - \tilde{b}(t, x, y^\ell)| \leq C(|x' - x| + |(y^\ell)' - y^\ell|).$$

en supposant de plus que $\tilde{b}$ est $(\ell - 1)$ -fois dérivable en y^ℓ uniformément par rapport à (t, x) et de dérivées lipschitziennes en (x, y^ℓ) uniformément par rapport à t .

Remarque 17. La condition (ii) est vérifiée par des drifts de la forme

$$b(t, x, \nu) = \sum_{j=1}^J F_j \left(t, x, \int_{\mathbb{R}^{q_2}} G_j(t, x, \int_{\mathbb{R}^d} H_j(t, x, z) \nu(dz), z') \lambda_j(dz') \right),$$

pour des entiers $J, q_1, q_2, q_3 \geq 1$, des fonctions régulières $F_j(t, x, \cdot) : \mathbb{R}^{q_3} \rightarrow \mathbb{R}^d$, $G_j(t, x, \cdot) : \mathbb{R}^{q_1} \times \mathbb{R}^{q_2} \rightarrow \mathbb{R}^{q_3}$, $H_j(t, x, \cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^{q_1}$ et des mesures positives λ_j sur $\mathbb{R}^{q_2}$, sous certaines conditions techniques, et par des combinaisons de tels drifts. Des exemples explicites de modèles de champ moyen où le drift est de la forme (ii) plutôt que (i) ou (iii) sont donnés par exemple dans [46, 162, 148, ?]. La condition (ii) est une condition intermédiaire entre (i) et (iii).

Remarque 18. La condition (iii) est plus forte que la condition (ii) : en effet, sous la condition (iii), b a alors une k -ième différentielle linéaire fonctionnelle pour tout $k \geq 1$, et pour $k > \ell$, $\delta_\nu^k b(t, x, \cdot, \nu) = 0$: par convention, en reprenant les notations du cas précédent, nous dirons que $k_b^* = \ell + 1$. Pour $1 \leq k \leq \ell$ (cf. Exemple 2.7 de l'article de Jourdain et Tse [108]), nous avons

$$\delta_\nu^k b(t, x, y^k, \nu) = \sum_{s \in \mathfrak{S}_\ell} \int_{(\mathbb{R}^d)^{(\ell-k)}} \tilde{b}(t, x, y_{s(1)}^\ell, \dots, y_{s(\ell)}^\ell) \nu^{\otimes(\ell-k)}(dy_{k+1}, \dots, dy_\ell),$$

ce qui donne par exemple, pour $\ell = 1$ (resp. $\ell = 2$), $\delta_\nu^k b(t, x, \cdot, \nu) = \begin{cases} \tilde{b}(t, x, \cdot) & \text{si } k = 1 \\ 0 & \text{si } k \geq 2 \end{cases}$

$$\left(\text{resp. } \delta_\nu^k b(t, x, \cdot, \nu) = \begin{cases} y \mapsto \int_{\mathbb{R}^d} (\tilde{b}(t, x, y, z) + \tilde{b}(t, x, z, y)) \nu(dz) & \text{si } k = 1 \\ (y, y') \mapsto \tilde{b}(t, x, y, y') + \tilde{b}(t, x, y', y) & \text{si } k = 2 \\ 0 & \text{si } k \geq 3 \end{cases} \right).$$

Exemples 19. (i) Un exemple de noyau d'interaction vérifiant la condition (iii) pour $\ell = 1$ est donné par notre modèle de référence (4), ou par notre modèle (15).

(ii) Un exemple de noyau d'interaction vérifiant la condition (iii) pour $\ell \geq 1$ est donné par

$$\tilde{b}(t, x, y^\ell) = \exp \left(- \left| x - \frac{1}{\ell} \sum_{i=1}^\ell y_i \right|^2 \right) \quad (\text{ou encore par } \tilde{b}(t, x, y^\ell) = \begin{cases} \exp \left(- \frac{1}{t} \left| x - \frac{1}{\ell} \sum_{i=1}^\ell y_i \right|^2 \right) & \text{si } t \geq T_c \\ 0 & \text{sinon} \end{cases})$$

pour un certain temps critique $T_c \in (0, T)$.

Remarque 20. En combinant les hypothèses de bornitude ponctuelle et de régularité, nous obtenons que b est au plus linéaire : dans le cadre paramétrique,

$$|b_0(\vartheta; t, x, \nu)| \leq \sup_{(\vartheta', t') \in \Theta \times [0, T]} |b_0(\vartheta'; t', 0, \delta_0)| + |b_0|_{\text{Lip}}(|x| + \mathbf{m}_1(\nu)),$$

et dans le cadre non-paramétrique,

$$|b(t, x, \nu)| \leq \sup_{t' \in [0, T]} |b(t', 0, \delta_0)| + |b|_{\text{Lip}}(|x| + \mathbf{m}_1(\nu)).$$

Enfin, dans le cadre paramétrique, nous ferons une hypothèse usuelle de régularité de b_0 en ϑ :

Hypothèse 21. Il existe $r_1, r_2 \geq 1$ et $C > 0$ tel que la fonction $\vartheta \mapsto b_0(\vartheta; t, x, \nu)$ est deux fois différentiable dans l'intérieur de Θ uniformément par rapport à (t, x, ν) , et pour tous $1 \leq \ell, \ell' \leq p$, pour tout point ϑ dans l'intérieur de Θ ,

$$\sup_{t \in [0, T]} (|\partial_{\vartheta_\ell} b_0(\vartheta; t, x, \nu)| + |\partial_{\vartheta_\ell \vartheta_{\ell'}}^2 b_0(\vartheta; t, x, \nu)|) \leq C(1 + |x|^{r_1} + \mathbf{m}_{r_2}(\nu)),$$

$$\sup_{t \in [0, T]} |\partial_{\vartheta_\ell} b_0(\vartheta; t, x', \nu') - \partial_{\vartheta_\ell} b_0(\vartheta; t, x, \nu)| \leq C(|x' - x| + \mathcal{W}_1(\nu', \nu)),$$

et nous noterons respectivement $r_1(b_0), r_2(b_0)$ et $|D_\vartheta^{(2)} b_0|_{\text{Lip}, \infty}$ les plus petites valeurs de $r_1, r_2 \geq 1$ et $C \geq 0$ pour lesquelles ceci est vrai.

Remarque 22. Cela implique que, pour tout $1 \leq \ell \leq p$,

$$\begin{aligned} & \sup_{t \in [0, T]} \left(|b_0(\vartheta'; t, x', \nu') - b_0(\vartheta; t, x, \nu)| + |\partial_{\vartheta_\ell} b_0(\vartheta'; t, x', \nu') - \partial_{\vartheta_\ell} b_0(\vartheta; t, x, \nu)| \right) \\ & \lesssim |x' - x| + \mathcal{W}_1(\nu', \nu) + |\vartheta' - \vartheta| (1 + |x|^{r_1(b_0)} + \mathbf{m}_{r_2(b_0)}(\nu)) . \end{aligned}$$

Exemples 23. Reprenons notre Exemple 12,

- (i) Pour les cas (4) et (14), $D_\vartheta^2 b_0 = 0$, et l'Hypothèse 21 est bien vérifiée pour $r_1(b_0) = r_2(b_0) = 1$.
- (ii) Pour le cas (15), $b_0(\vartheta; t, x, \nu) = 2 \int_{\mathbb{R}^d} (\vartheta_1 \vartheta_2 (y - x) e^{-\vartheta_2 |y - x|^2} - \vartheta_3 \vartheta_4 (y - x) e^{-\vartheta_4 |y - x|^2}) \nu(dy)$, et donc l'Hypothèse 21 est bien vérifiée pour $r_1(b_0) = r_2(b_0) = 1$.
- (iii) Enfin, pour le cas (16), si F est de classe $\mathcal{C}^1$ (resp. $\mathcal{C}^2$), nous avons

$$\begin{aligned} \partial_\vartheta b_0^i(\vartheta; t, x, \nu) &= \left(\int_{\mathbb{R}^d} g(y - x) \nu(dy) \right) \cdot \nabla F^i \left(\vartheta \int_{\mathbb{R}^d} g(y - x) \nu(dy) \right) , \\ \left(\text{resp. } \partial_\vartheta^2 b_0^i(\vartheta; t, x, \nu) &= \left\langle D^2 F^i \left(\vartheta \int_{\mathbb{R}^d} g(y - x) \nu(dy) \right) \left(\int_{\mathbb{R}^d} g(y - x) \nu(dy) \right), \left(\int_{\mathbb{R}^d} g(y - x) \nu(dy) \right) \right\rangle \right) , \end{aligned}$$

d'où le fait que, si DF est lipschitzienne (resp. si $D^2 F$ est lipschitzienne), nous avons

$$\begin{aligned} |\partial_\vartheta b_0(\vartheta; t, x', \nu') - \partial_\vartheta b_0(\vartheta; t, x, \nu)| &\lesssim (|x' - x| + \mathcal{W}_1(\nu', \nu)) (1 + |x| + \mathbf{m}_1(\nu)) , \\ \left(\text{resp. } |\partial_\vartheta^2 b_0(\vartheta; t, x', \nu') - \partial_\vartheta^2 b_0(\vartheta; t, x, \nu)| &\lesssim (|x' - x| + \mathcal{W}_1(\nu', \nu)) (1 + |x|^2 + \mathbf{m}_2(\nu)) \right) , \end{aligned}$$

et, si par exemple g est de plus bornée, l'Hypothèse 21 est vérifiée pour $r_1(b_0) = r_2(b_0) = 1$.

Pour résumer, dans le cadre paramétrique, nous nous placerons sous les Hypothèses 1, 4, 6, 8, 11 et 21, tandis que dans le cadre non-paramétrique, nous nous placerons sous les Hypothèses 2, 4, 6, 9 et 16.

2.2 Modèle statistique considéré

Nous pouvons réinterpréter notre modèle de la manière suivante. Définissons tout d'abord les ensembles utilisés pour décrire les propriétés des *mesures initiales* :

$$\begin{aligned} \mathfrak{M}_d &:= \bigcap_{\ell \geq 1} \mathcal{P}_\ell(\mathbb{R}^d) , \quad \mathfrak{M}_{d, \mathbf{m}} := \left\{ \mu_0 \in \mathfrak{M}_d , \forall \ell \geq 0, \mathbf{m}_{2\ell}(\mu_0) \leq \mathbf{m}_{2\ell} \right\} , \\ \mathfrak{M}_{d, \mathbf{m}}^* &:= \left\{ \mu_0 \in \mathfrak{M}_{d, \mathbf{m}} , \left(\left(\frac{\mathbf{m}_{2\ell}(\mu_0)}{\ell!} \right)^{1/\ell} \right)_{\ell \geq 1} \text{ bornée} \right\} , \end{aligned}$$

pour $d \in \mathbb{N}^*$, $\mathbf{m} = (\mathbf{m}_\ell)_{\ell \geq 0} \in (\mathbb{R}_+)^{\mathbb{N}}$. Lorsque nous considérerons le cas $\mu_0 \in \mathfrak{M}_{d, \mathbf{m}}^*$, nous supposons implicitement que la suite $((\frac{\mathbf{m}_{2\ell}}{\ell!})^{1/\ell})_{\ell \geq 1}$ est elle-même bornée.

Définissons ensuite les ensembles auxquels appartiennent les *fonctions de diffusion* :

$$\mathfrak{C}_{T, d} := \left\{ c : [0, T] \times \mathbb{R}^d \rightarrow \mathcal{S}_d^{++}(\mathbb{R}) \text{ mesurable ; } \sup_{\substack{t \in [0, T], x, x' \in \mathbb{R}^d \\ x \neq x'}} \frac{|\sigma(t, x') - \sigma(t, x)|}{|x' - x|} < \infty , \right.$$

$$0 < \inf_{\substack{t \in [0, T], x, y \in \mathbb{R}^d \\ y \neq 0}} \frac{y^\top c(t, x) y}{|y|^2} \leq \sup_{\substack{t \in [0, T], x, y \in \mathbb{R}^d \\ y \neq 0}} \frac{y^\top c(t, x) y}{|y|^2} < \infty \Big\} ,$$

ainsi que son sous-ensemble

$$\begin{aligned} \mathfrak{C}_{T, d, \mathfrak{s}_+, \mathfrak{s}_-, \mathfrak{c}_1} &:= \left\{ c \in \mathfrak{C}_{T, d} ; \sup_{t \in [0, T], x, x' \in \mathbb{R}^d} \frac{|\sigma(t, x') - \sigma(t, x)|}{|x' - x|} \leq \mathfrak{c}_1 , \right. \\ &\quad \left. \mathfrak{s}_- \leq \inf_{t \in [0, T], x, y \in \mathbb{R}^d, y \neq 0} \frac{y^\top c(t, x) y}{|y|^2} \leq \sup_{t \in [0, T], x, y \in \mathbb{R}^d, y \neq 0} \frac{y^\top c(t, x) y}{|y|^2} \leq \mathfrak{s}_+ \right\} , \end{aligned}$$

pour $T > 0$, $d \in \mathbb{N}^*$, $\mathfrak{s}_+, \mathfrak{s}_-, \mathfrak{c}_1 > 0$.

Enfin, définissons les ensembles auxquels appartiennent les *fonctions de drift* :

$$\begin{aligned} \mathfrak{B}_{T, d} &:= \left\{ b : [0, T] \times \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \rightarrow \mathbb{R}^d \text{ mesurable ;} \right. \\ &\quad \left. \sup_{t \in [0, T]} |b(t, 0, \delta_0)| < \infty , \sup_{\substack{t \in [0, T], x, x' \in \mathbb{R}^d, \nu, \nu' \in \mathcal{P}_1(\mathbb{R}^d) \\ (x, \nu) \neq (x', \nu')}} \frac{|b(t, x', \nu') - b(t, x, \nu)|}{|x' - x| + \mathcal{W}_1(\nu', \nu)} < \infty \right\} , \end{aligned}$$

ainsi que ses sous-ensembles

$$\begin{aligned} \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3} &:= \left\{ b \in \mathfrak{B}_{T, d} ; \sup_{t \in [0, T]} |b(t, 0, \delta_0)| \leq \mathfrak{c}_2 , k_b^* \leq \mathfrak{h} , m_b \leq \mathfrak{n} , \right. \\ &\quad \left. \max(|b|_{\text{Lip}}, |\delta_\nu^1 b|_{\text{Lip}, \infty}, \dots, |\delta_\nu^{k_b^*} b|_{\text{Lip}, \infty}, |\delta_\nu^{m_b} b|_{\text{Lip}}) \leq \mathfrak{c}_3 \right\} , \\ \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}^{\mathfrak{r}_1, \mathfrak{r}_2, \mathfrak{c}_4} &:= \left\{ b \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3} ; \exists \vartheta \in \Theta, b(t, x, \nu) = b_0(\vartheta; t, x, \nu), \forall (t, x, \nu) \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) , \right. \\ &\quad \left. r_1(b_0) \leq \mathfrak{r}_1, r_2(b_0) \leq \mathfrak{r}_2, |D_\vartheta^{(2)} b_0|_{\text{Lip}, \infty} \leq \mathfrak{c}_4 \right\} , \end{aligned}$$

pour $T > 0$, $d \in \mathbb{N}^*$, $\mathfrak{h}, \mathfrak{n} \in \mathbb{N}$, $\mathfrak{c}_2, \mathfrak{c}_3 \geq 0$, $\mathfrak{r}_1, \mathfrak{r}_2 \in \mathbb{N}^*$, $\mathfrak{c}_4 \geq 0$.

Nous allons considérer un ellipsoïde de notre modèle paramétré par

$$\begin{aligned} \mathfrak{b} &:= (T, d, \mathfrak{m} = (\mathfrak{m}_{2\ell})_{\ell \geq 0}, \mathfrak{s}_+, \mathfrak{s}_-, \mathfrak{c}_1, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3, \mathfrak{r}_1, \mathfrak{r}_2, \mathfrak{c}_4) \in \mathfrak{E} , \\ \mathfrak{E} &:= \mathbb{R}_+^* \times \mathbb{N}^* \times (\mathbb{R}_+)^{\mathbb{N}} \times (\mathbb{R}_+^*)^3 \times (\mathbb{N})^2 \times (\mathbb{R}_+^*)^2 \times (\mathbb{N}^*)^2 \times \mathbb{R}_+^* , \end{aligned}$$

et, dans toute la suite, de manière générale, la notation $A \lesssim B$ signifiera qu'il existe une constante $C = C_{\mathfrak{b}}$, dépendant uniquement de $\mathfrak{b}$, telle que $A \leq C_{\mathfrak{b}} B$.

Finalement, dans le cadre paramétrique (resp. non-paramétrique), nous considérerons le modèle

$$(b, c, \mu_0) \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}^{\mathfrak{r}_1, \mathfrak{r}_2, \mathfrak{c}_4} \times \mathfrak{C}_{T, d, \mathfrak{s}_+, \mathfrak{s}_-, \mathfrak{c}_1} \times \mathfrak{M}_{d, \mathfrak{m}} \quad (\text{resp. } (b, c, \mu_0) \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3} \times \mathfrak{C}_{T, d, \mathfrak{s}_+, \mathfrak{s}_-, \mathfrak{c}_1} \times \mathfrak{M}_{d, \mathfrak{m}}^*).$$

2.3 Quelques propriétés classiques & D'autres moins classiques

Sur l'EDP & l'EDS de McKean-Vlasov

En suivant le cadre fourni dans la monographie de Bogachev *et al.* [17] (section 6.7(iii), correspondant à l'article [143] de Manita et Shaposhnikov), le problème (2) est compris au sens faible suivant.

Définition 24. μ est dite solution de probabilité du problème (2) si μ est une mesure borélienne sur $[0, T] \times \mathbb{R}^d$, de la forme $\mu = (\mu_t)_{t \in [0, T]}$ i.e. $\mu(dt, dx) = \mu_t(dx)dt$, vérifiant

- (i) pour tout $t \in [0, T]$, μ_t est une mesure de probabilité sur $\mathbb{R}^d$;
- (ii) pour tous $1 \leq i, j \leq d$, les fonctions

$$(t, x) \in [0, T] \times \mathbb{R}^d \mapsto c_{ij}(t, x), (t, x) \in [0, T] \times \mathbb{R}^d \mapsto b^i(t, x, \mu_t)$$

sont mesurables, et dans $L^1([0, T] \times U, \mu)$ pour toute boule fermée $U \subset \mathbb{R}^d$;

- (iii) pour toute fonction test $\Phi \in \mathcal{C}_c^\infty(\mathbb{R}^d; \mathbb{R})$ et tout $t \in [0, T]$,

$$\int_{\mathbb{R}^d} \Phi d\mu_t - \int_{\mathbb{R}^d} \Phi d\mu_0 = \int_0^t \int_{\mathbb{R}^d} L_{c,b}^\mu(\Phi) d\mu_s ds ,$$

où l'opérateur $L_{c,b}^\mu$ est défini par :

$$L_{c,b}^\mu \Phi(t, x) := \frac{1}{2} \sum_{i,j=1}^d \partial_{x_i x_j}^2 \Phi(x) c_{ij}(t, x) + \sum_{i=1}^d \partial_{x_i} \Phi(x) b^i(t, x, \mu_t) .$$

Pour établir certaines propriétés sur $\bar{\mathbb{P}}_b$, nous aurons besoin également de l'outil suivant : pour $(c, \mu_0) \in \mathfrak{C}_{T,d,s_+,s_-,c_1} \times \mathfrak{M}_{d,m}$, et $b, b' \in \mathfrak{B}_{T,d,h,n,c_2,c_3}$, nous noterons $\bar{X}^{b',b}$ l'unique solution forte sur $(\mathcal{C}, \mathcal{F}_T)$, sous $\bar{\mathbb{P}}_b$, de

$$\begin{cases} d\bar{X}_t^{b',b} = b'(t, \bar{X}_t^{b',b}, \mu_t^{b'}) dt + \sigma(t, \bar{X}_t^{b',b}) d\bar{B}_t^b, & t \in [0, T], \\ \bar{X}_{|t=0}^{b',b} = X_0, \end{cases}$$

$\bar{X}^{b',b}$ est donc, par construction, couplé avec X et vérifie $\mathcal{L}_{\bar{\mathbb{P}}_b}(\bar{X}^{b',b}) = \bar{\mathbb{P}}_{b'}$ et $\mathcal{L}_{\bar{\mathbb{P}}_b}(\bar{X}_t^{b',b}) = \mu_t^{b'}$ pour tout $t \in [0, T]$. Lorsque $b, b' \in \mathfrak{B}_{T,d,h,n,c_2,c_3}^{r_1,r_2,r}$ sont de la forme $b = b_0(\vartheta; \cdot), b' = b_0(\vartheta'; \cdot), \vartheta, \vartheta' \in \Theta$, nous noterons également respectivement $\bar{X}^{\vartheta',\vartheta}$ à la place de $\bar{X}^{b',b}$.

Nous regroupons dans le théorème suivant des propriétés très utiles vérifiées par l'équation de McKean-Vlasov. La preuve en sera faite en Annexe 1.

Théorème 25. Soit $(c, \mu_0) \in \mathfrak{C}_{T,d,s_+,s_-,c_1} \times \mathfrak{M}_{d,m}$. Alors

- (i) Pour tout $b \in \mathfrak{B}_{T,d,h,n,c_2,c_3}$, la loi de probabilité $\bar{\mathbb{P}}_b$ sur $(\mathcal{C}, \mathcal{F}_T)$ est également bien définie, l'application $b \in \mathfrak{B}_{T,d,h,n,c_2,c_3} \mapsto \mu^b = \mathcal{S}(b, c, \mu_0)$ est bien définie, $\mu^b \in \mathcal{C}([0, T]; \mathcal{P}_1(\mathbb{R}^d))$ et $(\mu_t^b)_{t \in [0, T]}$ est l'unique solution de probabilité du problème de Cauchy (2) ;
- (ii) Pour tout entier $r \geq 1$, pour tout $b \in \mathfrak{B}_{T,d,h,n,c_2,c_3}$ (resp. $b \in \mathfrak{B}_{T,d,h,n,c_2,c_3}^{r_1,r_2,r}$),

$$\sup_{t \in [0, T]} \mathbf{m}_r(\mu_t^b) \leq \mathbb{E}_{\bar{\mathbb{P}}_b} \left[\sup_{t \in [0, T]} |X_t|^r \right] \lesssim 1 \quad \left(\text{resp.} \quad \sup_{(\vartheta, t) \in \Theta \times [0, T]} \mathbf{m}_r(\mu_t^\vartheta) \leq \sup_{\vartheta \in \Theta} \mathbb{E}_{\bar{\mathbb{P}}_\vartheta} \left[\sup_{t \in [0, T]} |X_t|^r \right] \lesssim 1 \right),$$

ce qui entraîne que, pour tout $(t, x) \in [0, T] \times \mathbb{R}^d$ (resp. $(\vartheta, t, x) \in \Theta \times [0, T] \times \mathbb{R}^d$)

$$|b(t, x, \mu_t^b)| \leq \sup_{t' \in [0, T]} |b(t', 0, \delta_0)| + |b|_{Lip}(|x| + \sup_{t' \in [0, T]} \mathbf{m}_1(\mu_{t'}^b))$$

$$\left(\text{resp. } |b_0(\vartheta; t, x, \mu_t^\vartheta)| \leq \sup_{(\vartheta', t') \in \Theta \times [0, T]} |b_0(\vartheta'; t', 0, \delta_0)| + |b_0|_{Lip}(|x| + \sup_{(\vartheta', t') \in \Theta \times [0, T]} \mathfrak{m}_1(\mu_{t'}^{\vartheta'}) \right),$$

et donc pour tous $x_0 \in \mathbb{R}^d$, $R > 0$, il existe $\kappa_5 = \kappa_5(x_0, R, \mathfrak{b}) > 0$, tel que

$$\sup_{t \in [0, T], |x - x_0| \leq R} |b(t, x, \mu_t^b)| \leq \kappa_5 \quad (\text{resp.} \quad \sup_{(\vartheta, t) \in \Theta \times [0, T], |x - x_0| \leq R} |b_0(\vartheta; t, x, \mu_t^\vartheta)| \leq \kappa_5);$$

(iii) Si $\mu_0 \in \mathfrak{M}_{d, \mathfrak{m}}^*$, il existe $\mathfrak{c}_5 = \mathfrak{c}_5(\mathfrak{b}) > 0$, tel que pour tout $b \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}$ et tout entier $q \geq 1$,

$$\sup_{t \in [0, T]} \mathbb{E}_{\mathbb{P}_b} [|X_t|^{2q}] \leq q! \mathfrak{c}_5^q;$$

en particulier, X_t est sous-gaussienne sous $\mathbb{P}_b$ (voir la définition 2.30 du Chapitre 2 et l'article [31]);

(iv) Pour tout $b \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}$, $t \in [0, T] \mapsto \mu_t^b \in \mathcal{P}_1(\mathbb{R}^d)$ est régulière :

$$\mathcal{W}_1(\mu_{t_2}^b, \mu_{t_1}^b) \leq \mathbb{E}_{\mathbb{P}_b} [|X_{t_2} - X_{t_1}|] \lesssim |t_2 - t_1|^{1/2} \quad \text{pour tous } t_1, t_2 \in [0, T];$$

(v) L'application $b \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3} \mapsto \mu^b = \mathcal{S}(b, c, \mu_0)$ est régulière en distance de Wasserstein $\mathcal{W}_1$ et en variation totale : pour tous $b_1, b_2 \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}$,

$$\sup_{s \in [0, t]} \mathcal{W}_1(\mu_s^{b_2}, \mu_s^{b_1}) \lesssim \left(\int_0^t \int_{\mathbb{R}^d} |b_2(s, x, \mu_s^{b_1}) - b_1(s, x, \mu_s^{b_1})|^2 \mu_s^{b_1}(dx) ds \right)^{1/2}, \quad t \in [0, T], \quad i = 1, 2,$$

et, étant donné que $\mathbb{P}_{b_1}$ et $\mathbb{P}_{b_2}$ sont équivalentes avec, pour tout $t \in [0, T]$,

$$\begin{aligned} \log \frac{d\mathbb{P}_{b_1}}{d\mathbb{P}_{b_2}}(X)_{|\mathcal{F}_t} &= \int_0^t \left((c^{-1}b_1)(s, X_s, \mu_s^{b_1}) - (c^{-1}b_2)(s, X_s, \mu_s^{b_2}) \right)^\top dX_s \\ &\quad - \frac{1}{2} \int_0^t \left(|(c^{-1/2}b_1)(s, X_s, \mu_s^{b_1})|^2 - |(c^{-1/2}b_2)(s, X_s, \mu_s^{b_2})|^2 \right) ds, \end{aligned}$$

$$\sup_{s \in [0, t]} \|\mu_s^{b_2} - \mu_s^{b_1}\|_{TV} \lesssim \left(\int_0^t \int_{\mathbb{R}^d} |b_2(s, x, \mu_s^{b_2}) - b_1(s, x, \mu_s^{b_1})|^2 \mu_s^{b_1}(dx) ds \right)^{1/2}, \quad t \in [0, T], \quad i = 1, 2;$$

dans le cadre paramétrique, $\vartheta \in \Theta \mapsto \mu^\vartheta$ est même lipschitzienne :

$$\max \left(\sup_{t \in [0, T]} \mathcal{W}_1(\mu_t^{\vartheta_2}, \mu_t^{\vartheta_1}); \sup_{t \in [0, T]} \|\mu_t^{\vartheta_2} - \mu_t^{\vartheta_1}\|_{TV} \right) \lesssim |\vartheta_2 - \vartheta_1|, \quad \text{pour tous } \vartheta_1, \vartheta_2 \in \Theta;$$

(vi) Pour tout $b \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}$, pour tout $t \in (0, T)$, μ_t^b admet une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$, que nous noterons également $\mu_t^b : \mu^b(dt, dx) = \mu_t^b(x) dx dt$; de plus, il existe un représentant de μ^b continu sur $(0, T) \times \mathbb{R}^d$, et même localement Hölder sur $(0, T) \times \mathbb{R}^d$, que nous noterons également μ^b : plus précisément, pour tout $\alpha, \beta \in (0, \frac{1}{2})$,

$$\mu^b \in \mathcal{H}_{loc}^{\alpha, \beta}((0, T) \times \mathbb{R}^d);$$

(vii) Dans le cadre paramétrique, nous pouvons, pour chaque $\vartheta \in \Theta$, choisir un représentant de la densité μ^ϑ tel que

$$((\vartheta, t, x) \in \Theta \times (0, T) \times \mathbb{R}^d \mapsto \mu_t^\vartheta(x)) \in \mathcal{C}(\Theta \times (0, T) \times \mathbb{R}^d; \mathbb{R});$$

l'application $\vartheta \in \Theta \mapsto \mu^\vartheta \in L^1((0, T) \times \mathbb{R}^d)$ est alors également continue (ce qui signifie que l'application $\vartheta \in \Theta \mapsto \mu^\vartheta$ est continue en variation totale);

(viii) Pour tout $b \in \mathfrak{B}_{T,d,h,n,c_2,c_3}$, μ^b est strictement positive en tout point :

$$\text{pour tout } (t, x) \in (0, T) \times \mathbb{R}^d, \mu_t^b(x) > 0 ;$$

de plus μ^b est localement bornée : pour tous $x_0 \in \mathbb{R}^d$, $R > 0$ et $[R_1, R_2] \subset (0, T)$, il existe $\kappa_3 = \kappa_3(x_0, R, R_1, R_2, b)$, $\kappa_4 = \kappa_4(x_0, R, R_1, R_2, b) > 0$ tel que

$$\kappa_4 \leq \inf_{t \in [R_1, R_2], |x - x_0| \leq R} \mu_t^b(x) \leq \sup_{t \in [R_1, R_2], |x - x_0| \leq R} \mu_t^b(x) \leq \kappa_3 ;$$

(ix) L'application $b \in \mathfrak{B}_{T,d,h,n,c_2,c_3} \cap \mathcal{C}([0, T] \times \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d); \mathbb{R}^d) \mapsto \mu^b = \mathcal{S}(b, c, \mu_0)$ est "injective" au sens où, pour $b_1, b_2 \in \mathfrak{B}_{T,d,h,n,c_2,c_3} \cap \mathcal{C}([0, T] \times \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d); \mathbb{R}^d)$ tel que $\mu^{b_1} = \mu^{b_2}$,

$$b_1(t_0, x_0, \mu_{t_0}^{b_1}) = b_2(t_0, x_0, \mu_{t_0}^{b_1}) \text{ pour tout } (t_0, x_0) \in [0, T] \times \mathbb{R}^d ;$$

en particulier, pour $b_i = b_0(\vartheta_i; \cdot)$, $\vartheta_i \in \Theta$, $i = 1, 2$, si $\mu^{\vartheta_1} = \mu^{\vartheta_2}$, alors

$$b_0(\vartheta_1; t_0, x_0, \mu_{t_0}^{\vartheta_1}) = b_0(\vartheta_2; t_0, x_0, \mu_{t_0}^{\vartheta_1}) \text{ pour tout } (t_0, x_0) \in [0, T] \times \mathbb{R}^d .$$

Preuve. La preuve est détaillée en Annexe 1 (sauf pour le point (iii) dont la preuve est donnée dans la Section 8.1 du Chapitre 2).

Les points (i) et (ii) sont prouvés en Section 1.1 et 1.3 (existence, unicité et moments uniformément bornés de μ^b), le point (iv) en Section 1.2 (régularité en temps de μ^b), le point (v) en Section 1.4 (régularité de $b \mapsto \mu^b$), le point (vi) en Section 1.5 (existence et régularité Hölder de la densité), le point (vii) en Section 1.9 (régularité dans le cadre paramétrique), le point (viii) en Section 1.6 (positivité de μ^b), le point (ix) en Section 1.7 ("injectivité" de $b \mapsto \mu^b$). $\square$

Pour μ_0 admettant une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ (que nous noterons également μ_0), nous allons à présent énoncer un ensemble d'hypothèses supplémentaires qui entraînent, associées à nos hypothèses de base, que μ^b est solution au sens classique du problème de Cauchy (2), c'est-à-dire que $\mu^b \in \mathcal{C}^{1,2}((0, T) \times \mathbb{R}^d; \mathbb{R})$ vérifie pour tout $(t, x) \in (0, T) \times \mathbb{R}^d$

$$\begin{aligned} \partial_t \mu_t^b(x) &= \frac{1}{2} \sum_{i,j=1}^d c_{ij}(t, x) \partial_{x_i x_j}^2 \mu_t^b(x) + \sum_{i=1}^d \left(\sum_{j=1}^d \partial_{x_j} c_{ij}(t, x) - \beta^i(t, x) \right) \partial_{x_i} \mu_t^b(x) \\ &\quad + \left(\frac{1}{2} \sum_{i,j=1}^d \partial_{x_i x_j}^2 c_{ij}(t, x) - \sum_{i=1}^d \partial_{x_i} \beta^i(t, x) \right) \mu_t^b(x) , \end{aligned}$$

où nous avons noté $\beta : (t, x) \in [0, T] \times \mathbb{R}^d \mapsto b(t, x, \mu_t^b)$. Nous verrons de plus que μ^b peut s'écrire à l'aide d'une *solution fondamentale* du problème, construite à l'aide d'une méthode dite *paramétrix* dont les prémisses sont dûs à Levi au début du XX^e siècle (voir [128], et les monographies [75], [118], [113]).

Le jeu d'hypothèses que nous considérerons est tiré de l'article d'Angiuli et Lorenzi [4], mais l'on peut faire remonter ce type d'hypothèses aux travaux fondateurs sur les EDP paraboliques à coefficients non bornés des années 60 (Nash, Aronson, Besala, Krzyzanski, Bodanko, Chen, Kuroda, Eidel'man, ...), et plus particulièrement à l'article d'Aronson et Besala [5]. Nous pouvons aussi mentionner les travaux des années 2000-2010 de Addona, Lunardi, Metafunne, Pallara, Rhandi ... (voir la monographie de Lunardi [139] pour une vision d'ensemble sur l'aspect *semi-groupes* de ce type de problèmes), et l'article de Deck et Kruse [63].

Hypothèse 26. Il existe $\delta \in (0, 1)$ et $C > 0$ tel que, pour tous $1 \leq i, j \leq d$,

(i) $\mu_0 \in \mathcal{C}_b(\mathbb{R}^d; \mathbb{R})$;

(ii) $c_{ij}, \beta^i \in \mathcal{C}_{loc}^{\frac{\delta}{2}, \delta}((0, T) \times \mathbb{R}^d; \mathbb{R})$;

(iii) $\partial_{x_i} c_{ij}, \partial_{x_i x_j}^2 c_{ij}, \partial_{x_i} \beta^i$ existent et sont dans $\mathcal{C}_{loc}^{\frac{\delta}{2}, \delta}((0, T) \times \mathbb{R}^d; \mathbb{R})$;

(iv) $|\partial_{x_i x_j}^2 c_{ij}(t, x)| \leq C$ pour tout $(t, x) \in (0, T) \times \mathbb{R}^d$.

Remarque 27. Etant donné que $(b, c) \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3} \times \mathfrak{C}_{T, d, \mathfrak{s}_+, \mathfrak{s}_-, \mathfrak{c}_1}$, nous savons déjà que

- c est lipschitzienne en $x \in \mathbb{R}^d$ uniformément par rapport à $t \in [0, T]$, donc a fortiori, pour tout $\delta \in (0, 1)$, c est localement δ -Hölder en $x \in \mathbb{R}^d$ et, de plus, $\partial_{x_j} c$ existe au sens généralisé et est borné par $|c|_{\text{Lip}}$;
- pour $t_0 \in [0, T]$, $(t, x) \mapsto b(t_0, x, \mu_t^b)$ est dans $\mathcal{H}^{\frac{1}{2}, 1}([0, T] \times \mathbb{R}^d)$, car b est lipschitzienne en (x, ν) et pour tout $t, t' \in [0, T]$, $\mathcal{W}_1(\mu_{t'}^b, \mu_t^b) \lesssim |t' - t|^{1/2}$, donc une condition suffisante pour qu'il existe $\delta \in (0, 1)$, tel que β soit dans $\mathcal{C}_{loc}^{\frac{\delta}{2}, \delta}((0, T) \times \mathbb{R}^d; \mathbb{R}^d)$ est

$$|b(t', x', \nu') - b(t, x, \nu)| \leq C(|t' - t|^{1/2} + |x' - x| + \mathcal{W}_1(\nu', \nu)) , \quad (17)$$

pour un certain $C > 0$, et dans ce cas, pour tout $\delta \in (0, 1)$, $\beta \in \mathcal{C}_{loc}^{\frac{\delta}{2}, \delta}((0, T) \times \mathbb{R}^d; \mathbb{R}^d)$.

- les dérivées partielles $\partial_{x_i} \beta^i, 1 \leq i \leq d$ existent au sens généralisé et sont bornées par $|b|_{\text{Lip}}$.

Nous disposons donc du Théorème suivant :

Théorème 28. Soit $(b, c, \mu_0) \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3} \times \mathfrak{C}_{T, d, \mathfrak{s}_+, \mathfrak{s}_-, \mathfrak{c}_1} \times \mathfrak{M}_{d, \mathfrak{m}}$. S'il existe $\delta \in (0, 1)$ tel que

(i) $(b, c, \mu_0) \in \mathcal{C}^{0, 1, 0}((0, T) \times \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d); \mathbb{R}^d) \times \mathcal{C}^{0, 2}((0, T) \times \mathbb{R}^d; \mathcal{M}_d(\mathbb{R})) \times \mathcal{C}_b(\mathbb{R}^d; \mathbb{R})$,

(ii) $c_{ij}, \partial_{x_i} c_{ij}, \partial_{x_i x_j}^2 c_{ij} \in \mathcal{C}_{loc}^{\frac{\delta}{2}, \delta}((0, T) \times \mathbb{R}^d; \mathbb{R})$ pour tous $1 \leq i, j \leq d$, et $D_{x_c}^2 c$ est borné,

(iii) b vérifie (17), $\nu \mapsto \partial_{x_i} b^i(t, x, \nu)$ est lipschitzienne uniformément par rapport à (t, x) , et $\partial_{x_i} \beta^i \in \mathcal{C}_{loc}^{\frac{\delta}{2}, \delta}((0, T) \times \mathbb{R}^d; \mathbb{R})$ pour tout $1 \leq i \leq d$,

alors μ^b est solution au sens classique du problème de Cauchy (2) et

$$\mu^b \in \mathcal{C}_b([0, T] \times \mathbb{R}^d) \cap \mathcal{C}_{loc}^{1+\frac{\delta}{2}, 2+\delta}((0, T) \times \mathbb{R}^d) .$$

Notons $D := \{(s, \xi; t, x) \in [0, T] \times \mathbb{R}^d \times [0, T] \times \mathbb{R}^d ; s \leq t\}$. Il existe de plus une fonction (dite solution fondamentale du problème (2)) $Z : D \mapsto \mathbb{R}_+^*$ tel que pour tout $(s, \xi) \in [0, T] \times \mathbb{R}^d$, $Z(s, \xi; \cdot, \cdot) \in \mathcal{C}_{loc}^{1+\frac{\delta}{2}, 2+\delta}((s, T) \times \mathbb{R}^d)$ et est solution de l'EDP (2) sur $(s, T) \times \mathbb{R}^d$, et pour tous $0 \leq s < t \leq T$, et $x \in \mathbb{R}^d$,

$$\mu_t^b(x) = \int_{\mathbb{R}^d} Z(s, \xi; t, x) \mu_s^b(\xi) d\xi .$$

En particulier, pour tout $(t, x) \in (0, T] \times \mathbb{R}^d$,

$$\mu_t^b(x) = \int_{\mathbb{R}^d} Z(0, \xi; t, x) \mu_0(\xi) d\xi .$$

Preuve. Voir l'article d'Angiuli et Lorenzi [4]. $\square$

Enfin, nous traitons du cas particulier dit de *Vlasov* (voir [173]), avec un *drift homogène en temps* et une *diffusion constante* de la forme suivante

$$b(t, x, \nu) = b(x, \nu) = \int_{\mathbb{R}^d} \tilde{b}(x, y) \nu(dy), \text{ où } \tilde{b}(x, y) = F(x - y) + G(x), \quad x, y \in \mathbb{R}^d,$$

$$\sigma(t, x) = \sigma I_d, \sigma > 0,$$

pour lequel le problème (2) se réécrit

$$\begin{cases} \partial_t \mu_t^b = \frac{\sigma^2}{2} \Delta \mu_t^b - \operatorname{div}((G + F \star \mu_t^b) \mu_t^b), & (t, x) \in (0, T) \times \mathbb{R}^d, \\ \mu_{|t=0}^b = \mu_0. \end{cases} \quad (18)$$

Théorème 29. *Supposons que*

- (i) $\mu_0 \in \mathcal{C}_b(\mathbb{R}^d; \mathbb{R})$,
 - (ii) G lipschitzienne et $G \in \mathcal{H}_{loc}^{\delta_G}(\mathbb{R}^d; \mathbb{R}^d)$, $\delta_G > 1$,
 - (iii) F est à support compact, $F \in \mathcal{H}_{loc}^{\delta_F}(\mathbb{R}^d; \mathbb{R}^d)$, $\delta_F > 1$,
- alors $\mu^b \in \mathcal{C}^{1,2}((0, T) \times \mathbb{R}^d) \cap \mathcal{C}_b([0, T] \times \mathbb{R}^d)$ est solution classique du problème (18). De plus,

- si $\delta_G \notin \mathbb{N}^*$,

$$\mu^b \in \mathcal{H}_{loc}^{\frac{1+\delta_G}{2}, 1+\delta_G}((0, T) \times \mathbb{R}^d),$$

- si $\delta_G \in \mathbb{N}^*$, pour tout $\delta \in (0, 1)$,

$$\mu^b \in \mathcal{H}_{loc}^{\lfloor \frac{\delta_G}{2} \rfloor + 1 + \delta, \lfloor \delta_G \rfloor + 1 + \delta}((0, T) \times \mathbb{R}^d).$$

Preuve. Voir l'Annexe 1.8. $\square$

Enfin, pour l'étude paramétrique de notre modèle dans le Chapitre 1, nous aurons besoin de l'outil suivant : pour $(c, \mu_0) \in \mathcal{C}_{T, d, s_+, s_-, c_1} \times \mathfrak{M}_{d, m}$, et $b_1, b'_1 \in \mathfrak{B}_{T, d, h, n, c_2, c_3}$, nous noterons $\mathbb{P}_{b_1, b'_1}$ l'unique loi sur $(\mathcal{C}, \mathcal{F}_T)$ sous laquelle $X_0 \sim \mu_0$ et

$$\left(\overline{B}_t^{b_1, b'_1} := \int_0^t c^{-1/2}(s, X_s) (dX_s - b_1(s, X_s, \mu_s^{b'_1}) ds) \right)_{t \in [0, T]}$$

est un mouvement brownien standard d -dimensionnel. Le processus canonique X sur $(\mathcal{C}, \mathcal{F}_T)$ est donc solution, sous $\mathbb{P}_{b_1, b'_1}$, de l'EDS *linéaire au sens de McKean*

$$\begin{cases} dX_t = b_1(t, X_t, \mu_t^{b'_1}) dt + \sigma(t, X_t) d\overline{B}_t^{b_1, b'_1}, & t \in [0, T], \\ \mathcal{L}(X_0) = \mu_0, \end{cases}$$

autrement dit, en notant $\beta^{b_1, b'_1} : (t, x) \in [0, T] \times \mathbb{R}^d \mapsto b_1(t, x, \mu_t^{b'_1})$, de

$$\begin{cases} dX_t = \beta^{b_1, b'_1}(t, X_t) dt + \sigma(t, X_t) d\overline{B}_t^{b_1, b'_1}, & t \in [0, T], \\ \mathcal{L}(X_0) = \mu_0, \end{cases}$$

ce qui nous permet d'une certaine manière de *découpler* le drift et le paramètre de mesure. Nous noterons, pour tout $t \in [0, T]$, $\mu_t^{b_1, b'_1} := \mathcal{L}_{\mathbb{P}_{b_1, b'_1}}(X_t)$, et nous retrouvons que, pour $b \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}$, $\bar{\mathbb{P}}_b = \bar{\mathbb{P}}_{b, b}$, $\bar{B}^b = \bar{B}^{b, b}$ et $\mu_t^b = \mu_t^{b, b}$.

Pour $b, b_1, b'_1 \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}$, une manière de coupler les lois $\bar{\mathbb{P}}_b$ et $\bar{\mathbb{P}}_{b_1, b'_1}$ sur $(\mathcal{C}, \mathcal{F}_T)$ est la suivante : nous noterons $\bar{X}^{(b_1, b'_1), b}$ l'unique solution forte, sous $\bar{\mathbb{P}}_b$, de

$$\begin{cases} d\bar{X}_t^{(b_1, b'_1), b} = b_1(t, \bar{X}_t^{(b_1, b'_1), b}, \mu_t^{b'_1})dt + \sigma(t, \bar{X}_t^{(b_1, b'_1), b})d\bar{B}_t^b, & t \in [0, T], \\ \bar{X}_{|t=0}^{(b_1, b'_1), b} = X_0, \end{cases}$$

et nous avons alors, par définition, que $\bar{X}^{(b_1, b'_1), b}$ est couplé avec le processus canonique X , $\mathcal{L}_{\bar{\mathbb{P}}_b}(\bar{X}^{(b_1, b'_1), b}) = \bar{\mathbb{P}}_{b_1, b'_1}$, et $\mu_t^{b_1, b'_1} = \mathcal{L}_{\bar{\mathbb{P}}_b}(\bar{X}_t^{(b_1, b'_1), b})$.

Lorsque $b, b_1, b'_1 \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}^{r_1, r_2, \mathfrak{r}}$ sont de la forme $b = b_0(\vartheta; \cdot)$, $b_1 = b_0(\vartheta_1; \cdot)$, $b'_1 = b_0(\vartheta'_1; \cdot)$, $\vartheta, \vartheta_1, \vartheta'_1 \in \Theta$, nous noterons également respectivement $\bar{\mathbb{P}}_{\vartheta_1, \vartheta'_1}$, $\bar{B}^{\vartheta_1, \vartheta'_1}$ et $\bar{X}^{(\vartheta_1, \vartheta'_1), \vartheta}$ à la place de $\bar{\mathbb{P}}_{b_1, b'_1}$, $\bar{B}^{b_1, b'_1}$, et $\bar{X}^{(b_1, b'_1), b}$.

Nous disposons alors des propriétés suivantes.

Théorème 30. *Soit $(c, \mu_0) \in \mathfrak{C}_{T, d, \mathfrak{s}_+, \mathfrak{s}_-, \mathfrak{c}_1} \times \mathfrak{M}_{d, \mathfrak{m}}$. Alors*

- (i) *Pour tous $b, b' \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}$, la loi de probabilité $\bar{\mathbb{P}}_{b, b'}$ sur $(\mathcal{C}, \mathcal{F}_T)$ est bien définie, l'application $(b, b') \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3} \times \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3} \mapsto \mu^{b, b'}$ est bien définie, et $\mu^{b, b'} \in \mathcal{C}([0, T]; \mathcal{P}_1(\mathbb{R}^d))$;*
- (ii) *Pour tout $b \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}^{r_1, r_2, \mathfrak{r}}$ et tout entier $q \geq 1$,*

$$\sup_{\vartheta, \vartheta_1, \vartheta'_1 \in \Theta} \sup_{t \in [0, T]} \mathbb{E}_{\bar{\mathbb{P}}_\vartheta} [|\bar{X}_t^{(\vartheta_1, \vartheta'_1), \vartheta}|^q] \lesssim 1 ;$$

- (iii) *La fonction $(\vartheta, \vartheta') \in \Theta \times \Theta \mapsto \mu^{\vartheta, \vartheta'}$ est régulière en distance de Wasserstein $\mathcal{W}_1$ et en variation totale : pour tout $b \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}^{r_1, r_2, \mathfrak{r}}$, tous $\vartheta, \vartheta_1, \vartheta'_1, \vartheta_2 \in \Theta$, et tout entier $q \geq 1$,*

$$\sup_{t \in [0, T]} \mathcal{W}_1(\mu_t^{\vartheta_2, \vartheta'_2}, \mu_t^{\vartheta_1, \vartheta'_1}) \leq \sup_{t \in [0, T]} \mathbb{E}_{\bar{\mathbb{P}}_\vartheta} [|\bar{X}_t^{(\vartheta_2, \vartheta'_2), \vartheta} - \bar{X}_t^{(\vartheta_1, \vartheta'_1), \vartheta}|^q]^{1/q} \lesssim |\vartheta_2 - \vartheta_1| + |\vartheta'_2 - \vartheta'_1| ,$$

$$\sup_{t \in [0, T]} \|\mu_t^{\vartheta_2, \vartheta'_2} - \mu_t^{\vartheta_1, \vartheta'_1}\|_{TV} \leq \|\bar{\mathbb{P}}_{\vartheta_2, \vartheta'_2} - \bar{\mathbb{P}}_{\vartheta_1, \vartheta'_1}\|_{TV} \lesssim |\vartheta_2 - \vartheta_1| + |\vartheta'_2 - \vartheta'_1| ;$$

en particulier, pour tous $\vartheta, \vartheta' \in \Theta$,

$$\max \left(\sup_{t \in [0, T]} \mathcal{W}_1(\mu_t^{\vartheta, \vartheta'}, \mu_t^{\vartheta}) ; \sup_{t \in [0, T]} \|\mu_t^{\vartheta, \vartheta'} - \mu_t^{\vartheta}\|_{TV} \right) \lesssim |\vartheta' - \vartheta| .$$

Preuve. Voir Annexe 1 □

Nous passons maintenant aux propriétés sur le système de particules en interaction et sa loi de probabilité, en particulier nous prouvons un résultat de type *limite de champ moyen*.

Sur le système de diffusions en interaction & sur $\mathbb{P}_b^N$

Pour établir certaines propriétés sur $\mathbb{P}_b^N$, nous aurons besoin des deux outils suivants : pour $(c, \mu_0) \in \mathfrak{C}_{T,d,s_+,s_-,c_1} \times \mathfrak{M}_{d,m}$, et $b, b' \in \mathfrak{B}_{T,d,h,n,c_2,c_3}$, nous noterons $\mathbb{X}^{N,b',b} = (X_t^{i,b',b})_{t \in [0,T], 1 \leq i \leq N}$ l'unique solution forte sur $(\mathcal{C}^N, \mathcal{F}_T^N)$, sous $\mathbb{P}_b^N$, de

$$\begin{cases} dX_t^{i,b',b} = b'(t, X_t^{i,b',b}, \mu_t^N(\mathbb{X}_t^{N,b',b}))dt + \sigma(t, X_t^{i,b',b})dB_t^{i,b}, & t \in [0, T], \quad 1 \leq i \leq N, \\ \mathbb{X}_{|t=0}^{N,b',b} = \mathbb{X}_0^N, \end{cases}$$

et $\bar{\mathbb{X}}^{N,b',b} = (\bar{X}_t^{i,b',b})_{t \in [0,T], 1 \leq i \leq N}$ l'unique solution forte sur $(\mathcal{C}^N, \mathcal{F}_T^N)$, sous $\mathbb{P}_b^N$, de

$$\begin{cases} d\bar{X}_t^{i,b',b} = b'(t, \bar{X}_t^{i,b',b}, \mu_t^{b'})dt + \sigma(t, \bar{X}_t^{i,b',b})d\bar{B}_t^{i,b}, & t \in [0, T], \quad 1 \leq i \leq N, \\ \bar{\mathbb{X}}_{|t=0}^{N,b',b} = \mathbb{X}_0^N. \end{cases}$$

Les processus $\mathbb{X}^{N,b',b}$ et $\bar{\mathbb{X}}^{N,b',b}$ sont donc, par construction, couplés avec $\mathbb{X}^N$, et vérifient $\mathcal{L}_{\mathbb{P}_b^N}(\mathbb{X}^{N,b',b}) = \mathbb{P}_{b'}^N$ et $\mathcal{L}_{\mathbb{P}_b^N}(\bar{\mathbb{X}}^{N,b',b}) = \bar{\mathbb{P}}_{b'}^{\otimes N}$.

Lorsque $b, b' \in \mathfrak{B}_{T,d,h,n,c_2,c_3}^{r_1,r_2,r}$ sont de la forme $b = b_0(\vartheta; \cdot)$, $b' = b_0(\vartheta'; \cdot)$, $\vartheta, \vartheta' \in \Theta$, nous noterons également respectivement $\mathbb{X}^{N,\vartheta',\vartheta}$ et $\bar{\mathbb{X}}^{N,\vartheta',\vartheta}$ à la place de $\mathbb{X}^{N,b',b}$ et $\bar{\mathbb{X}}^{N,b',b}$.

Nous avons les résultats suivants dont la preuve est faite dans l'Annexe 1 :

Théorème 31. Soit $(c, \mu_0) \in \mathfrak{C}_{T,d,s_+,s_-,c_1} \times \mathfrak{M}_{d,m}$. Alors

- (i) Pour tout $b \in \mathfrak{B}_{T,d,h,n,c_2,c_3}$, pour tout entier $N \geq 1$, $\mathbb{P}_b^N$ est bien définie ;
- (ii) Pour tout entier $r \geq 1$, pour tout $b \in \mathfrak{B}_{T,d,h,n,c_2,c_3}$ (resp. $b \in \mathfrak{B}_{T,d,h,n,c_2,c_3}^{r_1,r_2,r}$),

$$\sup_{N \geq 1} \sup_{1 \leq i \leq N} \mathbb{E}_{\mathbb{P}_b^N} \left[\sup_{t \in [0,T]} |X_t^i|^r \right] \lesssim 1 \quad \left(\text{resp. } \sup_{N \geq 1} \sup_{1 \leq i \leq N} \sup_{\vartheta \in \Theta} \mathbb{E}_{\mathbb{P}_b^N} \left[\sup_{t \in [0,T]} |X_t^i|^r \right] \lesssim 1 \right);$$

- (iii) Pour tout $b \in \mathfrak{B}_{T,d,h,n,c_2,c_3}$, $t \in [0, T] \mapsto \mu_t^N$ est régulière : pour tous $0 \leq t, t' \leq T$,

$$\sup_{N \geq 1} \mathbb{E}_{\mathbb{P}_b^N} [\mathcal{W}_1(\mu_t^N, \mu_{t'}^N)] \leq \sup_{N \geq 1} \mathbb{E}_{\mathbb{P}_b^N} [|X_t^N - X_{t'}^N|] \lesssim |t - t'|^{1/2}.$$

- (iv) Pour tous $b, b_1, b_2 \in \mathfrak{B}_{T,d,h,n,c_2,c_3}$, pour tout entier $N \geq 1$,

- $\mathbb{P}_b^N$ et $\bar{\mathbb{P}}_b^{\otimes N}$ sont équivalentes, et

$$\begin{aligned} \log \frac{d\mathbb{P}_b^N}{d\bar{\mathbb{P}}_b^{\otimes N}}(\mathbb{X}^N) &= \sum_{i=1}^N \int_0^T ((c^{-1}b)(t, X_t^i, \mu_t^N) - (c^{-1}b)(t, X_t^i, \mu_t^b))^\top dX_t^i \\ &\quad - \frac{1}{2} \sum_{i=1}^N (|(c^{-1/2}b)(t, X_t^i, \mu_t^N)|^2 - |(c^{-1/2}b)(t, X_t^i, \mu_t^b)|^2) dt \end{aligned}$$

- $\mathbb{P}_{b_1}^N$ et $\mathbb{P}_{b_2}^N$ sont équivalentes, et

$$\log \frac{d\mathbb{P}_{b_1}^N}{d\mathbb{P}_{b_2}^N}(\mathbb{X}^N) = \sum_{i=1}^N \int_0^T ((c^{-1}b_1)(t, X_t^i, \mu_t^N) - (c^{-1}b_2)(t, X_t^i, \mu_t^N))^\top dX_t^i$$

$$- \frac{1}{2} \sum_{i=1}^N (|(c^{-1/2}b_1)(t, X_t^i, \mu_t^N)|^2 - |(c^{-1/2}b_2)(t, X_t^i, \mu_t^N)|^2) dt$$

(v) Pour tout $b \in \mathfrak{B}_{T,d,h,n,c_2,c_3}$ (resp. $\vartheta \in \Theta$), la limite de champ moyen $\mu^N \xrightarrow{N \rightarrow \infty} \mu^b$ (resp. $\mu^N \xrightarrow{N \rightarrow \infty} \mu^\vartheta$) est vérifiée au sens où : pour tout entier $q \geq 1$,

$$\mathbb{E}_{\mathbb{P}_b^N} \left[\sup_{t \in [0,T]} \mathcal{W}_1(\mu_t^N, \mu_t^b)^q \right] \xrightarrow{N \rightarrow \infty} 0 \quad (\text{resp. } \sup_{\vartheta \in \Theta} \mathbb{E}_{\mathbb{P}_\vartheta^N} \left[\sup_{t \in [0,T]} \mathcal{W}_1(\mu_t^N, \mu_t^\vartheta)^q \right] \xrightarrow{N \rightarrow \infty} 0) ;$$

plus précisément, il existe $\delta > 0$ (dépendant de la dimension d de l'espace) tel que pour tous entiers $q \geq 1$ et $N \geq 1$,

$$\mathbb{E}_{\mathbb{P}_b^N} \left[\sup_{t \in [0,T]} \mathcal{W}_1(\mu_t^N, \mu_t^b)^q \right] \lesssim N^{-\delta q} \quad (\text{resp. } \sup_{\vartheta \in \Theta} \mathbb{E}_{\mathbb{P}_\vartheta^N} \left[\sup_{t \in [0,T]} \mathcal{W}_1(\mu_t^N, \mu_t^\vartheta)^q \right] \lesssim N^{-\delta q}) ;$$

de même,

$$\mathbb{E}_{\mathbb{P}_b^N} \left[\sup_{t \in [0,T]} \mathcal{W}_1(\mu_t^N, \mu^N(\bar{\mathbb{X}}_t^{N,b,b}))^q \right] \leq \sup_{1 \leq i \leq N} \mathbb{E}_{\mathbb{P}_b^N} \left[\sup_{t \in [0,T]} |X_t^i - \bar{X}_t^{i,b,b}|^q \right] \lesssim N^{-\delta q}$$

$$(\text{resp. } \sup_{\vartheta \in \Theta} \mathbb{E}_{\mathbb{P}_\vartheta^N} \left[\sup_{t \in [0,T]} \mathcal{W}_1(\mu_t^N, \mu^N(\bar{\mathbb{X}}_t^{N,\vartheta,\vartheta}))^q \right] \leq \sup_{1 \leq i \leq N} \sup_{\vartheta \in \Theta} \mathbb{E}_{\mathbb{P}_\vartheta^N} \left[\sup_{t \in [0,T]} |X_t^i - \bar{X}_t^{i,\vartheta,\vartheta}|^q \right] \lesssim N^{-\delta q}) ;$$

(vi) Pour tous $b, b' \in \mathfrak{B}_{T,d,h,n,c_2,c_3}$ (resp. $\vartheta, \vartheta' \in \Theta$), pour tous entiers $q \geq 1$ et $N \geq 1$,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_b^N} \left[\sup_{t \in [0,T]} \mathcal{W}_1(\mu_t^N, \mu^N(\bar{\mathbb{X}}_t^{N,b',b}))^q \right] &\leq \sup_{1 \leq i \leq N} \mathbb{E}_{\mathbb{P}_b^N} \left[\sup_{t \in [0,T]} |X_t^i - X_t^{i,b',b}|^q \right] \\ &\lesssim \sup_{1 \leq i \leq N} \left(\mathbb{E}_{\mathbb{P}_b^N} \left[\sup_{t \in [0,T]} |b(t, X_t^i, \mu_t^N) - b'(t, X_t^i, \mu_t^N)|^{2q} \right] \right)^{1/2} \end{aligned}$$

$$(\text{resp. } \mathbb{E}_{\mathbb{P}_\vartheta^N} \left[\sup_{t \in [0,T]} \mathcal{W}_1(\mu_t^N, \mu^N(\bar{\mathbb{X}}_t^{N,\vartheta',\vartheta}))^q \right] \leq \sup_{1 \leq i \leq N} \mathbb{E}_{\mathbb{P}_\vartheta^N} \left[\sup_{t \in [0,T]} |X_t^i - \bar{X}_t^{i,\vartheta',\vartheta}|^q \right] \lesssim |\vartheta - \vartheta'|^q) ;$$

(vii) Pour tous $b, b' \in \mathfrak{B}_{T,d,h,n,c_2,c_3}$ (resp. $\vartheta, \vartheta' \in \Theta$), pour tous entiers $q \geq 1$ et $N \geq 1$,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_b^N} \left[\sup_{t \in [0,T]} \mathcal{W}_1(\mu^N(\bar{\mathbb{X}}_t^{N,b',b}), \mu^N(\bar{\mathbb{X}}_t^{N,b,b}))^q \right] &\leq \sup_{1 \leq i \leq N} \mathbb{E}_{\mathbb{P}_b^N} \left[\sup_{t \in [0,T]} |\bar{X}_t^{i,b',b} - \bar{X}_t^{i,b,b}|^q \right] \\ &\lesssim \sup_{1 \leq i \leq N} \left(\mathbb{E}_{\mathbb{P}_b^N} \left[\sup_{t \in [0,T]} |b(t, \bar{X}_t^{i,b,b}, \mu^N(\bar{\mathbb{X}}_t^{N,b,b})) - b'(t, \bar{X}_t^{i,b,b}, \mu^N(\bar{\mathbb{X}}_t^{N,b,b}))|^{2q} \right] \right)^{1/2} \end{aligned}$$

$$\begin{aligned} (\text{resp. } \mathbb{E}_{\mathbb{P}_\vartheta^N} \left[\sup_{t \in [0,T]} \mathcal{W}_1(\mu^N(\bar{\mathbb{X}}_t^{N,\vartheta',\vartheta}), \mu^N(\bar{\mathbb{X}}_t^{N,\vartheta,\vartheta}))^q \right] &\leq \sup_{1 \leq i \leq N} \mathbb{E}_{\mathbb{P}_\vartheta^N} \left[\sup_{t \in [0,T]} |\bar{X}_t^{i,\vartheta',\vartheta} - \bar{X}_t^{i,\vartheta,\vartheta}|^q \right] \\ &\lesssim |\vartheta - \vartheta'|^q) . \end{aligned}$$

Preuve. Pour le point (iii), nous renvoyons à l'Annexe 1, pour le point (ii) nous renvoyons à l'Annexe 1 et au Chapitre 1, pour les points (v), (vi) et (vii), nous renvoyons au Chapitre 1. La preuve des autres points est détaillée en Annexe 2 : en Section 2.1 pour le point (i), en Section 2.2 pour le point (iv). $\square$

Maintenant que nous avons énoncé nos hypothèses de travail, ainsi que les propriétés probabilistes et analytiques qui nous seront utiles tout au long de cette thèse, nous pouvons passer à l'étude statistique de notre modèle.

3 Objectifs statistiques & Résultats obtenus

Notre objectif sera d'écrire un programme d'estimation de μ puis b , d'abord dans le cadre paramétrique, puis dans le cadre non-paramétrique. Pour cela, nous disposerons dans les deux cas de l'observation continue du système de particules $(\mathbb{X}_t^N = (X_t^1, \dots, X_t^N))_{t \in [0, T]}$. Nous considérerons donc, dans le cadre paramétrique, pour $(c, \mu_0) \in \mathfrak{C}_{T, d, \mathfrak{s}_+, \mathfrak{s}_-, \mathfrak{c}_1} \times \mathfrak{M}_{d, \mathfrak{m}}$ donné, les expériences statistiques suivantes :

$$\mathcal{E}^N = \left(\mathcal{C}^N, \mathcal{F}_T^N, (\mathbb{P}_b^N)_{b \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}^{r_1, r_2, \mathfrak{r}}} \right) = \left(\mathcal{C}^N, \mathcal{F}_T^N, (\mathbb{P}_\vartheta^N)_{\vartheta \in \Theta} \right),$$

l'expérience limite

$$\mathcal{H} = \left(\mathcal{C}, \mathcal{F}_T, (\overline{\mathbb{P}}_b)_{b \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}^{r_1, r_2, \mathfrak{r}}} \right) = \left(\mathcal{C}, \mathcal{F}_T, (\overline{\mathbb{P}}_\vartheta)_{\vartheta \in \Theta} \right),$$

et sa contrepartie N -tensorielle

$$\mathcal{H}^{\otimes N} = \left(\mathcal{C}^N, \mathcal{F}_T^N, (\overline{\mathbb{P}}_b^{\otimes N})_{b \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}^{r_1, r_2, \mathfrak{r}}} \right) = \left(\mathcal{C}^N, \mathcal{F}_T^N, (\overline{\mathbb{P}}_\vartheta^{\otimes N})_{\vartheta \in \Theta} \right),$$

la famille d'expériences limite $\mathcal{G}_{\vartheta_0}$ ($\vartheta_0 \in \Theta$) pour

$$\mathcal{G}_{\vartheta_0} = \left(\mathcal{C}, \mathcal{F}_T, (\overline{\mathbb{P}}_{\vartheta, \vartheta_0})_{\vartheta \in \Theta} \right)$$

et sa contrepartie N -tensorielle $\mathcal{G}_{\vartheta_0}^{\otimes N}$ ($\vartheta_0 \in \Theta$) pour

$$\mathcal{G}_{\vartheta_0}^{\otimes N} = \left(\mathcal{C}^N, \mathcal{F}_T^N, (\overline{\mathbb{P}}_{\vartheta, \vartheta_0}^{\otimes N})_{\vartheta \in \Theta} \right).$$

Dans le cadre non-paramétrique, pour $(c, \mu_0) \in \mathfrak{C}_{T, d, \mathfrak{s}_+, \mathfrak{s}_-, \mathfrak{c}_1} \times \mathfrak{M}_{d, \mathfrak{m}}^*$ donné, nous considérerons l'expérience statistique

$$\mathcal{E}^N = \left(\mathcal{C}^N, \mathcal{F}_T^N, (\mathbb{P}_b^N)_{b \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}} \right),$$

l'expérience limite

$$\mathcal{H} = \left(\mathcal{C}, \mathcal{F}_T, (\overline{\mathbb{P}}_b)_{b \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}} \right),$$

et l'expérience limite N -tensorielle

$$\mathcal{H}^{\otimes N} = \left(\mathcal{C}^N, \mathcal{F}_T^N, (\overline{\mathbb{P}}_b^{\otimes N})_{b \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}} \right).$$

La limite de champ moyen (voir le point (v) du Théorème 31) établit que l'expérience limite $\mathcal{H}^{\otimes N}$ est une approximation de l'expérience $\mathcal{E}^N$. Il est en fait même impossible, sous certaines conditions sur $\mathfrak{b}$, de discriminer statistiquement entre $\mathbb{P}_b^N$ et $\overline{\mathbb{P}}_b^{\otimes N}$ asymptotiquement. Plus précisément, nous avons le résultat suivant :

Proposition 32. *Pour $(c, \mu_0) \in \mathfrak{C}_{T, d, \mathfrak{s}_+, \mathfrak{s}_-, \mathfrak{c}_1} \times \mathfrak{M}_{d, \mathfrak{m}}^*$, il existe une constante $\mathfrak{f} = \mathfrak{f}_{\mathfrak{b}} > 0$ telle que*

$$\limsup_{N \rightarrow \infty} \sup_{b \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}} \mathbb{E}_{\overline{\mathbb{P}}_b^{\otimes N}} \left[\log \frac{d\overline{\mathbb{P}}_b^{\otimes N}}{d\mathbb{P}_b^N}(\mathbb{X}^N) \right] \leq \mathfrak{f} < \infty,$$

et donc, pour tout $b \in \mathfrak{B}_{T,d,h,n,c_2,c_3}$, il y a propagation du chaos :

$$\text{pour tout } 1 \leq K \leq N, \quad \sup_{t \in [0,T]} \|\mu_t^{K,N} - \mu_t^{\otimes K}\|_{TV} \leq \|\mathbb{P}_b^{K,N} - \bar{\mathbb{P}}_b^{\otimes K}\|_{TV} \lesssim \sqrt{\frac{K}{N}}.$$

Si de plus nous supposons que $\mathfrak{f} < 2$, alors nous avons

$$\limsup_{N \rightarrow \infty} \sup_{b \in \mathfrak{B}_{T,d,h,n,c_2,c_3}} \|\mathbb{P}_b^N - \bar{\mathbb{P}}_b^{\otimes N}\|_{TV} < 1.$$

Preuve Voir Annexe 3 □

Remarque 33. Cette dernière inégalité montre que, si $\mathfrak{b}$ vérifie certaines conditions préliminaires, alors il n'existe pas de *test statistique* de l'*hypothèse nulle*

$$H_0 : \mathbb{P}_b^N = \bar{\mathbb{P}}_b^{\otimes N}$$

contre l'*hypothèse alternative*

$$H_1 : \mathbb{P}_b^N \neq \bar{\mathbb{P}}_b^{\otimes N}$$

qui ait des *erreurs de première et seconde espèces* asymptotiquement arbitrairement petites quand $N \rightarrow \infty$. En effet, pour $b \in \mathfrak{B}_{T,d,h,n,c_2,c_3}$ fixé, un test qui rejette H_0 sur l'évènement $\mathcal{R}_N$ a pour erreur de première espèce $\bar{\mathbb{P}}_b^{\otimes N}(\mathcal{R}_N)$, et pour erreur de deuxième espèce $\mathbb{P}_b^N(\mathcal{R}_N^c)$, et la somme de ces deux erreurs vérifie donc

$$\bar{\mathbb{P}}_b^{\otimes N}(\mathcal{R}_N) + \mathbb{P}_b^N(\mathcal{R}_N^c) = (\mathbb{P}_b^N(\mathcal{R}_N) + \bar{\mathbb{P}}_b^{\otimes N}(\mathcal{R}_N^c)) - (\bar{\mathbb{P}}_b^{\otimes N}(\mathcal{R}_N) - \mathbb{P}_b^N(\mathcal{R}_N^c)) \geq 1 - \|\mathbb{P}_b^N - \bar{\mathbb{P}}_b^{\otimes N}\|_{TV}.$$

La somme des erreurs de première et seconde espèces de ce test est donc asymptotiquement minorée par une quantité strictement positive :

$$\liminf_{N \rightarrow \infty} (\bar{\mathbb{P}}_b^{\otimes N}(\mathcal{R}_N) + \mathbb{P}_b^N(\mathcal{R}_N^c)) > 0.$$

Nous présentons dans les Sections suivantes les résultats obtenus au cours de cette thèse, d'abord dans le cadre paramétrique, puis dans le cadre non-paramétrique. Nous comparerons ensuite ces résultats dans le cadre particulier des applications $\mathcal{S}$ qui s'expriment explicitement.

3.1 Estimation paramétrique

Nous présentons dans cette Section les résultats de Statistique paramétrique obtenus dans l'article [64], et nous nous appuyerons pour cela sur les outils et les résultats présentés dans les livres de Ibragimov et Hasminskii [98] et de Genon-Catalot et Picard [83]. Nous supposerons dans toute cette partie que

$$(b, c, \mu_0) \in \mathfrak{B}_{T,d,h,n,c_2,c_3}^{\mathfrak{r}_1, \mathfrak{r}_2, c_4} \times \mathfrak{C}_{T,d,s_+,s_-,c_1} \times \mathfrak{M}_{d,m},$$

et donc que

$$b \in \{b_0(\vartheta; \cdot); \vartheta \in \Theta\},$$

avec b_0 connue à ϑ près, et vérifiant les Hypothèses 8, 11 et 21.

Vraisemblance du modèle $\mathcal{E}^N$

Nous définissons la *log-vraisemblance du modèle $\mathcal{E}^N$* par

$$\ell^N(\vartheta; \mathbb{X}^N) = \sum_{i=1}^N \int_0^T (c^{-1}b_0)(\vartheta; t, X_t^i, \mu_t^N)^\top dX_t^i - \frac{1}{2} \sum_{i=1}^N \int_0^T |(c^{-1/2}b_0)(\vartheta; t, X_t^i, \mu_t^N)|^2 dt \quad (19)$$

pour tout $\vartheta \in \Theta$. En effet, d'après le Théorème de Girsanov, les lois $\mathbb{P}_\vartheta^N$ sont toutes absolument continues par rapport à $\mathbb{W}^N$, définie comme étant l'unique probabilité sur $(\mathcal{C}^N, \mathcal{F}_T^N)$ sous laquelle $\mathcal{L}(X_0^1, \dots, X_0^N) = \mu_0^{\otimes N}$ et les processus

$$\left(\int_0^t c^{-1/2}(s, X_s^i) dX_s^i \right)_{t \in [0, T]}, \quad 1 \leq i \leq N,$$

sont des mouvements browniens d -dimensionnels standards indépendants. De plus, pour tout $\vartheta \in \Theta$,

$$\frac{d\mathbb{P}_\vartheta^N}{d\mathbb{W}^N}(\mathbb{X}^N) = \exp(\ell^N(\vartheta; \mathbb{X}^N)) \quad \mathbb{W}^N\text{-presque sûrement.}$$

Nous noterons $\mathcal{L}^N(\cdot; \mathbb{X}^N) = \exp(\ell^N(\cdot; \mathbb{X}^N))$ pour la *vraisemblance du modèle $\mathcal{E}^N$* . En suivant la nomenclature de Ibragimov et Hasminski [98], nous avons

Proposition 34. $(\mathcal{E}^N)_{N \geq 1}$ est régulier, c'est-à-dire que

- (i) $\vartheta \mapsto \mathcal{L}^N(\vartheta; \mathbb{X}^N)$ est différentiable pour tout ϑ dans l'intérieur de Θ , $\mathbb{W}^N$ -presque sûrement,
- (ii) $\nabla_\vartheta \mathcal{L}^N(\vartheta; \mathbb{X}^N)$ est continue dans $L^2(\mathcal{C}^N, \mathbb{W}^N)$, pour tout ϑ dans l'intérieur de Θ ,
- (iii) l'information de Fisher est finie :

$$\mathbb{E}_{\mathbb{P}_\vartheta^N} [|\nabla_\vartheta \ell^N(\vartheta; \mathbb{X}^N)|^2] < \infty \text{ pour tout } \vartheta \text{ dans l'intérieur de } \Theta.$$

Pour tout ϑ dans l'intérieur de Θ , nous pouvons donc définir la matrice d'information de Fisher de $\mathcal{E}^N$ en ϑ par

$$\mathbb{I}_{\mathcal{E}^N}(\vartheta) := \mathbb{E}_{\mathbb{P}_\vartheta^N} [\nabla_\vartheta \ell^N(\vartheta; \mathbb{X}^N) \nabla_\vartheta \ell^N(\vartheta; \mathbb{X}^N)^\top],$$

qui peut aussi se réécrire

$$\begin{aligned} \mathbb{I}_{\mathcal{E}^N}(\vartheta) &= \sum_{i=1}^N \sum_{j=1}^d \mathbb{E}_{\mathbb{P}_\vartheta^N} \left[\int_0^T \nabla_\vartheta (c^{-1/2}b_0)^j(\vartheta; t, X_t^i, \mu_t^N) \nabla_\vartheta (c^{-1/2}b_0)^j(\vartheta; t, X_t^i, \mu_t^N)^\top dt \right] \\ &= \left(\sum_{i=1}^N \mathbb{E}_{\mathbb{P}_\vartheta^N} \left[\int_0^T \partial_{\vartheta_\ell} (c^{-1/2}b_0)(\vartheta; t, X_t^i, \mu_t^N)^\top \partial_{\vartheta_{\ell'}} (c^{-1/2}b_0)(\vartheta; t, X_t^i, \mu_t^N) dt \right] \right)_{1 \leq \ell, \ell' \leq p}. \end{aligned} \quad (20)$$

Exemples 35. Pour notre modèle de référence (4), en prenant $d = 1$ pour simplifier les expressions, la log-vraisemblance et son gradient s'expriment de la manière suivante :

$$\ell^N(\vartheta; \mathbb{X}^N) = \frac{1}{\sigma^2} \sum_{i=1}^N \int_0^T \left(\vartheta_1 X_t^i + \vartheta_2 + \frac{\vartheta_3}{N} \sum_{j=1}^N (X_t^j - X_t^i) \right) dX_t^i - \frac{1}{2\sigma} \sum_{i=1}^N \int_0^T \left(\vartheta_1 X_t^i + \vartheta_2 + \frac{\vartheta_3}{N} \sum_{j=1}^N (X_t^j - X_t^i) \right)^2 dt,$$

$$\nabla_{\vartheta} \ell^N(\vartheta; \mathbb{X}^N) = \begin{pmatrix} \frac{1}{\sigma^2} \sum_{i=1}^N \int_0^T X_t^i dX_t^i - \frac{1}{\sigma} \sum_{i=1}^N \int_0^T X_t^i (\vartheta_1 X_t^i + \vartheta_2 + \frac{\vartheta_3}{N} \sum_{j=1}^N (X_t^j - X_t^i)) dt \\ \frac{1}{\sigma^2} \sum_{i=1}^N (X_T^i - X_0^i) - \frac{1}{\sigma} \sum_{i=1}^N \int_0^T (\vartheta_1 X_t^i + \vartheta_2 + \frac{\vartheta_3}{N} \sum_{j=1}^N (X_t^j - X_t^i)) dt \\ \frac{1}{\sigma^2 N} \sum_{i,j=1}^N \int_0^T (X_t^j - X_t^i) dX_t^i - \frac{1}{\sigma N} \sum_{i,j=1}^N \int_0^T (X_t^j - X_t^i) (\vartheta_1 X_t^i + \vartheta_2 + \frac{\vartheta_3}{N} \sum_{k=1}^N (X_t^k - X_t^i)) dt \end{pmatrix}.$$

Notons à présent, pour tout ϑ dans l'intérieur de Θ ,

$$\begin{aligned} \mathbb{I}(\vartheta) &:= \sum_{j=1}^d \int_0^T \int_{\mathbb{R}^d} \nabla_{\vartheta} (c^{-1/2} b_0)^j(\vartheta; t, x, \mu_t^{\vartheta}) \nabla_{\vartheta} (c^{-1/2} b_0)^j(\vartheta; t, x, \mu_t^{\vartheta})^{\top} \mu_t^{\vartheta}(dx) dt \\ &= \left(\int_0^T \int_{\mathbb{R}^d} \partial_{\vartheta_{\ell}} (c^{-1/2} b_0)(\vartheta; t, x, \mu_t^{\vartheta})^{\top} \partial_{\vartheta_{\ell'}} (c^{-1/2} b_0)(\vartheta; t, x, \mu_t^{\vartheta}) \mu_t^{\vartheta}(dx) dt \right)_{1 \leq \ell, \ell' \leq p}. \end{aligned}$$

La fonction $\mathbb{I}$ à valeurs dans l'espace des matrices $\mathcal{M}_d(\mathbb{R})$ apparaît alors comme l'information de Fisher asymptotique normalisée de $\mathcal{E}^N$:

Proposition 36. *Pour tout ϑ dans l'intérieur de Θ ,*

$$N^{-1} \mathbb{I}_{\mathcal{E}^N}(\vartheta) \xrightarrow{N \rightarrow \infty} \mathbb{I}(\vartheta).$$

Preuve. Voir la Proposition 1.11 du Chapitre 1. □

Vraisemblance du modèle limite $\mathcal{H}$ & de sa contrepartie N -tensorielle $\mathcal{H}^{\otimes N}$

Nous définissons la *log-vraisemblance du modèle $\mathcal{H}$* par

$$\bar{\ell}(\vartheta; X) := \int_0^T (c^{-1} b_0)(\vartheta; t, X_t, \mu_t^{\vartheta})^{\top} dX_t - \frac{1}{2} \int_0^T |(c^{-1/2} b_0)(\vartheta; t, X_t, \mu_t^{\vartheta})|^2 dt, \quad \vartheta \in \Theta.$$

En effet, d'après le Théorème de Girsanov, les lois $\bar{\mathbb{P}}_{\vartheta}$, $\vartheta \in \Theta$, sont toutes absolument continues par rapport à W , définie comme étant l'unique loi de probabilité sur $(\mathcal{C}, \mathcal{F}_T)$ sous laquelle $\mathcal{L}(X_0) = \mu_0$ et sous laquelle le processus $\left(\int_0^t c^{-1/2}(s, X_s) dX_s \right)_{t \in [0, T]}$ est un mouvement brownien standard d -dimensionnel. De plus, pour tout $\vartheta \in \Theta$,

$$\frac{d\bar{\mathbb{P}}_{\vartheta}}{dW}(X) = \exp(\bar{\ell}(\vartheta; X)) \quad W\text{-presque sûrement.}$$

Nous noterons $\bar{\mathcal{L}}(\cdot; X) := \exp(\bar{\ell}(\cdot; X))$ pour la *vraisemblance du modèle $\mathcal{H}$* .

De même, pour tout entier $N \geq 1$, nous définissons la *log-vraisemblance du modèle $\mathcal{H}^{\otimes N}$* par

$$\bar{\ell}^N(\vartheta; \mathbb{X}^N) := \sum_{i=1}^N \bar{\ell}(\vartheta; X^i) = \sum_{i=1}^N \int_0^T (c^{-1} b_0)(\vartheta; t, X_t^i, \mu_t^{\vartheta})^{\top} dX_t^i$$

$$-\frac{1}{2} \sum_{i=1}^N \int_0^T |(c^{-1/2}b_0)(\vartheta; t, X_t^i, \mu_t^\vartheta)|^2 dt, \quad \vartheta \in \Theta,$$

puisque, d'après le Théorème de Girsanov, les lois $\bar{\mathbb{P}}_\vartheta^{\otimes N}$, $\vartheta \in \Theta$, sont toutes absolument continues par rapport à $\mathbb{W}^N$, et, pour tout $\vartheta \in \Theta$,

$$\frac{d\bar{\mathbb{P}}_\vartheta^{\otimes N}}{d\mathbb{W}^N}(\mathbb{X}^N) = \exp(\bar{\ell}^N(\vartheta; \mathbb{X}^N)) \quad \mathbb{W}^N\text{-presque sûrement.}$$

Nous noterons $\bar{\mathcal{L}}^N(\cdot; \mathbb{X}^N) := \exp(\bar{\ell}^N(\cdot; \mathbb{X}^N))$ pour la *vraisemblance du modèle* $\mathcal{H}^{\otimes N}$.

Introduisons une hypothèse supplémentaire nécessaire à l'étude de la régularité du modèle $\mathcal{H}$.

Hypothèse 37. La fonction $\vartheta \mapsto b(\vartheta; t, x, \mu_t^\vartheta)$ est différentiable sur l'intérieur de Θ , uniformément par rapport à $(t, x) \in (0, T) \times \mathbb{R}^d$.

L'expérience limite $\mathcal{H}$ vérifie les propriétés suivantes.

Proposition 38. Sous l'Hypothèse supplémentaire 37, nous avons que

- (i) $\mathcal{H}$ (resp. $\mathcal{H}^{\otimes N}$) est un modèle régulier ;
- (ii) l'information de Fisher $\mathbb{I}_{\mathcal{H}}$ de $\mathcal{H}$ est égale à l'information de Fisher normalisée $N^{-1}\mathbb{I}_{\mathcal{H}^{\otimes N}}$ de $\mathcal{H}^{\otimes N}$, et est donnée, pour tout ϑ dans l'intérieur de Θ , par

$$\begin{aligned} \mathbb{I}_{\mathcal{H}}(\vartheta) &= \mathbb{E}_{\bar{\mathbb{P}}_\vartheta} [\nabla_\vartheta \bar{\ell}(\vartheta; X) \nabla_\vartheta \bar{\ell}(\vartheta; X)^\top] \\ &= \sum_{j=1}^d \int_0^T \int_{\mathbb{R}^d} \nabla_\vartheta [(c^{-1/2}b_0)^j(\vartheta; t, x, \mu_t^\vartheta)] \nabla_\vartheta [(c^{-1/2}b_0)^j(\vartheta; t, x, \mu_t^\vartheta)]^\top \mu_t^\vartheta(dx) dt \\ &= \left(\int_0^T \int_{\mathbb{R}^d} \partial_{\vartheta_\ell} [(c^{-1/2}b_0)(\vartheta; t, x, \mu_t^\vartheta)]^\top \partial_{\vartheta_{\ell'}} [(c^{-1/2}b_0)(\vartheta; t, x, \mu_t^\vartheta)] \mu_t^\vartheta(dx) dt \right)_{1 \leq \ell, \ell' \leq p}. \end{aligned}$$

Remarque 39. Dans l'expression de $\mathbb{I}_{\mathcal{H}}(\vartheta)$ la dérivée en ϑ est prise sur la fonction "globale" $\vartheta \mapsto (c^{-1/2}b_0)(\vartheta; \cdot, \cdot, \mu_t^\vartheta)$, et donc, en toute généralité, $\mathbb{I}_{\mathcal{H}}(\vartheta)$ n'est pas égale à $\mathbb{I}(\vartheta)$.

Vraisemblance de la famille de modèles limite $\mathcal{G}_{\vartheta_0}$ ($\vartheta_0 \in \Theta$) & de sa contrepartie N -tensorielle $\mathcal{G}_{\vartheta_0}^{\otimes N}$ ($\vartheta_0 \in \Theta$)

Pour tout $\vartheta_0 \in \Theta$, nous définissons la *log-vraisemblance du modèle* $\mathcal{G}_{\vartheta_0}$ par

$$\bar{\ell}_{\vartheta_0}(\vartheta; X) := \int_0^T (c^{-1}b_0)(\vartheta; t, X_t, \mu_t^{\vartheta_0})^\top dX_t - \frac{1}{2} \int_0^T |(c^{-1/2}b_0)(\vartheta; t, X_t, \mu_t^{\vartheta_0})|^2 dt, \quad \vartheta \in \Theta.$$

En effet, d'après le Théorème de Girsanov, les lois $\bar{\mathbb{P}}_{\vartheta, \vartheta_0}$, $\vartheta \in \Theta$, sont toutes absolument continues par rapport à W , et de plus, pour tout $\vartheta \in \Theta$,

$$\frac{d\bar{\mathbb{P}}_{\vartheta, \vartheta_0}}{dW}(X) = \exp(\bar{\ell}_{\vartheta_0}(\vartheta; X)) \quad W\text{-presque sûrement.}$$

Nous noterons $\bar{\mathcal{L}}_{\vartheta_0}(\cdot; X) := \exp(\bar{\ell}_{\vartheta_0}(\cdot; X))$ pour la *vraisemblance du modèle* $\mathcal{G}_{\vartheta_0}$.

De même, pour tout $\vartheta_0 \in \Theta$ et tout entier $N \geq 1$, nous définissons la *log-vraisemblance du modèle* $\mathcal{G}_{\vartheta_0}^{\otimes N}$ par

$$\begin{aligned} \bar{\ell}_{\vartheta_0}^N(\vartheta; \mathbb{X}^N) &:= \sum_{i=1}^N \bar{\ell}_{\vartheta_0}(\vartheta; X^i) = \sum_{i=1}^N \int_0^T (c^{-1}b_0)(\vartheta; t, X_t^i, \mu_t^{\vartheta_0})^\top dX_t^i \\ &\quad - \frac{1}{2} \sum_{i=1}^N \int_0^T |(c^{-1/2}b_0)(\vartheta; t, X_t^i, \mu_t^{\vartheta_0})|^2 dt, \quad \vartheta \in \Theta, \end{aligned}$$

puisque, d'après le Théorème de Girsanov, les lois $\bar{\mathbb{P}}_{\vartheta, \vartheta_0}^{\otimes N}, \vartheta \in \Theta$, sont toutes absolument continues par rapport à $\mathbb{W}^N$, et, pour tout $\vartheta \in \Theta$,

$$\frac{d\bar{\mathbb{P}}_{\vartheta, \vartheta_0}^{\otimes N}}{d\mathbb{W}^N}(\mathbb{X}^N) = \exp(\bar{\ell}_{\vartheta_0}^N(\vartheta; \mathbb{X}^N)) \quad \mathbb{W}^N\text{-presque sûrement.}$$

Nous noterons $\bar{\mathcal{L}}_{\vartheta_0}^N(\cdot; \mathbb{X}^N) := \exp(\bar{\ell}_{\vartheta_0}^N(\cdot; \mathbb{X}^N))$ pour la *vraisemblance du modèle* $\mathcal{G}_{\vartheta_0}^{\otimes N}$.

La famille d'expériences limite $\mathcal{G}_{\vartheta_0}(\vartheta_0 \in \Theta)$ vérifie les propriétés suivantes.

Proposition 40. *Pour tout $\vartheta_0 \in \Theta$,*

- (i) $\mathcal{G}_{\vartheta_0}$ (resp. $\mathcal{G}_{\vartheta_0}^{\otimes N}$) est un modèle régulier ;
- (ii) l'information de Fisher $\mathbb{I}_{\mathcal{G}_{\vartheta_0}}$ de $\mathcal{G}_{\vartheta_0}$ est égale à l'information de Fisher normalisée $N^{-1}\mathbb{I}_{\mathcal{G}_{\vartheta_0}^{\otimes N}}$ de $\mathcal{G}_{\vartheta_0}^{\otimes N}$, et est donnée, pour tout ϑ dans l'intérieur de Θ , par

$$\begin{aligned} \mathbb{I}_{\mathcal{G}_{\vartheta_0}}(\vartheta) &:= \mathbb{E}_{\bar{\mathbb{P}}_{\vartheta, \vartheta_0}}[\nabla_{\vartheta} \bar{\ell}_{\vartheta_0}(\vartheta; X) \nabla_{\vartheta} \bar{\ell}_{\vartheta_0}(\vartheta; X)^\top] \\ &= \sum_{j=1}^d \int_0^T \int_{\mathbb{R}^d} \nabla_{\vartheta} (c^{-1/2}b_0)^j(\vartheta; t, x, \mu_t^{\vartheta_0}) \nabla_{\vartheta} (c^{-1/2}b_0)^j(\vartheta; t, x, \mu_t^{\vartheta_0})^\top \mu_t^{\vartheta, \vartheta_0}(dx) dt \\ &= \left(\int_0^T \int_{\mathbb{R}^d} \partial_{\vartheta_\ell} (c^{-1/2}b_0)(\vartheta; t, x, \mu_t^{\vartheta_0})^\top \partial_{\vartheta_{\ell'}} (c^{-1/2}b_0)(\vartheta; t, x, \mu_t^{\vartheta_0}) \mu_t^{\vartheta, \vartheta_0}(dx) dt \right)_{1 \leq \ell, \ell' \leq p}; \end{aligned}$$

- (iii) $\vartheta \mapsto \mathbb{I}_{\mathcal{G}_{\vartheta_0}}(\vartheta)$ est lipschitzienne.

Remarque 41. Dans l'expression de $\mathbb{I}_{\mathcal{G}_{\vartheta_0}}(\vartheta)$ la dérivée en ϑ est cette fois prise sur la fonction "découplée" $\vartheta \mapsto (c^{-1/2}b_0)(\vartheta; \cdot, \cdot, \mu_t^{\vartheta_0})$, et étant donné que $\mu^{\vartheta, \vartheta_0} \xrightarrow{\vartheta_0 \rightarrow \vartheta} \mu^\vartheta$ (voir le Théorème 30), nous avons, pour tout ϑ dans l'intérieur de Θ ,

$$\mathbb{I}_{\mathcal{G}_{\vartheta_0}}(\vartheta) \xrightarrow{\vartheta_0 \rightarrow \vartheta} \mathbb{I}(\vartheta).$$

Identifiabilité & Non-dégénérescence de l'Information de Fisher

Identifiabilité

Avoir une paramétrisation bien posée de notre modèle est une condition nécessaire à toute inférence statistique dans ce modèle pour obtenir des estimateurs au moins consistants. Par *paramétrisation bien posée*, nous entendons l'*identifiabilité* du modèle au sens suivant.

Définition 42. Le modèle $(\mathcal{E}^N)_{N \geq 1}$ (resp. $\mathcal{H}, \mathcal{G}_{\vartheta_0}(\vartheta_0 \in \Theta)$) est dit identifiable (globalement sur Θ) si $\vartheta \mapsto (\mathbb{P}_{\vartheta}^N)_{N \geq 1}$ (resp. $\vartheta \mapsto \bar{\mathbb{P}}_{\vartheta}, \vartheta \mapsto \bar{\mathbb{P}}_{\vartheta, \vartheta_0}(\vartheta_0 \in \Theta)$) est injective sur Θ .

Les hypothèses suivantes garantissent l'identifiabilité (globale) de ces différents modèles sur Θ .

Hypothèse 43. (i) Pour tout $\vartheta \in \Theta$, pour tout entier $N \geq 1$, pour $\mathbb{P}_{\vartheta}^N$ -presque tout ω , pour tout $\vartheta' \neq \vartheta$, les fonctions $t \mapsto \left(b(\vartheta; t, X_t^i(\omega), \mu_t^N(\omega))\right)_{1 \leq i \leq N}$ et $t \mapsto \left(b(\vartheta'; t, X_t^i(\omega), \mu_t^{(N)}(\omega))\right)_{1 \leq i \leq N}$ ne sont pas dt-presque partout égales ;

(ii) Pour tout $\vartheta \in \Theta$, pour $\bar{\mathbb{P}}_{\vartheta}$ -presque tout ω , pour tout $\vartheta' \neq \vartheta$, les fonctions $t \mapsto b(\vartheta; t, X_t(\omega), \mu_t^{\vartheta})$ et $t \mapsto b(\vartheta'; t, X_t(\omega), \mu_t^{\vartheta'})$ ne sont pas dt-presque partout égales ;

(iii) Soit $\vartheta_0 \in \Theta$. Pour $\vartheta \in \Theta$, pour $\bar{\mathbb{P}}_{\vartheta, \vartheta_0}$ -presque tout ω , pour tout $\vartheta' \neq \vartheta$, les fonctions $t \mapsto b(\vartheta; t, X_t(\omega), \mu_t^{\vartheta_0})$ et $t \mapsto b(\vartheta'; t, X_t(\omega), \mu_t^{\vartheta_0})$ ne sont pas dt-presque partout égales.

Nous avons donc que le point (i) (resp. (ii), (iii)) de l'Hypothèse 43 garantit l'identifiabilité globale sur Θ de $(\mathcal{E}^N)_{N \geq 1}$ (resp. $\mathcal{H}, \mathcal{G}_{\vartheta_0}$).

Cette Hypothèse 43, difficile à vérifier en pratique, est relativement standard dans la littérature de la Statistique des processus stochastiques et minimale (voir par exemple [79]). Vérifions le par exemple dans le cas du modèle $\mathcal{H}$. D'après le Théorème de Girsanov, pour deux paramètres $\vartheta, \vartheta' \in \Theta$, les lois $\bar{\mathbb{P}}_{\vartheta}$ et $\bar{\mathbb{P}}_{\vartheta'}$ sont absolument continues l'une par rapport à l'autre et

$$\begin{aligned} \log \frac{d\bar{\mathbb{P}}_{\vartheta}}{d\bar{\mathbb{P}}_{\vartheta'}}(X) &= \int_0^T \left((c^{-1}b_0)(\vartheta; s, X_s, \mu_s^{\vartheta}) - (c^{-1}b_0)(\vartheta'; s, X_s, \mu_s^{\vartheta'}) \right)^{\top} dX_s \\ &\quad - \frac{1}{2} \int_0^T \left(|(c^{-1/2}b_0)(\vartheta; s, X_s, \mu_s^{\vartheta})|^2 - |(c^{-1/2}b_0)(\vartheta'; s, X_s, \mu_s^{\vartheta'})|^2 \right) ds. \end{aligned}$$

Supposons alors que l'Hypothèse 43 (ii) soit en défaut pour un certain $\vartheta' \in \Theta, \vartheta' \neq \vartheta$. Cela impliquerait donc que, $\bar{\mathbb{P}}_{\vartheta}$ -presque sûrement, $\frac{d\bar{\mathbb{P}}_{\vartheta}}{d\bar{\mathbb{P}}_{\vartheta'}}(X) = 1$, c'est-à-dire $\bar{\mathbb{P}}_{\vartheta} = \bar{\mathbb{P}}_{\vartheta'}$, ce qui contredirait l'identifiabilité (globale) de $\mathcal{H}$ dans Θ .

Remarque 44. L'Hypothèse 43 est par exemple satisfaite dès que $\vartheta \mapsto ((t, x, \nu) \mapsto b_0(\vartheta; t, x, \nu))$ est injective.

Exemples 45. Ceci est le cas dans notre modèle de référence (4). En effet, supposons que pour $\vartheta, \vartheta' \in \Theta$, nous ayons $b_0(\vartheta; \cdot) = b_0(\vartheta'; \cdot)$, c'est-à-dire pour tous x, ν ,

$$\vartheta_1 x + \vartheta_2 + \vartheta_3 \int_{\mathbb{R}^d} (y - x) \nu(dy) = \vartheta'_1 x + \vartheta'_2 + \vartheta'_3 \int_{\mathbb{R}^d} (y - x) \nu(dy),$$

alors, en considérant successivement les cas $(x, \nu) = (0, \delta_0)$, $x = 0$ et $\int_{\mathbb{R}^d} y \nu(dy) \neq 0$, et enfin $x \neq 0$, nous obtenons $\vartheta = \vartheta'$.

Ceci est aussi à rapprocher du résultat d'injectivité de l'application $b \mapsto \mathcal{S}(b, c, \mu_0)$ du Théorème 25 (ix).

Nous traitons dans le paragraphe suivant de l'identifiabilité locale de $(\mathcal{E}^N)_{N \geq 1}$ et de ses liens avec les propriétés locales de l'information de Fisher asymptotique normalisée $\mathbb{I}$ de $(\mathcal{E}^N)_{N \geq 1}$.

**Identifiabilité locale & Non-dégénérescence faible asymptotique
de $(\mathcal{E}^N)_{N \geq 1}$**

Soit ϑ dans l'intérieur de Θ . Nous dirons que

Définition 46. Le modèle $(\mathcal{E}^N)_{N \geq 1}$ est localement identifiable en ϑ si $\vartheta' \mapsto (\mathbb{P}_{\vartheta'}^N)_{N \geq 1}$ est injective dans un voisinage de ϑ .

Nous dirons également, suivant en cela une dénomination classique, que

Définition 47. Le modèle $(\mathcal{E}^N)_{N \geq 1}$ est asymptotiquement non-dégénéré faiblement en ϑ si $\mathbb{I}(\vartheta)$ est inversible.

Nous disposons alors du résultat classique suivant qui lie ces deux notions.

Proposition 48 (Théorème 1 de [167], voir aussi [50], [181]). *Soit ϑ dans l'intérieur de Θ . Si $(\mathcal{E}^N)_{N \geq 1}$ est asymptotiquement non-dégénéré faiblement en ϑ , alors $(\mathcal{E}^N)_{N \geq 1}$ est localement identifiable en ϑ .*

Cette propriété usuelle de non-dégénérescence faible asymptotique de $(\mathcal{E}^N)_{N \geq 1}$ est cependant insuffisante, en toute généralité, dans notre cadre de travail : nous avons besoin de la définition plus forte suivante pour vérifier les hypothèses de la théorie de Ibragimov et Hasminski et obtenir ainsi des propriétés fines sur l'estimateur du maximum de vraisemblance.

Non-dégénérescence forte asymptotique de $(\mathcal{E}^N)_{N \geq 1}$

Définition 49. Nous dirons que le modèle $(\mathcal{E}^N)_{N \geq 1}$ est asymptotiquement non-dégénéré fortement sur Θ si

$$\inf_{[\vartheta, \vartheta'] \subset \Theta} \det \left(\sum_{j=1}^d \int_0^T \int_{\mathbb{R}^d} \nabla_{\vartheta} (c^{-1/2} b_0)^j ([\vartheta, \vartheta']; t, x, \mu_t^{\vartheta}) \nabla_{\vartheta} (c^{-1/2} b_0)^j ([\vartheta, \vartheta']; t, x, \mu_t^{\vartheta})^{\top} \mu_t^{\vartheta}(dx) dt \right) > 0, \quad (21)$$

où pour une fonction ϕ définie sur Θ , $\phi([\vartheta, \vartheta']) := \int_0^1 \phi(\vartheta + \lambda(\vartheta' - \vartheta)) d\lambda$.

Remarque 50. Si $(\mathcal{E}^N)_{N \geq 1}$ est asymptotiquement non-dégénéré fortement sur Θ , nous avons

$$\inf_{\vartheta \in \Theta} \det \mathbb{I}(\vartheta) > 0, \quad (22)$$

ce qui est équivalent au fait que le modèle $(\mathcal{E}^N)_{N \geq 1}$ est asymptotiquement non-dégénéré faiblement en tout point de Θ , puisque d'après la Proposition 38, $\vartheta \mapsto \mathbb{I}(\vartheta)$ est continue, et par hypothèse Θ est compact.

Nous donnons dans la sous-section suivantes des conditions suffisantes pour la non-dégénérescence asymptotique forte de $(\mathcal{E}^N)_{N \geq 1}$, d'abord dans un cadre général, puis dans des modèles particuliers.

Conditions suffisantes pour la non-dégénérescence asymptotique forte de $(\mathcal{E}^N)_{N \geq 1}$

Une difficulté spécifique à l'étude de la non-dégénérescence asymptotique du modèle $(\mathcal{E}^N)_{N \geq 1}$ réside dans la présence dans l'expression de $\mathbb{I}(\vartheta)$ de $(\mu_t^\vartheta)_{0 \leq t \leq T}$. Or, cette dernière n'est jamais explicite, hormis dans quelques cas particuliers précédemment cités, avec par exemple dans le modèle (4) une structure de dépendance en μ à travers ses moments qui permet aussi d'obtenir une formule explicite pour $\mathbb{I}(\vartheta)$. Dans ce cas particulier, nous trouvons directement, grâce à cette expression, des conditions suffisantes sur Θ et μ_0 pour que le modèle $(\mathcal{E}^N)_{N \geq 1}$ soit asymptotiquement non-dégénéré fortement (voir la Section 4 du Chapitre 1).

Dans le cas général, (21) peut cependant être testée de la manière suivante : la Définition 49 étant équivalente au fait de montrer que

$$\inf_{[\vartheta, \vartheta'] \subset \Theta} \min_{|z|=1} \sum_{j=1}^d \int_0^T \int_{\mathbb{R}^d} (\nabla_{\vartheta}(c^{-1/2}b)^j([\vartheta, \vartheta']; t, x, \mu_t^\vartheta)^\top z)^2 \mu_t^\vartheta(dx) dt > 0,$$

et étant donné que d'après les points (vi) et (viii) du Théorème 25, il existe une version continue et strictement positive de la densité de μ ,

$$((\vartheta, t, x) \in \Theta \times (0, T) \times \mathbb{R}^d \mapsto \mu_t^\vartheta(x)) \in \mathcal{C}(\Theta \times (0, T) \times \mathbb{R}^d; \mathbb{R}),$$

par un simple argument de continuité, il est donc suffisant de montrer qu'il n'existe pas de segment $[\vartheta, \vartheta'] \subset \Theta$ et de vecteur $z \in \mathbb{R}^p$, $|z| = 1$, tels que la fonction

$$x \mapsto \int_0^T \sum_{j=1}^d (\nabla_{\vartheta}(c^{-1/2}b_0)^j([\vartheta, \vartheta']; t, x, \mu_t^\vartheta)^\top z)^2 dt$$

tende vers 0 quand $|x| \rightarrow \infty$, ou même, dès que l'on a de la continuité en t quand $t \rightarrow 0^+$, qu'une des fonctions

$$x \mapsto \nabla_{\vartheta}(c^{-1/2}b_0)^j([\vartheta, \vartheta']; 0, x, \mu_0)^\top z, \quad j = 1, \dots, d$$

ne s'annule pas sur tout $\mathbb{R}^d$. Ce dernier critère, que l'on formalise dans la proposition suivante, a l'avantage d'éviter le terme μ_t^ϑ pour $t > 0$.

Proposition 51. *Supposons que les fonctions*

$$t \mapsto \nabla_{\vartheta}(c^{-1/2}b_0)^j([\vartheta, \vartheta']; t, x, \mu_t^\vartheta), \quad j = 1, \dots, d$$

sont toutes continues en $t = 0$ pour tout $[\vartheta, \vartheta'] \subset \Theta$ et presque pour tout $x \in \mathbb{R}^d$.

Si, pour tout $[\vartheta, \vartheta'] \subset \Theta$ et tout $z \in \mathbb{R}^p$, $|z| = 1$, une des fonctions

$$x \mapsto \nabla_{\vartheta}(c^{-1/2}b_0)^j([\vartheta, \vartheta']; 0, x, \mu_0)^\top z, \quad j = 1, \dots, d \tag{23}$$

ne s'annule pas sur tout $\mathbb{R}^d$, alors $(\mathcal{E}^N)_{N \geq 1}$ est asymptotiquement non-dégénéré fortement.

Dans la Section 4 du Chapitre 1, nous appliquons ce critère aux Exemples 12.

Maintenant que nous avons traité des questions d'identifiabilité et de non-dégénérescence de notre modèle, nous pouvons, dans la section suivante, passer à l'étude de sa régularité.

La propriété LAN

La propriété de *normalité asymptotique locale* (*local asymptotic normality* ou LAN) d'un modèle statistique caractérise sa régularité : elle exprime le fait que l'expérience ressemble localement à un *shift gaussien* dans une certaine échelle optimale donnée par l'information de Fisher. Plus précisément, nous considérons la définition suivante.

Définition 52. La suite d'expériences $(\mathcal{E}^N)_{N \geq 1}$ satisfait la propriété LAN en $\vartheta \in \Theta$, avec pour taux d'information $N\mathbb{I}(\vartheta)$, si pour $u \in \mathbb{R}^p$

$$\log \frac{d\mathbb{P}_{\vartheta + (N\mathbb{I}(\vartheta))^{-1/2}u}^N}{d\mathbb{P}_{\vartheta}^N}(\mathbb{X}^N) = u^\top \xi_{\vartheta}^N - \frac{1}{2}|u|^2 + r_N(\vartheta, u) \quad (24)$$

où ξ_{ϑ}^N converge en loi sous $\mathbb{P}_{\vartheta}^N$ vers une variable gaussienne standard dans $\mathbb{R}^p$, et $r_N(\vartheta, u) \rightarrow 0$ en $\mathbb{P}_{\vartheta}^N$ -probabilité.

Remarque 53. (24) n'a de sens que si $\vartheta + (N\mathbb{I}(\vartheta))^{-1/2}u$ est bien définie, ce qui est le cas si $(\mathcal{E}^N)_{N \geq 1}$ est non-dégéné faiblement asymptotiquement en tout point de Θ (voir la Définition 3.1), et si $\vartheta + (N\mathbb{I}(\vartheta))^{-1/2}u$ est dans Θ , ce qui est le cas pour ϑ dans l'intérieur de Θ et pour N assez grand.

Nous avons alors le théorème suivant.

Théorème 54. Si l'on suppose que $(\mathcal{E}^N)_{N \geq 1}$ est asymptotiquement identifiable et fortement non-dégéné sur Θ alors, pour tout ϑ dans l'intérieur de Θ , la suite d'expériences statistiques $(\mathcal{E}^N)_{N \geq 1}$ est localement asymptotiquement normale en ϑ avec pour taux d'information $N\mathbb{I}(\vartheta)$.

Le Théorème 54 est le résultat le plus fort que l'on peut obtenir sur la structure locale de $(\mathcal{E}^N)_{N \geq 1}$. Nous pouvons le reformuler de la manière suivante. Soit ϑ_0 un point dans l'intérieur de Θ . Alors pour tous $N \geq 1$ et $\vartheta \in \Theta$, en prenant $v_N := (N\mathbb{I}(\vartheta_0))^{1/2}(\vartheta - \vartheta_0)$, nous avons $\vartheta = \vartheta_0 + (N\mathbb{I}(\vartheta_0))^{-1/2}v_N$. Nous pouvons donc reparamétriser $(\mathcal{E}^N)_{N \geq 1}$ localement autour de ϑ_0 , sous la forme

$$\mathcal{E}_{\vartheta_0}^N := \{\mathbb{P}_{\vartheta_0 + (N\mathbb{I}(\vartheta_0))^{-1/2}v_N}^N ; v_N \in V_N\},$$

avec $V_N := \{v_N \in \mathbb{R}^p ; \vartheta_0 + (N\mathbb{I}(\vartheta_0))^{-1/2}v_N \in \Theta\}$. Alors, d'après la propriété LAN, la log-vraisemblance est de la forme

$$\log \frac{d\mathbb{P}_{\vartheta_0 + (N\mathbb{I}(\vartheta_0))^{-1/2}v_N}^N}{d\mathbb{P}_{\vartheta_0}^N}(\mathbb{X}^N) = v^\top \xi_{\vartheta_0}^N - \frac{1}{2}|v|^2 + r_N(\vartheta_0, v_N),$$

avec $\xi_{\vartheta_0}^N$ qui converge en loi sous $\mathbb{P}_{\vartheta_0}^N$ vers une variable gaussienne standard dans $\mathbb{R}^p$, et $r_N(\vartheta_0, v_N) \rightarrow 0$ en $\mathbb{P}_{\vartheta_0}^N$ -probabilité. Donc, $\mathbb{P}_{\vartheta_0}^N$ -asymptotiquement, la log-vraisemblance ressemble à la log-vraisemblance de l'expérience la plus simple possible, à savoir un *shift gaussien* sur $\mathbb{R}^p$,

$$\mathcal{T}^N := \{v_N + \xi ; v_N \in V_N\}$$

où ξ suit une $\mathcal{N}_p(0, \text{Id}_{\mathbb{R}^p})$, puisque pour tout $z \in \mathbb{R}^p$,

$$\log \frac{(2\pi)^{-p/2} \exp\left(-\frac{|z-v_N|^2}{2}\right)}{(2\pi)^{-p/2} \exp\left(-\frac{|z|^2}{2}\right)} = v_N^\top z - \frac{1}{2}|v_N|^2.$$

Bornes inférieures minimax asymptotiques exactes

Dans toute la suite, nous appellerons *fonction de perte polynomiale sur $\mathbb{R}^p$* toute fonction $w : \mathbb{R}^p \rightarrow [0, \infty)$ symétrique, admettant un majorant polynomial, et telle que, pour tout $c > 0$, l'ensemble $\{w < c\}$ est convexe.

Pour ϑ dans l'intérieur de Θ et pour une fonction de perte polynomiale w sur $\mathbb{R}^p$, nous définissons le risque associé à w pour un estimateur $\hat{\vartheta}_N$ dans $\mathcal{E}^N$:

$$\mathcal{R}_w^N(\hat{\vartheta}_N; \vartheta) := \mathbb{E}_{\mathbb{P}_{\vartheta}^N} [w((N\mathbb{I}(\vartheta))^{1/2}(\hat{\vartheta}_N - \vartheta))] ,$$

et pour tout sous-ensemble ouvert non vide Θ_0 de Θ ,

$$\mathcal{R}_w^N(\hat{\vartheta}_N; \Theta_0) := \sup_{\vartheta \in \Theta_0} \mathcal{R}_w^N(\hat{\vartheta}_N; \vartheta) .$$

La première conséquence simple de la propriété LAN est donnée en termes de *bornes inférieures minimax asymptotiques exactes* par le résultat suivant.

Corollaire 55. *Soit ϑ dans l'intérieur de Θ et w une fonction de perte polynomiale sur $\mathbb{R}^p$. Si l'on suppose que $(\mathcal{E}^N)_{N \geq 1}$ est asymptotiquement identifiable et asymptotiquement fortement non-dégénéré sur un voisinage ouvert de θ alors, pour tout estimateur $\hat{\vartheta}_N$ dans $\mathcal{E}^N$, pour tout $\delta > 0$ assez petit, nous avons*

$$\liminf_{N \rightarrow \infty} \sup_{\vartheta' \in \overline{B}(\vartheta; \delta)} \mathbb{E}_{\mathbb{P}_{\vartheta'}^N} [w((N\mathbb{I}(\vartheta))^{1/2}(\hat{\vartheta}_N - \vartheta'))] \geq (2\pi)^{-p/2} \int_{\mathbb{R}^p} w(x) \exp(-\tfrac{1}{2}|x|^2) dx .$$

Le Corollaire 55 est une simple application du Théorème de convolution de Hájek (voir par exemple [14] et le Théorème II.9.1 de [98]), étant donnée la propriété LAN du Théorème 54 (cf. Théorème II.12.1 [98]),

Remarque 56. *Suivant en cela la Remarque II.12.1 de [98], le Corollaire 55 nous fournit une borne inférieure asymptotique minimax locale précise :*

$$\liminf_{N \rightarrow \infty} \inf_{\hat{\vartheta}_N} \sup_{\vartheta' \in \overline{B}(\vartheta; \delta)} \mathbb{E}_{\mathbb{P}_{\vartheta'}^N} [w((N\mathbb{I}(\vartheta))^{1/2}(\hat{\vartheta}_N - \vartheta'))] \geq (2\pi)^{-p/2} \int_{\mathbb{R}^p} w(x) \exp(-\tfrac{1}{2}|x|^2) dx .$$

La propriété LAN a également de nombreuses conséquences en termes de propriétés fortes de l'estimateur du maximum de vraisemblance, c'est l'objet de la sous-section suivante. En particulier, l'estimateur du maximum de vraisemblance atteint la borne minimax du Corollaire précédent.

Estimateur du maximum de vraisemblance

Exemples 57. Dans le cadre du modèle (4), pour $d = 1$, nous obtenons que l'estimateur du maximum de vraisemblance est $\hat{\vartheta}^{N, \text{mle}} = (\hat{\vartheta}_1^{N, \text{mle}}, \hat{\vartheta}_2^{N, \text{mle}}, \hat{\vartheta}_3^{N, \text{mle}})$ défini par

$$\hat{\vartheta}_1^{N, \text{mle}} = \frac{\frac{T(\sigma-1)}{N\sigma} \sum_{i=1}^N \int_0^T X_t^i dX_t^i - \frac{1}{\sigma N} \sum_{i=1}^N \int_0^T X_t^i dt \cdot \frac{1}{N} \sum_{i=1}^N (X_T^i - X_0^i) + \frac{T}{\sigma N^2} \sum_{i,j=1}^N \int_0^T X_t^j dX_t^i}{\frac{T(N+1)}{N^2} \sum_{i=1}^N \int_0^T (X_t^i)^2 dt - \left(\frac{1}{N} \sum_{i=1}^N \int_0^T X_t^i dt \right)^2} ,$$

$$\begin{aligned}\hat{\vartheta}_2^{N, \text{mle}} &= \frac{1}{\sigma NT} \sum_{i=1}^N (X_T^i - X_0^i) - \hat{\vartheta}_1^{N, \text{mle}} \cdot \frac{1}{NT} \sum_{i=1}^N \int_0^T X_t^i dt, \\ \hat{\vartheta}_3^{N, \text{mle}} &= \frac{\frac{1}{N^2} \sum_{i,j=1}^N \int_0^T X_t^j dX_t^i - \frac{1}{N} \sum_{i=1}^N \int_0^T X_t^i dX_t^i}{\sigma \left(\frac{1}{N} \sum_{i=1}^N \int_0^T (X_t^i)^2 dt - \frac{1}{N^2} \sum_{i,j=1}^N \int_0^T X_t^i X_t^j dt \right)} - \hat{\vartheta}_1^{N, \text{mle}},\end{aligned}$$

et nous pouvons directement vérifier sa consistance et sa normalité asymptotique grâce à cette forme explicite et aux propriétés de limites de champ moyen.

Dans le cadre général, nous avons prouvé certaines propriétés de l'estimateur du maximum de vraisemblance, qui reposent sur une version uniforme de la propriété LAN du Théorème 54 (voir le Chapitre 1) : les résultats obtenus sont plus fins que les usuels développements asymptotiques faibles.

Théorème 58. *Supposons que le modèle $(\mathcal{E}^N)_{N \geq 1}$ est asymptotiquement identifiable. Alors, pour N assez grand, l'estimateur du maximum de vraisemblance dans $\mathcal{E}^N$, i.e. la solution $\hat{\vartheta}_N^{\text{mle}}$ de*

$$\mathcal{L}^N(\hat{\vartheta}_N^{\text{mle}}; \mathbb{X}^N) = \sup_{\vartheta \in \Theta} \mathcal{L}^N(\vartheta; \mathbb{X}^N) \quad (25)$$

est bien définie. De plus, si $(\mathcal{E}^N)_{N \geq 1}$ est asymptotiquement non-dégénéré fortement, nous obtenons les bornes supérieures asymptotiques suivantes.

(i) *Pour tout $\vartheta \in \Theta$, $\hat{\vartheta}_N^{\text{mle}}$ est consistant et asymptotiquement normal :*

$$\sqrt{N}(\hat{\vartheta}_N^{\text{mle}} - \vartheta) \rightarrow \mathcal{N}_p(0, \mathbb{I}(\vartheta)^{-1})$$

en $\mathbb{P}_{\vartheta}^N$ -loi quand $N \rightarrow \infty$;

(ii) *Pour toute fonction de perte polynomiale w et tout ϑ dans l'intérieur de Θ , nous avons l'optimalité minimax locale asymptotique exacte :*

$$\limsup_{N \rightarrow \infty} \sup_{\vartheta' \in \overline{B}(\vartheta; \delta)} \mathbb{E}_{\mathbb{P}_{\vartheta'}^N} [w((N\mathbb{I}(\vartheta))^{1/2}(\hat{\vartheta}_N^{\text{mle}} - \vartheta'))] \xrightarrow{\delta \rightarrow 0} (2\pi)^{-p/2} \int_{\mathbb{R}^p} w(x) \exp(-\frac{1}{2}|x|^2) dx ;$$

(iii) *Pour toute fonction de perte polynomiale w et tout sous-ensemble ouvert non vide Θ_0 de Θ , nous avons l'optimalité minimax asymptotique globale : quand $N \rightarrow \infty$*

$$\mathcal{R}_w^N(\hat{\vartheta}_N^{\text{mle}}; \Theta_0) = \inf_{\vartheta_N} \mathcal{R}_w^N(\vartheta_N; \Theta_0)(1 + o(1)) .$$

Nous retrouvons ainsi les propriétés asymptotiques classiques d'un estimateur du maximum de vraisemblance.

Observations discrètes

Enfin, nous avons complété l'article [64] par une étude théorique, dans le cadre paramétrique, lorsque nous disposons de M observations discrètes de chacune des N particules à la place des observations continues, et nous avons obtenu les résultats qui suivent dans la *double asymptotique* $N, M \rightarrow \infty$ (les preuves de ces résultats sont données dans l'Annexe 4). Cette étude s'appuie sur la nomenclature et les résultats du Chapitre 3 de [54], ainsi que sur des techniques classiques de discrétisation et d'estimation en *haute fréquence* (voir *e.g.* les livres de Jacod et de ses co-auteurs [2, 101, 102]).

Soit ϑ_0 dans l'intérieur de Θ . Pour $M \in \mathbb{N}^*$, nous supposons donc que nous disposons maintenant des observations discrètes $\mathbb{X}^{N,M} := (X_{t_j^M}^i)_{1 \leq i \leq N, 0 \leq j \leq M} \in (\mathbb{R}^d)^{N \times (M+1)}$ de $\mathbb{P}_{\vartheta_0}^N$, aux instants $t_j^M = j\Delta_M, j = 0, \dots, M$, séparés par un pas constant $\Delta_M := \frac{T}{M}$.

Nous définissons tout d'abord la fonction de *log-vraisemblance approchée* (ou *approximate log-likelihood*) suivante, basée sur le *schéma d'approximation d'Euler-Maruyama*

$$\begin{aligned} \tilde{\ell}^{N,M}(\vartheta; \mathbb{X}^{N,M}) &= \sum_{i=1}^N \sum_{j=1}^M (c^{-1}b_0)(\vartheta; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N)^\top (X_{t_j^M}^i - X_{t_{j-1}^M}^i) \\ &\quad - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^M |(c^{-1/2}b_0)(\vartheta; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N)|^2 \Delta_M, \end{aligned}$$

et ensuite l'*estimateur du maximum de vraisemblance approché* (ou *approximate maximum likelihood estimator*, AMLE) associé $\hat{\vartheta}_{N,M}^{\text{amle}}$ par

$$\tilde{\ell}^{N,M}(\hat{\vartheta}_{N,M}^{\text{amle}}; \mathbb{X}^{N,M}) = \sup_{\vartheta \in \Theta} \tilde{\ell}^{N,M}(\vartheta; \mathbb{X}^{N,M}).$$

Notons $\tilde{U}^{N,M}(\vartheta; y) := -2\tilde{\ell}^{N,M}(\vartheta; y)$, $y = (y_j^i)_{1 \leq i \leq N, 0 \leq j \leq M}$. Nous utiliserons le *processus de contraste*

$$U^{N,M}(\vartheta; \mathbb{X}^{N,M}) := \frac{1}{N} (\tilde{U}^{N,M}(\vartheta; \mathbb{X}^{N,M}) - \tilde{U}^{N,M}(\vartheta_0; \mathbb{X}^{N,M})),$$

ce qui signifie que $U^{N,M}(\vartheta; \mathbb{X}^{N,M})$ est égal à

$$\begin{aligned} & - \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M \left((c^{-1}b_0)(\vartheta; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N) - (c^{-1}b_0)(\vartheta_0; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N) \right)^\top (X_{t_j^M}^i - X_{t_{j-1}^M}^i) \\ & + \frac{1}{2} \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M \left(|(c^{-1/2}b_0)(\vartheta; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N)|^2 - |(c^{-1/2}b_0)(\vartheta_0; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N)|^2 \right) \Delta_M, \end{aligned}$$

et ce qui entraîne bien sûr que $\hat{\vartheta}_{N,M}^{\text{amle}} \in \underset{\vartheta \in \Theta}{\operatorname{argmin}} U^{N,M}(\vartheta; \mathbb{X}^{N,M})$.

Nous obtenons alors la *consistance* de l'estimateur du maximum de vraisemblance approché sous une hypothèse de régularité plus forte en t de b_0 et c :

Théorème 59. *Supposons que*

(i) l'expérience statistique $(\mathcal{E}^N)_{N \geq 1}$ est asymptotiquement non-dégénérée faiblement,
 (ii) pour un certain $\gamma \in (0, 1]$, b_0 et c sont γ -Hölder en t uniformément par rapport à (ϑ, x, ν) ,
 alors $\hat{\vartheta}_{N,M}^{\text{amle}}$ est consistant :

$$\hat{\vartheta}_{N,M}^{\text{amle}} \rightarrow \vartheta_0$$

en $\mathbb{P}_{\vartheta_0}^N$ -probabilité, quand $N, M \rightarrow \infty$.

Preuve. Voir Annexe 4, Section 4.1. □

Enfin, sous des hypothèses de régularité un peu plus restrictives sur b_0 et c , ainsi qu'une condition du type $N \ll M$ (caractéristique de l'estimation statistique en haute fréquence), nous obtenons la *normalité asymptotique* de l'estimateur du maximum de vraisemblance approché.

Théorème 60. *Sous les hypothèses du Théorème 59, si nous supposons de plus que*

(i) pour un certain $\gamma' \in (0, 1]$, pour tout $1 \leq k \leq p$,

$$\partial_{\vartheta_k}(c^{-1}b_0) \text{ est } \gamma'\text{-Hölder en } t \text{ uniformément par rapport à } (\vartheta, x, \nu),$$

(ii) b_0 est deux fois continûment différentiable en ϑ uniformément par rapport à (t, x, ν) ,

$$(iii) \quad NM^{-2\gamma} \xrightarrow{N, M \rightarrow \infty} 0,$$

alors $\hat{\vartheta}_{N,M}^{\text{amle}}$ est asymptotiquement normal :

$$\sqrt{N}(\hat{\vartheta}_{N,M}^{\text{amle}} - \vartheta_0) \rightarrow \mathcal{N}_p(0, \mathbb{I}(\vartheta_0)^{-1})$$

en loi sous $\mathbb{P}_{\vartheta_0}^N$, quand $N, M \rightarrow \infty$.

Preuve. Voir Annexe 4, Section 4.2. □

Nous présentons maintenant dans les sections suivantes les résultats de Statistique non-paramétrique obtenus dans l'article [65]. Dans ces sections, nous nous appuyerons sur les outils et les résultats présentés dans les livres de Tsybakov [182] et Comte [47], et nous nous placerons sous les hypothèses

$$(b, c, \mu_0) \in \mathfrak{T}(\mathbf{b}) := \mathfrak{B}_{T,d,\mathfrak{h},\mathfrak{n},\mathfrak{c}_2,\mathfrak{c}_3} \times \mathfrak{C}_{T,d,\mathfrak{s}_+,\mathfrak{s}_-,\mathfrak{c}_1} \times \mathfrak{M}_{d,\mathfrak{m}}^*, \quad \mathbf{b} \in \mathfrak{E}.$$

Commençons par l'estimation de la solution $\mu^b = \mathcal{S}(b, c, \mu_0)$ de (2).

3.2 Estimation non-paramétrique de μ^b

Rappel des hypothèses

- (i) $\mu_0 \in \bigcap_{\ell \geq 0} \mathcal{P}_\ell(\mathbb{R}^d)$, avec $\mathfrak{m}_{2\ell}(\mu_0) \leq \mathfrak{m}_{2\ell}$ et la suite $((\frac{\mathfrak{m}_{2\ell}}{\ell!})^{1/\ell})_{\ell \geq 1}$ bornée ;
- (ii) $\sigma_-^2 |y|^2 \leq y^\top c(t, x) y \leq \sigma_+^2 |y|^2$, $0 < \mathfrak{s}_- \leq \sigma_- \leq \sigma_+ \leq \mathfrak{s}_+$;
- (iii) $|\sigma(t, x') - \sigma(t, x)| \leq |\sigma|_{\text{Lip}} |x' - x|$, $|\sigma|_{\text{Lip}} \leq \mathfrak{c}_1$;
- (iv) $\sup_{t \in [0, T]} |b(t, 0, \delta_0)| \leq \mathfrak{c}_2$;
- (v) $|b(t, x', \nu') - b(t, x, \nu)| \leq |b|_{\text{Lip}} (|x' - x| + \mathcal{W}_1(\nu', \nu))$;
- (vi) b est au plus k_b^* fois linéairement différentiable en ν et de décomposition à au plus m_b termes, avec $k_b^* \leq \mathfrak{h}$, $m_b \leq \mathfrak{n}$;
- (vii) $\max(|b|_{\text{Lip}}, |\delta_\nu^1 b|_{\text{Lip}, \infty}, \dots, |\delta_\nu^{k_b^*} b|_{\text{Lip}, \infty}, |\delta_\nu^{k_b^*} b|_{\text{Lip}}) \leq \mathfrak{c}_3$.

Objectif principal & Quantification de la qualité des estimateurs

Rappelons que, sous ces hypothèses, d'après le Théorème 25, $\mu^b = \mathcal{S}(b, c, \mu_0)$ admet une densité continue et strictement positive

$$\mu^b : (t, x) \in (0, T) \times \mathbb{R}^d \mapsto \mu_t^b(x) \in \mathbb{R}_+^* .$$

Nous voulons, pour tout $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$, estimer $\mu_{t_0}^b(x_0)$ à partir des observations continues $\mathbb{X}^N = (\mathbb{X}_t^N)_{t \in [0, T]}$. Pour un estimateur quelconque $\tilde{\mu}^N$ de μ^b , construit à partir de $\mathbb{X}^N$, *i.e.* fonction mesurable de $\mathbb{X}^N$, nous définissons le *risque quadratique* de $\tilde{\mu}^N$ en (t_0, x_0) par

$$\mathcal{R}_{(t_0, x_0)}(\tilde{\mu}^N, \mu^b) := \mathbb{E}_{\mathbb{P}_b^N} [(\tilde{\mu}_{t_0}^N(x_0) - \mu_{t_0}^b(x_0))^2] .$$

Estimateurs à noyaux

Parmi la classe des estimateurs $\tilde{\mu}_{t_0}^N(x_0)$ de $\mu_{t_0}^b(x_0)$ construits à partir de $\mathbb{X}_{t_0}^N$, nous nous intéressons plus particulièrement à la famille des *estimateurs à noyaux*

$$\{\hat{\mu}_h^N(t_0, x_0) ; h \in \mathcal{H}_1\}, \mathcal{H}_1 \subset (0, 1) ,$$

définis de la manière suivante :

$$\hat{\mu}_h^N(t_0, x_0) := \frac{1}{N} \sum_{i=1}^N K_h(x_0 - X_{t_0}^i) = \int_{\mathbb{R}^d} K_h(x_0 - x) \mu_{t_0}^N(dx) ,$$

où $K_h(x) := h^{-d} K(\frac{x_0 - x}{h})$, K étant un *noyau*, c'est-à-dire une fonction intégrable de $\mathbb{R}^d$ dans $\mathbb{R}$ vérifiant la condition $\int_{\mathbb{R}^d} K(y) dy = 1$, et $\mathcal{H}_1$ est l'ensemble (pour le moment quelconque) des *fenêtres* possibles.

Remarque 61. Par définition, pour toute fenêtre $h > 0$, K_h est lui aussi un noyau.

Exemples 62. Des exemples de telles fonctions K est donné par les densités de probabilités sur $\mathbb{R}^d$, par exemple par le noyau gaussien $K(y) = (2\pi)^{-d/2} e^{-|y|^2/2}$, ou le noyau de Silverman $K(y) = \frac{1}{2} e^{-|y|/2} \sin(\frac{|y|}{\sqrt{2}} + \frac{\pi}{4})$ (voir *e.g.* [182]).

Cette méthode de l'*estimation par noyau d'une densité* $\mu_{t_0}^b$ à partir d'un système de particules $(X_{t_0}^i)_{1 \leq i \leq N}$, dite *méthode de Parzen-Rosenblatt*, généralise la méthode d'estimation par *histogramme* : l'estimation est maintenant continue grâce au *paramètre de lissage* ou *fenêtre* $h > 0$. Rosenblatt introduisit cette méthode en dimension $d = 1$ dans l'article [166], pour estimer au point x_0 la densité $\mu_{t_0}^b$ d'un échantillon indépendant et identiquement distribué $(X_{t_0}^i)_{1 \leq i \leq N}$ par $\frac{1}{N} \sum_{i=1}^N \frac{1}{2h} \mathbb{1}_{x_0 - h < X_{t_0}^i \leq x_0 + h}$, *i.e.* en prenant pour noyau le *noyau uniforme*, $K(y) = \frac{1}{2} \mathbb{1}_{[-1, 1]}(y)$, puis Parzen généralisa cette méthode au cas d'un noyau K quelconque dans l'article [163].

Ici cependant, nos particules, bien que de même loi $\mu_{t_0}^{1, N}$, ne sont en toute généralité pas indépendantes, puisque leurs dynamiques interagissent via la mesure empirique du système entier μ^N dans la fonction de drift b . Pour que les N trajectoires X^i , $1 \leq i \leq N$, soient indépendantes, il faudrait que b ne dépende pas de la mesure, autrement dit que $b(t, x, \nu) = b(t, x)$. Dans ce cas particulier, $\mu^{1, N} = \mu^b$, et nous retombons sur un problème classique en Statistique non-paramétrique (voir *e.g.* [182]). Mais lorsque b dépend effectivement de la mesure (ce que nous supposons par

défaut dans cette thèse), les particules dépendent les unes des autres, et $\mu_{t_0}^{1,N}$ n'est plus qu'une approximation de $\mu_{t_0}^b$.

Comme pour le choix du nombre de classes dans un histogramme, le choix de la fenêtre h est essentiel pour la qualité de l'approximation. Un h petit (correspondant à un grand nombre de classes dans un histogramme) amène à un *sous-lissage* : l'estimation va osciller trop fortement autour de la valeur réelle. Au contraire, un h grand (correspondant à un petit nombre de classes dans un histogramme) amène à un *sur-lissage* : l'estimation n'est alors pas assez précise et ne permet pas de bien capter les détails. Il faut donc trouver un équilibre entre ces deux écueils et chercher le compromis optimal, *i.e.* ici une fenêtre $h \in \mathcal{H}_1$ optimale en un sens à déterminer. C'est l'objet de la section suivante.

Décomposition du risque quadratique ponctuel

Nous allons décomposer notre risque quadratique ponctuel en un terme de *biais* et un terme de *variance*. Par définition, nous avons pour $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$, $h \in \mathcal{H}_1$,

$$\mathcal{R}_{(t_0, x_0)}(\hat{\mu}_h^N, \mu^b) = \mathbb{E}_{\mathbb{P}_b^N} [(\hat{\mu}_h^N(t_0, x_0) - \mu_{t_0}^b(x_0))^2] .$$

Nous avons alors (au moins) deux possibilités pour décomposer $\mathcal{R}_{(t_0, x_0)}(\hat{\mu}_h^N, \mu^b)$: utiliser

$$\begin{aligned} \mathcal{R}_{(t_0, x_0)}(\hat{\mu}_h^N, \mu^b) &= \mathbb{E}_{\mathbb{P}_b^N} [(\hat{\mu}_h^N(t_0, x_0) - \mathbb{E}_{\mathbb{P}_b^N} [\hat{\mu}_h^N(t_0, x_0)])^2] + (\mathbb{E}_{\mathbb{P}_b^N} [\hat{\mu}_h^N(t_0, x_0)] - \mu_{t_0}^b(x_0))^2 \\ &= \text{Var}_{\mathbb{P}_b^N} [\hat{\mu}_h^N(t_0, x_0)] + \left(\int_{\mathbb{R}^d} K_h(x_0 - x) \mu_{t_0}^{1,N}(dx) - \mu_{t_0}^b(x_0) \right)^2 , \end{aligned}$$

ce qui, étant donné la présence de $\mu^{1,N}$ et l'interdépendance des particules dans le cas général, ne nous simplifie pas la tâche, hormis dans le cas $b(t, x, \nu) = b(t, x)$, où les particules sont alors *i.i.d.* de loi $\mu^{1,N} = \mu^b$ et où nous retrouvons la décomposition biais-variance usuelle

$$\mathcal{R}_{(t_0, x_0)}(\hat{\mu}_h^N, \mu^b) = \frac{1}{N} \text{Var}_{\mathbb{P}_b^N} [K_h(x_0 - X_{t_0}^1)] + \left(\int_{\mathbb{R}^d} K_h(x_0 - x) \mu_{t_0}^b(dx) - \mu_{t_0}^b(x_0) \right)^2 ;$$

ou bien plutôt utiliser (c'est ce que nous ferons) la décomposition approximative suivante

$$\mathcal{R}_{(t_0, x_0)}(\hat{\mu}_h^N, \mu^b) \leq 2\mathbb{E}_{\mathbb{P}_b^N} \left[\left(\int_{\mathbb{R}^d} K_h(x_0 - x) (\mu_{t_0}^N - \mu_{t_0}^b)(dx) \right)^2 \right] + 2 \left(\int_{\mathbb{R}^d} K_h(x_0 - x) \mu_{t_0}^b(dx) - \mu_{t_0}^b(x_0) \right)^2 .$$

Nous appellerons terme de *biais* (*au carré*) et terme de *variance* respectivement

$$\left(\int_{\mathbb{R}^d} K_h(x_0 - x) \mu_{t_0}^b(dx) - \mu_{t_0}^b(x_0) \right)^2 \text{ et } \mathbb{E}_{\mathbb{P}_b^N} \left[\left(\int_{\mathbb{R}^d} K_h(x_0 - x) (\mu_{t_0}^N - \mu_{t_0}^b)(dx) \right)^2 \right] .$$

Nous allons traiter ces deux termes séparément, en commençant par le terme de variance.

Remarque 63. Nous nous conformons à la nomenclature statistique habituelle en conservant ces termes de biais et variance mais, dans notre cadre de travail, ces termes ne recouvrent pas la réalité qu'ils sont censés définir. En effet, d'une part, $\int_{\mathbb{R}^d} K_h(x_0 - x) \mu_{t_0}^b(dx) - \mu_{t_0}^b(x_0)$ n'est (en toute généralité) pas le biais de notre estimateur $\hat{\mu}_h^N(t_0, x_0)$ au sens de $\mathbb{E}_{\mathbb{P}_b^N} [\hat{\mu}_h^N(t_0, x_0)] - \mu_{t_0}^b(x_0)$ puisqu'étant donné que $\mu^{1,N}$ n'est qu'une approximation de μ^b ,

$$\mathbb{E}_{\mathbb{P}_b^N} [\hat{\mu}_h^N(t_0, x_0)] = \int_{\mathbb{R}^d} K_h(x_0 - x) \mu_{t_0}^{1,N}(dx)$$

n'a pas de raison d'être égal à $\int_{\mathbb{R}^d} K_h(x_0 - x) \mu_{t_0}^b(dx)$. Et d'autre part, pour la même raison, $\mathbb{E}_{\mathbb{P}_b^N} \left[\left(\int_{\mathbb{R}^d} K_h(x_0 - x) (\mu_{t_0}^N - \mu_{t_0}^b)(dx) \right)^2 \right]$ n'est (en toute généralité) pas la variance de notre estimateur $\hat{\mu}_h^N(t_0, x_0)$ au sens de $\mathbb{E}_{\mathbb{P}_b^N} \left[\left(\hat{\mu}_h^N(t_0, x_0) - \mathbb{E}_{\mathbb{P}_b^N} [\hat{\mu}_h^N(t_0, x_0)] \right)^2 \right]$.

Dans toute la suite, nous ferons l'hypothèse suivante sur le noyau K que nous utilisons :

Hypothèse 64. K est borné et à support compact.

Exemples 65. Un exemple de tel noyau K est donné par le noyau uniforme $K(y) = \frac{1}{2} \mathbb{1}_{[-1,1]}(y)$.

Terme de variance

Pour traiter ce terme, nous avons utilisé tout d'abord le fait que

$$\mathbb{E}_{\mathbb{P}_b^N} \left[\left(\int_{\mathbb{R}^d} K_h(x_0 - x) (\mu_{t_0}^N - \mu_{t_0}^b)(dx) \right)^2 \right] = \int_0^\infty \mathbb{P}_b^N \left(\left| \int_{\mathbb{R}^d} K_h(x_0 - x) (\mu_{t_0}^N - \mu_{t_0}^b)(dx) \right| \geq z^{1/2} \right) dz ,$$

ce qui nous a amené à chercher comment majorer de manière fine les probabilités de déviations

$$\mathbb{P}_b^N \left(\int_{\mathbb{R}^d} \phi(x) (\mu_{t_0}^N - \mu_{t_0}^b)(dx) \geq z^{1/2} \right) ,$$

pour $\phi(x) = \pm K_h(x_0 - x)$. Nous avons dans ce but établi le résultat d'*inégalité de concentration de type Bernstein* suivant (nous y reviendrons en détails dans la Section 3.6) :

Théorème 66. Il existe $\kappa_1 = \kappa_1(\mathbf{b}), \kappa_2 = \kappa_2(\mathbf{b}) > 0$ tels que pour toute fonction bornée $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$,

$$\mathbb{P}_b^N \left(\int_{\mathbb{R}^d} \phi(y) (\mu_{t_0}^N - \mu_{t_0}^b)(dy) \geq z \right) \leq \kappa_1 \exp \left(- \kappa_2 \frac{Nz^2}{|\phi|_{L^2(\mu_{t_0}^b)}^2 + |\phi|_\infty z} \right) , \text{ pour tout } z \geq 0 .$$

Comme, par hypothèse, K est bornée, nous avons $|\phi|_\infty = h^{-d} |K|_\infty$, et comme K est également à support compact $\text{supp}(K)$ et $h \in \mathcal{H}_1 \subset (0, 1)$,

$$|\phi|_{L^2(\mu_{t_0}^b)}^2 = h^{-2d} \int_{(x_0 - h \text{supp}(K))} \left| K \left(\frac{x_0 - x}{h} \right) \right|^2 \mu_{t_0}^b(x) dx \leq h^{-2d} \int_{(x_0 - \text{supp}(K))} \left| K \left(\frac{x_0 - x}{h} \right) \right|^2 \mu_{t_0}^b(x) dx ,$$

et comme nous savons, d'après le Théorème 25, que μ^b est localement bornée autour de (t_0, x_0) par $\kappa_3 = \kappa_3(\mu^b, t_0, x_0) > 0$,

$$|\phi|_{L^2(\mu_{t_0}^b)}^2 \leq h^{-2d} \int_{(x_0 - \text{supp}(K))} \left| K \left(\frac{x_0 - x}{h} \right) \right|^2 \kappa_3 dx \leq \kappa_3 h^{-d} |K|_2^2 .$$

Nous obtenons donc que

$$\mathbb{E}_{\mathbb{P}_b^N} \left[\left(\int_{\mathbb{R}^d} K_h(x_0 - x) (\mu_{t_0}^N - \mu_{t_0}^b)(dx) \right)^2 \right] \leq \int_0^\infty 2\kappa_1 \exp \left(- \kappa_2 \frac{N h^d z}{\kappa_3 |K|_2^2 + |K|_\infty z^{1/2}} \right) dz$$

$$\begin{aligned} &\lesssim \kappa_1 \max \left(\left(\frac{\kappa_2 N h^d}{\kappa_3 |K|_2^2} \right)^{-1}, \left(\frac{\kappa_2 N h^d}{|K|_\infty} \right)^{-2} \right) \\ &\lesssim (N h^d)^{-1} + (N h^d)^{-2} . \end{aligned}$$

Nous nous laissons à présent la possibilité de choisir un ensemble de fenêtres dépendant de N et nous le notons $\mathcal{H}_1^N$. Nous ferons l'hypothèse supplémentaire suivante :

Hypothèse 67. $\mathcal{H}_1^N \subset [(N^{-1}(\log N)^2)^{1/d}; 1]$ et $\text{Card}(\mathcal{H}_1^N) \lesssim N$.

Remarque 68. Cela implique donc que, pour $h \in \mathcal{H}_1^N$, $(N h^d)^{-1} \lesssim 1$.

Proposition 69. *Sous les Hypothèses 64 et 67,*

$$\mathbb{E}_{\mathbb{P}_b^N} \left[\left(\int_{\mathbb{R}^d} K_h(x_0 - x) (\mu_{t_0}^N - \mu_{t_0}^b)(dx) \right)^2 \right] \lesssim (N h^d)^{-1} . \quad (26)$$

Passons maintenant au terme de biais (au carré) $\left(\int_{\mathbb{R}^d} K_h(x_0 - x) \mu_{t_0}^b(dx) - \mu_{t_0}^b(x_0) \right)^2$.

Terme de biais

Ce terme se traite via la régularité de $\mu_{t_0}^b$ localement en x_0 . Dans cette section, comme dans l'article [65], nous ne considérons comme espaces quantifiant cette régularité que des *espaces de type Hölder* définis de la manière suivante.

Définition 70. *Soit un réel $\beta > 0$, nous noterons $\lfloor \beta \rfloor$ le plus petit entier strictement inférieur à β . Nous dirons que $\mu_{t_0}^b$ est localement β -Hölder en x_0 , et nous noterons $\mu_{t_0}^b \in \mathcal{H}^\beta(x_0)$, s'il existe un voisinage (relativement compact) $\mathcal{U} = \mathcal{U}_{x_0}$ de x_0 dans $\mathbb{R}^d$ tel que*

(i) $\mu_{t_0}^b$ est $\lfloor \beta \rfloor$ fois dérivable sur $\mathcal{U}$,

(ii) *il existe $C > 0$, tel que pour tout $s = (s_1, \dots, s_d) \in \mathbb{N}^d$ vérifiant $|s| := \sum_{i=1}^d s_i \leq \lfloor \beta \rfloor$, pour tout $x, x' \in \mathcal{U}$,*

$$|D_x^s \mu_{t_0}^b(x') - D_x^s \mu_{t_0}^b(x)| \leq C |x' - x|^{\beta - \lfloor \beta \rfloor}$$

où

$$D_x^s \mu_{t_0}^b(x) = \partial_{x_1}^{s_1} \dots \partial_{x_d}^{s_d} \mu_{t_0}^b(x) .$$

Nous noterons $[\mu_{t_0}^b]_{\mathcal{H}^\beta(x_0)}$ le plus petit réel $C \geq 0$ telle que la condition ci-dessus est vérifiée, et nous munirons $\mathcal{H}^\beta(x_0)$ de la norme

$$|\mu_{t_0}^b|_{\mathcal{H}^\beta(x_0)} := \sup_{x \in \mathcal{U}} |\mu_{t_0}^b(x)| + [\mu_{t_0}^b]_{\mathcal{H}^\beta(x_0)} .$$

Remarque 71. Il serait plus juste de noter cet espace $\mathcal{H}^\beta(U_{x_0})$, mais nous garderons la notation ci-dessus par souci de simplicité.

Nous allons maintenant définir l'*ordre du noyau* K , autrement dit un entier caractérisant la capacité de K à prendre en compte des informations de plus en plus précises sur la fonction à estimer.

Définition 72. Nous dirons que K est un noyau d'ordre $o + 1$ pour un certain $o \in \mathbb{N}^*$, si

$$\int_{\mathbb{R}^d} K(y) y^s dy = 0, \quad \text{pour tout } s \in \mathbb{N}^d \text{ tel que } |s| \leq o,$$

où, pour $y = (y_1, \dots, y_d) \in \mathbb{R}^d$, $y^s := y_1^{s_1} \dots y_d^{s_d}$.

Remarque 73. Un tel noyau existe toujours, nous pouvons par exemple définir K par :

$$K(y) := \prod_{i=1}^d \tilde{K}(y_i), \quad y = (y_1, \dots, y_d) \in \mathbb{R}^d, \quad \text{avec } \tilde{K}(z) := \sum_{k=0}^o \phi_k(0) \phi_k(z) \mathbf{1}_{|z| \leq 1}, \quad z \in \mathbb{R},$$

où $(\phi_k)_{k \in \mathbb{N}}$ est la base orthonormale de $L^2([-1, 1])$ constituée des polynômes de Legendre (cf. [182]).

Supposons à présent que $\mu_{t_0}^b \in \mathcal{H}^\beta(x_0)$ pour un certain réel $\beta > 0$, et que K est un noyau d'ordre $o + 1$ pour un certain entier o . Nous avons alors, en utilisant que $\int_{\mathbb{R}^d} K(y) dy = 1$,

$$\int_{\mathbb{R}^d} K_h(x_0 - x) \mu_{t_0}^b(dx) - \mu_{t_0}^b(x_0) = \int_{\mathbb{R}^d} K(y) (\mu_{t_0}^b(x_0 - hy) - \mu_{t_0}^b(x_0)) dy,$$

et par application du théorème de Taylor, en notant pour $s \in \mathbb{N}^d$, $s! = s_1! \dots s_d!$, ceci vaut

$$\begin{aligned} & \int_{\mathbb{R}^d} K(y) \sum_{0 \leq |s| \leq (\lfloor \beta \rfloor - 1) \wedge (o-1)} \frac{1}{s!} D_x^s \mu_{t_0}^b(x_0) (-h)^{|s|} y_1^{s_1} \dots y_d^{s_d} dy_1 \dots dy_d \\ & + \int_{\mathbb{R}^d} K(y) \sum_{|s| = \lfloor \beta \rfloor \wedge o} \frac{|s|}{s!} \int_0^1 (1-\lambda)^{|s|-1} D_x^s \mu_{t_0}^b(x_0 - \lambda hy) (-h)^{|s|} y_1^{s_1} \dots y_d^{s_d} d\lambda dy_1 \dots dy_d. \end{aligned}$$

Or, par définition de l'ordre de K , pour tout $|s| \leq o$, $\int_{\mathbb{R}^d} K(y) y_1^{s_1} \dots y_d^{s_d} dy_1 \dots dy_d = 0$. Donc le premier terme est nul, et le second se réécrit

$$\sum_{|s| = \lfloor \beta \rfloor \wedge o} \frac{|s|}{s!} (-h)^{|s|} \int_0^1 (1-\lambda)^{|s|-1} \int_{\mathbb{R}^d} K(y) (D_x^s \mu_{t_0}^b(x_0 - \lambda hy) - D_x^s \mu_{t_0}^b(x_0)) y_1^{s_1} \dots y_d^{s_d} dy_1 \dots dy_d d\lambda.$$

En utilisant le fait que $\mu_{t_0}^b \in \mathcal{H}^\beta(x_0)$, nous obtenons finalement

$$\begin{aligned} & \left| \int_{\mathbb{R}^d} K_h(x_0 - x) \mu_{t_0}^b(dx) - \mu_{t_0}^b(x_0) \right| \\ & \leq \sum_{|s| = \lfloor \beta \rfloor \wedge o} \frac{|s|}{s!} h^{|s|} \int_{\mathbb{R}^d} |K(y)| [\mu_{t_0}^b]_{\mathcal{H}^\beta(x_0)} h^{\beta - \lfloor \beta \rfloor} |y|^{\beta - \lfloor \beta \rfloor} |y_1|^{s_1} \dots |y_d|^{s_d} dy_1 \dots dy_d, \\ & \leq h^{\beta - \lfloor \beta \rfloor + \lfloor \beta \rfloor \wedge o} [\mu_{t_0}^b]_{\mathcal{H}^\beta(x_0)} (\lfloor \beta \rfloor \wedge o)^2 \int_{\mathbb{R}^d} |K(y)| |y|^{\beta - \lfloor \beta \rfloor + \lfloor \beta \rfloor \wedge o} dy. \end{aligned}$$

Proposition 74. Si K est d'ordre $o + 1$ où $o \in \mathbb{N}^*$, et $\mu_{t_0}^b \in \mathcal{H}^\beta(x_0)$ où $\beta > 0$,

$$\left| \int_{\mathbb{R}^d} K_h(x_0 - x) \mu_{t_0}^b(dx) - \mu_{t_0}^b(x_0) \right|^2 \lesssim h^{2(\beta - \lfloor \beta \rfloor + \lfloor \beta \rfloor \wedge o)} \lesssim h^{2(\beta \wedge o)} \quad \text{pour tout } h \in (0, 1).$$

Risque quadratique

En regroupant les résultats des Propositions 69 et 74, nous obtenons la borne supérieure suivante sur le risque quadratique de $\hat{\mu}_h^N$ en (t_0, x_0) .

Proposition 75. *Sous les Hypothèses 64 et 67, si K est d'ordre $o+1$ où $o \in \mathbb{N}^*$, et $\mu_{t_0}^b \in \mathcal{H}^\beta(x_0)$ où $\beta > 0$, alors pour tout $h \in \mathcal{H}_1^N$,*

$$\mathcal{R}_{(t_0, x_0)}(\hat{\mu}_h^N, \mu^b) \lesssim h^{2(\beta \wedge o)} + (Nh^d)^{-1},$$

et l'infimum du terme de droite étant atteint en $\hat{h}_{\beta, o}^N := N^{-\frac{1}{2(\beta \wedge o) + d}}$, qui est dans $\mathcal{H}_1^N$ pour N assez grand, nous avons

$$\inf_{h \in \mathcal{H}_1^N} \mathcal{R}_{(t_0, x_0)}(\hat{\mu}_h^N, \mu^b) \leq \mathcal{R}_{(t_0, x_0)}(\hat{\mu}_{\hat{h}_{\beta, o}^N}^N, \mu^b) \lesssim N^{-\frac{2(\beta \wedge o)}{2(\beta \wedge o) + d}}. \quad (27)$$

Risque minimax sur les classes de Hölder

Nous allons maintenant estimer μ^b en (t_0, x_0) , donc en tant que fonction du temps et de l'espace, et nous définissons, pour quantifier sa régularité en ces deux variables, la classe de fonctions suivante.

Classe de Hölder anisotrope

Définition 76. *Soient deux réels $\alpha, \beta > 0$. Nous dirons que μ^b est localement (α, β) -Hölder en (t_0, x_0) , et nous noterons $\mu^b \in \mathcal{H}^{\alpha, \beta}(t_0, x_0)$, s'il existe des voisinages relativement compacts $\mathcal{U}_{t_0}$ de t_0 dans $(0, T)$ et $\mathcal{U}_{x_0}$ de x_0 dans $\mathbb{R}^d$ tel que $\mu^b(x) \in \mathcal{H}^\alpha(t_0)$ uniformément en $x \in \mathcal{U}_{x_0}$ et $\mu_t^b(\cdot) \in \mathcal{H}^\beta(x_0)$ uniformément en $t \in \mathcal{U}_{t_0}$, i.e.*

- il existe $C_1 > 0$, tel que pour tout entier $0 \leq s \leq \lfloor \alpha \rfloor$, pour tous $t, t' \in \mathcal{U}_{t_0}$, et $x \in \mathcal{U}_{x_0}$,

$$|\partial_t^s \mu_{t'}^b(x) - \partial_t^s \mu_t^b(x)| \leq C_1 |t' - t|^{\alpha - \lfloor \alpha \rfloor},$$

- il existe $C_2 > 0$, tel que pour tout $s \in \mathbb{N}^d$, $0 \leq |s| \leq \lfloor \beta \rfloor$, pour tous $x, x' \in \mathcal{U}_{x_0}$, et $t \in \mathcal{U}_{t_0}$,

$$|D_x^s \mu_{t_0}^b(x') - D_x^s \mu_{t_0}^b(x)| \leq C_2 |x' - x|^{\beta - \lfloor \beta \rfloor}.$$

Nous munirons $\mathcal{H}^{\alpha, \beta}(t_0, x_0)$ de la norme

$$|\mu^b|_{\mathcal{H}^{\alpha, \beta}(t_0, x_0)} := \sup_{x \in \mathcal{U}_{x_0}} |\mu^b(\cdot, x)|_{\mathcal{H}^\alpha(t_0)} + \sup_{t \in \mathcal{U}_{t_0}} |\mu^b(t, \cdot)|_{\mathcal{H}^\beta(x_0)}.$$

Soient deux réels $\alpha, \beta > 0$, nous définissons la classe $\mathcal{S}^{\alpha, \beta}(t_0, x_0)$ de solutions de l'équation de McKean-Vlasov par

$$\mathcal{S}_b^{\alpha, \beta}(t_0, x_0) := \left\{ (b, c, \mu_0) \in \mathfrak{T}(\mathbf{b}) ; \mu^b = \mathcal{S}(b, c, \mu_0) \in \mathcal{H}^{\alpha, \beta}(t_0, x_0) \right\},$$

et nous définissons, pour $L > 0$, l'ellipsoïde

$$\mathcal{S}_{b, L}^{\alpha, \beta}(t_0, x_0) := \left\{ (b, c, \mu_0) \in \mathcal{S}_b^{\alpha, \beta}(t_0, x_0) ; |\mu^b|_{\mathcal{H}^{\alpha, \beta}(t_0, x_0)} \leq L \right\}.$$

Ce dernier ensemble va nous servir de modèle de régularité pour la fonction inconnue μ^b .

Risque minimax, Vitesse de convergence optimale & Estimateur oracle

Nous définissons le *risque quadratique sur la classe de régularité* $\mathcal{S}_{\mathbf{b},L}^{\alpha,\beta}(t_0, x_0)$, pour un estimateur $\tilde{\mu}^N$ de μ^b construit à partir de $\mathbb{X}^N$ par

$$\mathcal{R}(\tilde{\mu}^N; \mathcal{S}_{\mathbf{b},L}^{\alpha,\beta}(t_0, x_0)) := \sup_{(b,c,\mu_0) \in \mathcal{S}_{\mathbf{b},L}^{\alpha,\beta}(t_0, x_0)} \mathcal{R}_{(t_0, x_0)}(\tilde{\mu}^N, \mu^b) .$$

D'après (27),

$$\mathcal{R}(\hat{\mu}_{\hat{h}_{\beta,o}^N}^N; \mathcal{S}_{\mathbf{b},L}^{\alpha,\beta}(t_0, x_0)) \lesssim N^{-\frac{2(\beta \wedge o)}{2(\beta \wedge o) + d}} ,$$

ce qui signifie que la *vitesse de convergence* de l'estimateur $\hat{\mu}_{\hat{h}_{\beta,o}^N}^N$ est $v_{\beta,o}^N := N^{-\frac{(\beta \wedge o)}{2(\beta \wedge o) + d}}$, et, si nous choisissons un noyau K d'ordre $o + 1 = \lfloor \beta \rfloor + 2$, nous obtenons alors, en notant $\hat{h}_{\beta}^N := \hat{h}_{\beta, \lfloor \beta \rfloor + 1}^N$,

$$\mathcal{R}(\hat{\mu}_{\hat{h}_{\beta}^N}^N; \mathcal{S}_{\mathbf{b},L}^{\alpha,\beta}(t_0, x_0)) \lesssim N^{-\frac{2\beta}{2\beta + d}} ,$$

et donc que la vitesse de convergence de l'estimateur $\hat{\mu}_{\hat{h}_{\beta}^N}^N$ est $v_{\beta}^N := v_{\beta, \lfloor \beta \rfloor + 1}^N = N^{-\frac{\beta}{2\beta + d}}$.

Ce résultat soulève plusieurs questions. Peut-on améliorer cette vitesse de convergence en prenant un autre estimateur que $\hat{\mu}_{\hat{h}_{\beta}^N}^N$? Quel est la meilleure vitesse de convergence possible? Comment choisir un estimateur si nous ne connaissons pas (α, β) ?

Pour répondre à ces questions, nous introduisons la notion de *risque minimax sur la classe de régularité* $\mathcal{S}_{\mathbf{b},L}^{\alpha,\beta}(t_0, x_0)$:

$$\mathcal{R}^*(\mathcal{S}_{\mathbf{b},L}^{\alpha,\beta}(t_0, x_0)) := \inf_{\tilde{\mu}^N} \mathcal{R}(\tilde{\mu}^N; \mathcal{S}_{\mathbf{b},L}^{\alpha,\beta}(t_0, x_0)) ,$$

où l'infimum est pris sur les estimateurs $\tilde{\mu}^N$ construits à partir de $\mathbb{X}_{t_0}^N$. L'*estimateur oracle*, $\tilde{\mu}^{N, \text{oracle}}$, sur $\mathcal{S}_{\mathbf{b},L}^{\alpha,\beta}(t_0, x_0)$, est défini par :

$$\mathcal{R}(\tilde{\mu}^{N, \text{oracle}}; \mathcal{S}_{\mathbf{b},L}^{\alpha,\beta}(t_0, x_0)) = \mathcal{R}^*(\mathcal{S}_{\mathbf{b},L}^{\alpha,\beta}(t_0, x_0)) .$$

Dans la Section 7.3 du Chapitre 2, nous avons établi la *borne inférieure* suivante sur le risque minimax.

Théorème 77. *Pour tous $\beta, L > 0$, $\mathbf{b} \in \mathfrak{E}$, il existe une constante $C_{\beta, L, \mathbf{b}} > 0$ telle que, pour N assez grand,*

$$\mathcal{R}^*(\mathcal{S}_{\mathbf{b},L}^{\alpha,\beta}(t_0, x_0)) \geq C_{\beta, L, \mathbf{b}} N^{-\frac{2\beta}{2\beta + d}} .$$

Cela implique que l'estimateur $\hat{\mu}_{\hat{h}_{\beta}^N}^N$ atteint bien la *vitesse de convergence optimale minimax* $v_{\beta}^{N, \star} := N^{-\frac{\beta}{2\beta + d}}$ associée à la classe $\mathcal{S}_{\mathbf{b},L}^{\alpha,\beta}(t_0, x_0)$. Nous pouvons donc dire (qu'à une constante près) $\hat{\mu}_{\hat{h}_{\beta}^N}^N$ se comporte comme l'estimateur oracle.

Cependant, en toute généralité, la régularité β de la solution $\mu_{t_0}^b$ est inconnue, et donc l'estimateur oracle est lui-même inconnu. Nous pouvons cependant espérer construire un estimateur $\hat{\mu}^N$ qui se comporte asymptotiquement presque de la même manière que lui, au sens où

$$\mathcal{R}(\hat{\mu}^N; \mathcal{S}_{\mathbf{b},L}^{\alpha,\beta}(t_0, x_0)) = \mathcal{R}^*(\mathcal{S}_{\mathbf{b},L}^{\alpha,\beta}(t_0, x_0))(1 + o(1)) ,$$

et ceci est l'objet de la section suivante.

Estimateur minimax adaptatif & Méthode de Goldenshluger-Lepski

Définition 78. *Un estimateur $\tilde{\mu}^N$, construit à partir de $\mathbb{X}^N$, est dit adaptatif au sens minimax sur $\{\mathcal{S}_{\mathbf{b},L}^{\alpha,\beta}(t_0, x_0); \alpha, \beta, L > 0, \mathbf{b} \in \mathfrak{E}\}$ si pour tous $\alpha, \beta, L > 0, \mathbf{b} \in \mathfrak{E}$, il existe une constante $C_{\alpha,\beta,\mathbf{b},L,(t_0,x_0)} > 0$ tel que, pour N assez grand,*

$$\mathcal{R}(\tilde{\mu}^N; \mathcal{S}_{\mathbf{b},L}^{\alpha,\beta}(t_0, x_0)) \leq C_{\alpha,\beta,\mathbf{b},L,(t_0,x_0)} N^{-\frac{2\beta}{2\beta+d}} .$$

Pour construire un (presque) tel estimateur nous allons utiliser une variante de la méthode de Goldenshluger-Lepski pour l'estimation ponctuelle (voir les articles fondateurs de cette méthode [87, 88, 89], ainsi que [47]; la variante utilisée ici est celle introduite dans [27]) : cette méthode *adaptative* va nous permettre de choisir une fenêtre optimale, et ce uniquement à partir des données $\mathbb{X}^N$, sans connaissance préalable sur la régularité de $\mu_{t_0}^b$.

Dans toute la suite, nous supposons que le noyau K utilisé dans la définition de $\hat{\mu}_h^N$ vérifie l'Hypothèse 64 et que $\mathcal{H}_1^N$ vérifie l'Hypothèse 67.

Définissons tout d'abord le *biais à l'échelle* $h \in \mathcal{H}_1^N$ de μ au point (t_0, x_0) par :

$$\mathcal{B}_h^N(\mu)(t_0, x_0) := \sup_{h' \leq h, h' \in \mathcal{H}_1^N} \left| \int_{\mathbb{R}^d} K_{h'}(x_0 - x) \mu_{t_0}^b(x) dx - \mu_{t_0}^b(x_0) \right|. \quad (28)$$

Nous avons alors

$$\left(\int_{\mathbb{R}^d} K_h(x_0 - x) \mu_{t_0}(dx) - \mu_{t_0}(x_0) \right)^2 \leq \mathcal{B}_h^N(\mu)(t_0, x_0)^2 .$$

Définissons à présent le *proxy* suivant pour la variance de notre estimateur

$$\mathbf{V}_h^N = \varpi_1 |K|_2^2 (\log N) (Nh^d)^{-1}, \quad (29)$$

où $\varpi_1 > 0$ est un *hyperparamètre* à fixer de manière préliminaire. Nous avons alors, d'après la Proposition 69,

$$\mathbb{E}_{\mathbb{P}_b^N} \left[\left(\int_{\mathbb{R}^d} K_h(x_0 - x) (\mu_{t_0}^N - \mu_{t_0}^b)(dx) \right)^2 \right] \lesssim \mathbf{V}_h^N ,$$

et finalement, pour tout $h \in \mathcal{H}_1^N$,

$$\mathcal{R}_{(t_0,x_0)}(\hat{\mu}_h^N, \mu^b) \lesssim \mathcal{B}_h^N(\mu)(t_0, x_0)^2 + \mathbf{V}_h^N .$$

Il nous reste maintenant à choisir une fenêtre *optimale* $\hat{h}^N$, au sens où elle vérifie une *inégalité oracle*, c'est-à-dire que notre fenêtre *fait toujours aussi bien que l'oracle* au sens :

$$\mathcal{R}_{(t_0,x_0)}(\hat{\mu}_{\hat{h}^N}^N, \mu^b) \lesssim \inf_{h \in \mathcal{H}_1^N} \mathcal{R}_{(t_0,x_0)}(\hat{\mu}_h^N, \mu^b) .$$

En réalité, notre estimateur va vérifier l'*inégalité oracle approximative* suivante

$$\mathcal{R}_{(t_0, x_0)}(\widehat{\mu}_{\widehat{h}^N}^N, \mu^b) \lesssim \inf_{h \in \mathcal{H}_1^N} (\mathcal{B}_h^N(\mu)(t_0, x_0)^2 + \mathcal{V}_h^N) .$$

Nous choisissons notre fenêtre optimale $\widehat{h}^N$ uniquement à partir des données $\mathbb{X}_{t_0}^N$ (et donc sans rien savoir de la régularité de $\mu_{t_0}^b$) de la manière suivante :

$$\widehat{h}^N \in \operatorname{argmin}_{h \in \mathcal{H}_1^N} (\mathcal{A}_h^N + \mathcal{V}_h^N), \quad (30)$$

où $\mathcal{A}_h^N$ est une approximation du biais au carré de l'estimateur, et est définie par

$$\mathcal{A}_h^N := \max_{h' \leq h, h' \in \mathcal{H}_1^N} \left\{ (\widehat{\mu}_h^N(t_0, x_0) - \widehat{\mu}_{h'}^N(t_0, x_0))^2 - (\mathcal{V}_h^N + \mathcal{V}_{h'}^N) \right\}_+,$$

ce qui revient donc à comparer deux-à-deux les estimateurs possibles $\{\widehat{\mu}_h^N(t_0, x_0); h \in \mathcal{H}_1^N\}$.

Notre *estimateur de Goldenshluger-Lepski* de $\mu_{t_0}^b(x_0)$ est alors

$$\widehat{\mu}_{\text{GL}}^N(t_0, x_0) := \widehat{\mu}_{\widehat{h}^N}^N(t_0, x_0) .$$

Remarque 79. Cet estimateur dépend du noyau K et de l'hyperparamètre ϖ_1 . En pratique, l'hyperparamètre ϖ_1 doit être réglé de manière préliminaire par d'autres méthodes. De telles approches dans le contexte des méthodes de type Lepski ont été récemment introduites par Lacour *et al.* [117, 183]. Ce défaut est néanmoins commun à toutes les méthodes non-paramétriques qui dépendent du choix d'une fenêtre à partir des données.

Nous obtenons, comme annoncé ci-dessus, l'inégalité oracle approximative suivante, qui permet donc d'évaluer la performance de notre estimateur de $\mu_{t_0}^b(x_0)$.

Théorème 80. *Sous les Hypothèses 64 et 67, supposons que $\widehat{\mu}_{\text{GL}}^N(t_0, x_0)$ est calibré avec $\varpi_1 \geq 16\kappa_2^{-1}\kappa_3$, alors il existe une constante $C_{\mathbf{b}, K, (t_0, x_0)} > 0$ telle que, pour N assez grand, nous avons*

$$\mathcal{R}_{(t_0, x_0)}(\widehat{\mu}_{\text{GL}}^N, \mu^b) \leq C_{\mathbf{b}, K, (t_0, x_0)} \min_{h \in \mathcal{H}_1^N} (\mathcal{B}_h^N(\mu)(t_0, x_0)^2 + \mathcal{V}_h^N) .$$

Enfin, pour terminer notre étude de l'estimation non-paramétrique de μ^b , nous nous intéressons à son *estimation minimax adaptative dans les espaces de Hölder*. Nous obtenons la *borne supérieure adaptative* suivante.

Théorème 81. *Prenons comme espace des fenêtres*

$$\mathcal{H}_1^N := \left\{ e^{-k}; 1 \leq k \leq \frac{\log N - 2 \log(\log N)}{d} \right\} .$$

Pour tout $\mathbf{b} \in \mathfrak{E}$, si $\widehat{\mu}_{\text{GL}}^N(t_0, x_0)$ est construit à partir d'un noyau K vérifiant l'Hypothèse 64 et d'ordre $o + 1 \geq 1$, alors, pour tous $\beta, L > 0$, il existe une constante $C_{\mathbf{b}, K, \beta, L, (t_0, x_0)} > 0$ telle que, pour N assez grand, nous avons

$$\mathcal{R}(\widehat{\mu}_{\text{GL}}^N(t_0, x_0); \mathcal{S}_{\mathbf{b}, L}^{\alpha, \beta}(t_0, x_0)) \leq C_{\mathbf{b}, K, \beta, L, (t_0, x_0)} \left(\frac{\log N}{N} \right)^{2(\beta \wedge o)/(2(\beta \wedge o) + d)} .$$

Nous obtenons (à un terme logarithmique supplémentaire près, conséquence du phénomène de *Lepski-Low*, classique pour l'estimation ponctuelle, voir [127, 135]) le taux de convergence optimal usuel $N^{-(\beta \wedge o)/(2(\beta \wedge o) + d)}$ d'estimation de $\mu_{t_0}^b(x_0)$ comme une fonction de la variable d'état d -dimensionnelle x_0 , pour un t_0 fixé, à partir d'un système de N particules *i.i.d.* $(X_{t_0}^i)_{1 \leq i \leq N}$ de loi $\mu_{t_0}^b$ (ce qui, rappelons-le, serait le cas de nos observations $\mathbb{X}_{t_0}^N$ si b ne dépendait pas de la mesure) et d'un noyau d'ordre $o + 1$. Ce taux de convergence ne dépend bien sûr pas de la régularité de $t \mapsto \mu_t^b(x_0)$.

Nous passons maintenant à l'estimation non-paramétrique de la fonction de drift b . Nous nous appuyerons pour cela sur les méthodes utilisées et les résultats obtenus dans la section précédente. En particulier, toutes les remarques d'ordre général sur les estimateurs à noyaux que nous avons faites précédemment s'appliqueront également *mutatis mutandis* dans la section qui suit.

3.3 Estimation non-paramétrique de $b(\cdot, \cdot, \mu^b)$

Objectif principal & Quantification de la qualité des estimateurs

Nous voulons, pour tout $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$, estimer $b(t_0, x_0, \mu_{t_0}^b) \in \mathbb{R}^d$ à partir des observations continues $\mathbb{X}^N = (\mathbb{X}_t^N)_{t \in [0, T]}$.

Remarque 82. C'est donc bien la fonction $(t, x) \in (0, T) \times \mathbb{R}^d \mapsto b(t, x, \mu_t^b)$ que nous estimons, et non pas la fonction globale $(t, x, \nu) \in (0, T) \times \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \mapsto b(t, x, \nu)$, qui reste en toute généralité inaccessible dans le cadre d'estimation considéré ici. Cependant, dans le cas particulier du modèle de Vlasov, pour lequel $b(t, x, \nu) = b(x, \nu) = G(x) + F \star \nu(x)$, nous pouvons obtenir une estimation globale de la fonction $(t, x, \nu) \mapsto b(t, x, \nu)$: ceci sera l'objet de la Section 3.5 (voir également le Chapitre 2).

Soit $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$. Pour un estimateur quelconque $\tilde{b}^N$ de la fonction $(t, x) \in (0, T) \times \mathbb{R}^d \mapsto b(t, x, \mu_t^b)$, construit à partir de $\mathbb{X}^N$, *i.e.* fonction mesurable de $\mathbb{X}^N$ à valeurs dans $\mathbb{R}^d$, nous définissons le *risque quadratique ponctuel* de $\tilde{b}^N$ en (t_0, x_0) , par

$$\mathcal{R}_{(t_0, x_0)}(\tilde{b}^N, b(\cdot, \cdot, \mu^b)) := \mathbb{E}_{\mathbb{P}_b^N} \left[\left| \tilde{b}^N(t_0, x_0) - b(t_0, x_0, \mu_{t_0}^b) \right|^2 \right].$$

Posons $\pi^b : (t, x) \in (0, T) \times \mathbb{R}^d \mapsto b(t, x, \mu_t^b) \mu_t^b(x) \in \mathbb{R}^d$. Nous allons en fait commencer par estimer $\pi^b(t_0, x_0)$, et nous évaluerons la qualité de notre estimateur de la manière suivante. Pour un estimateur quelconque $\tilde{\pi}^N$ de la fonction π^b , construit à partir de $\mathbb{X}^N$, nous définissons le *risque quadratique ponctuel* de $\tilde{\pi}^N$ en (t_0, x_0) , par

$$\mathcal{R}_{(t_0, x_0)}(\tilde{\pi}^N, \pi^b) := \mathbb{E}_{\mathbb{P}_b^N} \left[\left| \tilde{\pi}^N(t_0, x_0) - \pi^b(t_0, x_0) \right|^2 \right].$$

Estimateurs à noyaux

Estimateurs à noyaux de $\pi^b(t_0, x_0)$

Parmi la classe des estimateurs $\tilde{\pi}^N(t_0, x_0)$ de $\pi^b(t_0, x_0)$ construits à partir de $\mathbb{X}^N$, nous nous intéressons plus particulièrement à la famille des estimateurs à noyaux

$$\{\hat{\pi}_{\mathbf{h}}^N(t_0, x_0) ; \mathbf{h} = (h_1, h_2) \in \mathcal{H}_2\}, \quad \mathcal{H}_2 \subset (0, 1) \times (0, 1),$$

où

$$\widehat{\pi}_{\mathbf{h}}^N(t_0, x_0) := \int_{[0, T] \times \mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) \pi^N(dt, dx) \quad (31)$$

est défini à l'aide de $\pi^N(dt, dx) := \frac{1}{N} \sum_{i=1}^N dX_t^i \delta_{X_t^i}(dx)$, qui est caractérisé par son action sur les fonctions tests $\Phi : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$:

$$\int_{[0, T] \times \mathbb{R}^d} \Phi(t, x) \pi^N(dt, dx) = \frac{1}{N} \sum_{i=1}^N \int_0^T \Phi(t, X_t^i) dX_t^i ,$$

et à l'aide d'un *noyau tensoriel*

$$(H \otimes K)_{\mathbf{h}} : (t, x) \in (0, T) \times \mathbb{R}^d \mapsto H_{h_1}(t) K_{h_2}(x) ,$$

qui est constitué d'un noyau temporel $H : (0, T) \rightarrow \mathbb{R}$ *i.e.* tel que $\int_0^T H(t) dt = 1$, et d'un noyau d'espace $K : \mathbb{R}^d \rightarrow \mathbb{R}$ *i.e.* tel que $\int_{\mathbb{R}^d} K(x) dx = 1$, et pour lesquels

$$H_{h_1}(t) := h_1^{-1} H\left(\frac{t}{h_1}\right) , \quad K_{h_2}(x) := h_2^{-d} K\left(\frac{x}{h_2}\right) .$$

Estimateurs à noyaux de $b(t_0, x_0, \mu_{t_0}^b)$

Parmi la classe des estimateurs $\widetilde{b}^N(t_0, x_0)$ de $b(t_0, x_0, \mu_{t_0}^b)$ construits à partir de $\mathbb{X}^N$, nous nous intéressons plus particulièrement à la famille des estimateurs à noyaux

$$\{\widehat{b}_{h, \mathbf{h}}^N(t_0, x_0)_{\varpi} ; h \in \mathcal{H}_1, \mathbf{h} = (h_1, h_2) \in \mathcal{H}_2\} , \quad \mathcal{H}_1 \subset (0, 1) , \mathcal{H}_2 \subset (0, 1) \times (0, 1) , \varpi > 0 ,$$

définis de la manière suivante :

$$\widehat{b}_{h, \mathbf{h}}^N(t_0, x_0)_{\varpi} := \frac{\widehat{\pi}_{\mathbf{h}}^N(t_0, x_0)}{\widehat{\mu}_h^N(t_0, x_0) \vee \varpi} ,$$

où $\varpi > 0$ est un *hyperparamètre de seuil* qui empêche l'estimateur d'exploser pour de petites valeurs de $\widehat{\mu}_h^N(t_0, x_0)$.

Notre estimateur de $b(t_0, x_0, \mu_{t_0}^b)$ est donc un *estimateur quotient*, de $\pi^b(t_0, x_0) = b(t_0, x_0, \mu_{t_0}^b) \mu_{t_0}^b(x_0)$ par $\mu_{t_0}^b(x_0)$, tout à fait dans l'esprit des estimateurs de type *Nadaraya-Watson pour la régression* [157]. Et nous allons voir dans la section suivante pourquoi la performance de notre estimateur de $b(t_0, x_0, \mu_{t_0}^b)$ va donc être de l'ordre de la pire performance entre l'estimateur de $\mu_{t_0}^b(x_0)$ et celui de $\pi^b(t_0, x_0)$.

Décomposition du risque quadratique ponctuel pour l'estimation de $b(\cdot, \cdot, \mu^b)$

Soit $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$, $(h, \mathbf{h}) \in \mathcal{H}_1 \times \mathcal{H}_2$, et $R, R_1, R_2 > 0$ tel que $[R_1, R_2] \subset (0, T)$. D'après le Théorème 25, nous savons que μ^b est localement bornée autour de (t_0, x_0) :

$$\kappa_4 \leq \inf_{t \in [R_1, R_2], |x - x_0| \leq R} \mu_t^b(x) \leq \sup_{t \in [R_1, R_2], |x - x_0| \leq R} \mu_t^b(x) \leq \kappa_3 .$$

En utilisant la décomposition suivante,

$$\widehat{b}_{h,\mathbf{h}}^N(t_0, x_0)_{\varpi} - b(t_0, x_0, \mu_{t_0}^b) = \frac{\pi^b(t_0, x_0)(\mu_{t_0}^b(x_0) - \widehat{\mu}_h^N(t_0, x_0) \vee \varpi)}{\mu_{t_0}^b(x_0)\widehat{\mu}_h^N(t_0, x_0) \vee \varpi} + \frac{(\widehat{\pi}_h^N(t_0, x_0) - \pi^b(t_0, x_0))\mu_{t_0}^b(x_0)}{\mu_{t_0}^b(x_0)\widehat{\mu}_h^N(t_0, x_0) \vee \varpi},$$

nous obtenons, dès que $\varpi \leq \kappa_4$,

$$|\widehat{b}_{h,\mathbf{h}}^N(t_0, x_0)_{\varpi} - b(t_0, x_0, \mu_{t_0}^b)| \lesssim |\widehat{\mu}_h^N(t_0, x_0) - \mu_{t_0}^b(x_0)| + |\widehat{\pi}_h^N(t_0, x_0) - \pi^b(t_0, x_0)|,$$

et donc, comme nous pouvions le supposer,

$$\mathcal{R}_{(t_0, x_0)}(\widehat{b}_{h,\mathbf{h}}^N, b(\cdot, \cdot, \mu_{t_0}^b)) \lesssim \mathcal{R}_{(t_0, x_0)}(\widehat{\mu}_h^N, \mu^b) + \mathcal{R}_{(t_0, x_0)}(\widehat{\pi}_h^N, \pi^b).$$

Etant donné que nous connaissons déjà l'ordre de $\mathcal{R}_{(t_0, x_0)}(\widehat{\mu}_h^N, \mu^b)$, il nous reste à étudier le terme $\mathcal{R}_{(t_0, x_0)}(\widehat{\pi}_h^N, \pi^b)$.

Décomposition du risque quadratique ponctuel pour l'estimation de π^b

Pour tout $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$, $h \in \mathcal{H}_2$, comme par définition,

$$\mathcal{R}_{(t_0, x_0)}(\widehat{\pi}_h^N, \pi^b) = \mathbb{E}_{\mathbb{P}_b^N} \left[|\widehat{\pi}_h^N(t_0, x_0) - \pi^b(t_0, x_0)|^2 \right],$$

nous utilisons la *décomposition biais-variance approximative* suivante

$$\begin{aligned} \mathcal{R}_{(t_0, x_0)}(\widehat{\pi}_h^N, \pi^b) &\lesssim \mathbb{E}_{\mathbb{P}_b^N} \left[\left| \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) (\pi^N(dt, dx) - \pi^b(t, x) dx dt) \right|^2 \right] \\ &\quad + \left| \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) \pi^b(t, x) dx dt - \pi^b(t_0, x_0) \right|^2. \end{aligned}$$

Nous appelons terme de *biais* (au carré) le terme

$$\left| \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) \pi^b(t, x) dx dt - \pi^b(t_0, x_0) \right|^2,$$

et terme de *variance* le terme

$$\mathbb{E}_{\mathbb{P}_b^N} \left[\left| \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) (\pi^N(dt, dx) - \pi^b(t, x) dx dt) \right|^2 \right].$$

Comme précédemment pour l'estimation de $\mu_{t_0}^b(x_0)$, cette nomenclature est inexacte formellement mais nous la gardons par habitude statistique, et nous traitons ces deux termes séparément.

Dans toute la suite, nous ferons les hypothèses suivantes sur les noyaux H et K :

Hypothèse 83. H (resp. K) est borné et à support compact dans $(0, T)$ (resp. $\mathbb{R}^d$).

Terme de variance

Pour traiter ce terme, nous le redécomposons d'abord de la manière suivante :

$$\begin{aligned}
 & \mathbb{E}_{\mathbb{P}_b^N} \left[\left| \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) (\pi^N(dt, dx) - \pi^b(t, x) dx dt) \right|^2 \right] \\
 & \lesssim \mathbb{E}_{\mathbb{P}_b^N} \left[\left| \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) b(t, x, \mu_t^b) (\mu_t^N - \mu_t^b)(dx) dt \right|^2 \right] \\
 & + \mathbb{E}_{\mathbb{P}_b^N} \left[\left| \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) (b(t, x, \mu_t^N) - b(t, x, \mu_t^b)) \mu_t^N(dx) dt \right|^2 \right] \\
 & + \mathbb{E}_{\mathbb{P}_b^N} \left[\left| \frac{1}{N} \sum_{i=1}^N \int_0^T (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - X_t^i) \sigma(t, X_t^i) dB_t^{i,b} \right|^2 \right].
 \end{aligned}$$

$$\textbf{Premier terme} : \mathbb{E}_{\mathbb{P}_b^N} \left[\left| \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) b(t, x, \mu_t^b) (\mu_t^N - \mu_t^b)(dx) dt \right|^2 \right]$$

Ce premier terme se traite de la même manière que le terme de variance dans la décomposition du risque quadratique de $\hat{\mu}_{\mathbf{h}}^N(t_0, x_0)$, en utilisant l'*inégalité de concentration de type Bernstein* plus générale suivante que nous avons établie dans [65] (nous y reviendrons dans la Section 3.6) :

Théorème 84. *Il existe $\kappa_1 = \kappa_1(\mathbf{b}), \kappa_2 = \kappa_2(\mathbf{b}) > 0$, tel que pour toute mesure de probabilité $\rho(dt)$ sur $[0, T]$ et toute fonction bornée $\Phi : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$,*

$$\mathbb{P}_b^N \left(\int_{[0, T] \times \mathbb{R}^d} \Phi(t, x) (\mu_t^N - \mu_t^b)(dx) \otimes \rho(dt) \geq z \right) \leq \kappa_1 \exp \left(-\kappa_2 \frac{Nz^2}{|\Phi|_{L^2(\mu_t^b(dx) \otimes \rho(dt))}^2 + |\Phi|_{\infty} z} \right), z \geq 0.$$

Remarque 85. Nous en déduisons bien le Théorème 109 pour $t_0 \in (0, T)$ quelconque en prenant $\rho(dt) := \delta_{t_0}(dt)$, et $\Phi : (t, x) \in [0, T] \times \mathbb{R}^d \mapsto \phi(x) \in \mathbb{R}$ bornée.

Nous appliquons le Théorème 84 à la mesure de probabilité $\rho(dt) = T^{-1} \mathbb{1}_{[0, T]}(t) dt$ et, pour $1 \leq k \leq d$, à la fonction

$$\Phi(t, x) := \pm (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) b^k(t, x, \mu_t^b).$$

En utilisant le fait que μ^b et b sont bornées localement d'après le Théorème 25, et comme H et K sont supposées bornées et à supports compacts (respectivement $\text{supp}(H)$ et $\text{supp}(K)$), nous avons

$$|\Phi|_{\infty} \leq (h_1 h_2^d)^{-1} |H \otimes K|_{\infty} \kappa_5,$$

$$|\Phi|_{L^2(\mu_t^b(dx) \otimes \rho(dt))}^2 \leq (h_1 h_2^d)^{-1} T^{-1} |H \otimes K|_2^2 \kappa_5^2 \kappa_3,$$

avec $\kappa_5 = \kappa_5(x_0, \text{supp}(K), \mathbf{b})$, $\kappa_3 = \kappa_3(t_0, x_0, \text{supp}(H), \text{supp}(K), \mathbf{b})$. Cela donne donc

$$\begin{aligned}
 & \mathbb{E}_{\mathbb{P}_b^N} \left[\left| \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) b(t, x, \mu_t^b) (\mu_t^N - \mu_t^b)(dx) dt \right|^2 \right] \\
 & \leq \int_0^{\infty} 2d\kappa_1 \exp \left(-\frac{\kappa_2 T^{-1} N h_1 h_2^d z}{\kappa_3 \kappa_5^2 |H \otimes K|_2^2 + \kappa_5 |H \otimes K|_{\infty} z^{1/2}} \right) dz \\
 & \lesssim \kappa_1 \max \left(\left(\frac{\kappa_2 T^{-1} N h_1 h_2^d}{\kappa_3 \kappa_5^2 |H \otimes K|_2^2} \right)^{-1}, \left(\frac{\kappa_2 T^{-1} N h_1 h_2^d}{\kappa_5 |H \otimes K|_{\infty}} \right)^{-2} \right) \\
 & \lesssim (N h_1 h_2^d)^{-1} + (N h_1 h_2^d)^{-2}.
 \end{aligned}$$

Deuxième terme : $\mathbb{E}_{\mathbb{P}_b^N} \left[\left| \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) (b(t, x, \mu_t^N) - b(t, x, \mu_t^b)) \mu_t^N(dx) dt \right|^2 \right]$

Pour traiter ce terme, nous séparons, grâce à l'inégalité de Cauchy-Schwarz, la partie en $(H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x)$ et la partie en $b(t, x, \mu_t^N) - b(t, x, \mu_t^b)$. Le deuxième terme est donc inférieur à

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_b^N} \left[\left(\int_0^T \int_{\mathbb{R}^d} |(H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x)|^2 \mu_t^N(dx) dt \right)^2 \right]^{1/2} \\ \times \mathbb{E}_{\mathbb{P}_b^N} \left[\left(\int_0^T \int_{\mathbb{R}^d} |b(t, x, \mu_t^N) - b(t, x, \mu_t^b)|^2 \mu_t^N(dx) dt \right)^2 \right]^{1/2}. \end{aligned}$$

D'une part, en décomposant μ_t^N en $\mu_t^b + (\mu_t^N - \mu_t^b)$ et en utilisant les mêmes techniques que précédemment, nous montrons que

$$\mathbb{E}_{\mathbb{P}_b^N} \left[\left(\int_0^T \int_{\mathbb{R}^d} |(H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x)|^2 \mu_t^N(dx) dt \right)^2 \right]^{1/2} \lesssim (h_1 h_2^d)^{-1} + N^{-1/2} (h_1 h_2^d)^{-3/2} + N^{-1} (h_1 h_2^d)^{-2}.$$

D'autre part, par symétrie du système de particules,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_b^N} \left[\left(\int_0^T \int_{\mathbb{R}^d} |b(t, x, \mu_t^N) - b(t, x, \mu_t^b)|^2 \mu_t^N(dx) dt \right)^2 \right]^{1/2} &\lesssim \sup_{t \in [0, T]} \mathbb{E}_{\mathbb{P}_b^N} \left[|b(t, X_t^N, \mu_t^N) - b(t, X_t^N, \mu_t^b)|^4 \right]^{1/2} \\ &\lesssim \left(\kappa_8^4 N^{-2} + \int_{\kappa_8^4 N^{-2}}^{\infty} \sup_{t \in [0, T]} \mathbb{P}_b^N (|b(t, X_t^N, \mu_t^N) - b(t, X_t^N, \mu_t^b)| \geq u^{1/4}) du \right)^{1/2} \lesssim N^{-1}, \end{aligned}$$

où nous avons utilisé l'inégalité de déviation suivante sur $|b(t, X_t^N, \mu_t^N) - b(t, X_t^N, \mu_t^b)|$, prouvée dans [65] (voir également le Chapitre 2) :

Lemme 86. *Il existe $\kappa_6 = \kappa_6(\mathbf{b}), \kappa_7 = \kappa_7(\mathbf{b}), \kappa_8 = \kappa_8(\mathbf{b}) > 0$ tel que, pour N assez grand*

$$\sup_{t \in [0, T]} \mathbb{P}_b^N (|b(t, X_t^N, \mu_t^N) - b(t, X_t^N, \mu_t^b)| \geq u) \leq \kappa_6 \exp \left(- \frac{\kappa_7 N u^2}{1 + N^{1/2} u} \right), u \geq \kappa_8 N^{-1/2}.$$

Nous obtenons finalement l'inégalité suivante sur le deuxième terme :

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_b^N} \left[\left| \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) (b(t, x, \mu_t^N) - b(t, x, \mu_t^b)) \mu_t^N(dx) dt \right|^2 \right] \\ \lesssim (N h_1 h_2^d)^{-1} + (N h_1 h_2^d)^{-3/2} + (N h_1 h_2^d)^{-2}. \end{aligned}$$

Troisième terme : $\mathbb{E}_{\mathbb{P}_b^N} \left[\left| \frac{1}{N} \sum_{i=1}^N \int_0^T (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - X_t^i) \sigma(t, X_t^i) dB_t^{i,b} \right|^2 \right]$

Enfin, ce troisième et dernier terme étant plus petit, à une constante multiplicative près, que

$$N^{-1} \mathbb{E}_{\mathbb{P}_b^N} \left[\int_0^T \int_{\mathbb{R}^d} |(H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x)|^2 \mu_t^N(dx) dt \right],$$

en décomposant comme ci-dessus μ_t^N en $\mu_t^b + (\mu_t^N - \mu_t^b)$, nous trouvons que

$$\mathbb{E}_{\mathbb{P}_b^N} \left[\left| \frac{1}{N} \sum_{i=1}^N \int_0^T (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - X_t^i) \sigma(t, X_t^i) dB_t^{i,b} \right|^2 \right] \lesssim (N h_1 h_2^d)^{-1} + (N h_1 h_2^d)^{-3/2} + (N h_1 h_2^d)^{-2}.$$

Conclusion pour le terme de variance

En additionnant les majorations des trois termes composant la variance, nous obtenons

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_b^N} \left[\left| \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) (\pi^N(dt, dx) - \pi^b(t, x) dx dt) \right|^2 \right] \\ \lesssim (Nh_1 h_2^d)^{-1} + (Nh_1 h_2^d)^{-3/2} + (Nh_1 h_2^d)^{-2} . \end{aligned}$$

Nous nous laissons à présent la possibilité de choisir un ensemble de fenêtres dépendant de N et nous le notons $\mathcal{H}_2^N$. Nous ferons l'hypothèse supplémentaire suivante :

Hypothèse 87.

$$\mathcal{H}_2^N \subset [(N^{-1}(\log N)^2)^{1/(d+1)}, (\log N)^{-2}] \times [(N^{-1}(\log N)^2)^{1/(d+1)}, 1] \text{ et } \text{Card}(\mathcal{H}_2^N) \lesssim N .$$

Remarque 88. Cela implique donc que, pour $\mathbf{h} = (h_1, h_2) \in \mathcal{H}_2^N$, $(Nh_1 h_2^d)^{-1} \lesssim 1$.

Proposition 89. *Sous les Hypothèses 83 et 87,*

$$\mathbb{E}_{\mathbb{P}_b^N} \left[\left| \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) (\pi^N(dt, dx) - \pi^b(t, x) dx dt) \right|^2 \right] \lesssim (Nh_1 h_2^d)^{-1} . \quad (32)$$

Passons maintenant au terme de biais (au carré)

$$\left| \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) \pi^b(t, x) dx dt - \pi^b(t_0, x_0) \right|^2 .$$

Terme de biais

Ce terme se traite de la même manière que le terme de biais de μ^b , via la régularité hölderienne locale de π^b dans un voisinage de (t_0, x_0) . Etant donné que, pour tout $\mathbf{h} = (h_1, h_2) \in \mathcal{H}_2$, nous disposons de la décomposition

$$\begin{aligned} & \left| \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) \pi^b(t, x) dx dt - \pi^b(t_0, x_0) \right| \\ & \leq \int_{\mathbb{R}^d} |K(y)| \left| \int_0^T H(s) (\pi^b(t_0 - h_1 s, x_0 - h_2 y) - \pi^b(t_0, x_0 - h_2 y)) ds \right| dy \\ & \quad + \int_0^T |H(s)| \left| \int_{\mathbb{R}^d} K(y) (\pi^b(t_0, x_0 - h_2 y) - \pi^b(t_0, x_0)) dy \right| ds , \end{aligned}$$

nous avons le résultat suivant.

Proposition 90. *Supposons que $\pi^b \in \mathcal{H}^{\alpha, \beta}(t_0, x_0)$, pour certains $\alpha, \beta > 0$, et que H (resp. K) est un noyau d'ordre $o_H + 1 \in \mathbb{N}^*$ (resp. $o_K + 1 \in \mathbb{N}^*$). Soit $\mathbf{h} = (h_1, h_2) \in \mathcal{H}_2$. Alors*

$$\left| \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) \pi^b(t, x) dx dt - \pi^b(t_0, x_0) \right| \lesssim h_2^{\beta \wedge o_K} + h_1^{\alpha \wedge o_H} .$$

Risque quadratique

En regroupant les résultats des Propositions 89 et 90, nous obtenons la borne supérieure suivante sur le risque quadratique de $\widehat{\pi}_{\mathbf{h}}^N$ en (t_0, x_0) .

Proposition 91. *Sous les Hypothèses 83 et 87, si H (resp. K) est un noyau d'ordre $o_H + 1 \in \mathbb{N}^*$ (resp. $o_K + 1 \in \mathbb{N}^*$), et $\pi^b \in \mathcal{H}^{\alpha, \beta}(t_0, x_0)$, pour certains $\alpha, \beta > 0$, alors pour tout $\mathbf{h} \in \mathcal{H}_2^N$,*

$$\mathcal{R}_{(t_0, x_0)}(\widehat{\pi}_{\mathbf{h}}^N, \pi^b) \lesssim h_1^{\beta \wedge o_H} + h_2^{\beta \wedge o_K} + (Nh_1 h_2^d)^{-1},$$

et l'infimum du terme de droite étant atteint en

$$\widehat{\mathbf{h}}_{\alpha, \beta, o_H, o_K}^N := (\widehat{h}_{1, \alpha, \beta, o_H, o_K}^N, \widehat{h}_{2, \alpha, \beta, o_H, o_K}^N) = \left(N^{-1 / \left(2(\beta \wedge o_K) + d + \frac{\beta \wedge o_K}{\alpha \wedge o_H} \right)}, N^{-1 / \left(2(\alpha \wedge o_H) + 1 + d \frac{\alpha \wedge o_H}{\beta \wedge o_K} \right)} \right),$$

qui est dans $\mathcal{H}_2^N$ pour N assez grand, nous avons

$$\inf_{\mathbf{h} \in \mathcal{H}_2^N} \mathcal{R}_{(t_0, x_0)}(\widehat{\pi}_{\mathbf{h}}^N, \pi^b) \leq \mathcal{R}_{(t_0, x_0)}(\widehat{\pi}_{\widehat{\mathbf{h}}_{\alpha, \beta, o_H, o_K}^N}^N, \pi^b) \lesssim N^{-\frac{2\mathfrak{s}_d(\alpha, \beta, o_H, o_K)}{2\mathfrak{s}_d(\alpha, \beta, o_H, o_K)+1}}, \quad (33)$$

$$\text{où } \mathfrak{s}_d(\alpha, \beta, o_H, o_K) := \left(\frac{1}{\alpha \wedge o_H} + \frac{d}{\beta \wedge o_K} \right)^{-1}.$$

Risque quadratique ponctuel pour l'estimation de $b(\cdot, \cdot, \mu^b)$

Maintenant que nous avons majoré le risque quadratique pour l'estimation de π^b , il nous reste à majorer le risque quadratique pour l'estimation de $b(\cdot, \cdot, \mu^b)$.

Proposition 92. *Sous les Hypothèses 64, 67 83, et 87, si nous supposons que l'estimateur de $\mu_{t_0}^b(x_0)$ est construit à l'aide d'un noyau d'ordre $o + 1 \geq 1$, et que l'estimateur de $\pi^b(t_0, x_0)$ est construit à l'aide de noyaux d'ordre $o_H + 1, o_K + 1 \geq 1$, et que de plus $\mu^b, \pi^b \in \mathcal{H}^{\alpha, \beta}(t_0, x_0)$, alors nous avons, pour tout $(h, \mathbf{h}) \in \mathcal{H}_1^N \times \mathcal{H}_2^N$,*

$$\begin{aligned} \mathcal{R}_{(t_0, x_0)}(\widehat{b}_{h, \mathbf{h}}^N, b(\cdot, \cdot, \mu^b)) &\lesssim \mathcal{R}_{(t_0, x_0)}(\widehat{\mu}_h^N, \mu^b) + \mathcal{R}_{(t_0, x_0)}(\widehat{\pi}_{\mathbf{h}}^N, \pi^b) \\ &\lesssim h^{2(\beta \wedge o)} + (Nh^d)^{-1} + h_1^{\beta \wedge o_H} + h_2^{\beta \wedge o_K} + (Nh_1 h_2^d)^{-1}. \end{aligned}$$

Finalement,

$$\inf_{(h, \mathbf{h}) \in \mathcal{H}_1^N \times \mathcal{H}_2^N} \mathcal{R}_{(t_0, x_0)}(\widehat{b}_{h, \mathbf{h}}^N, b(\cdot, \cdot, \mu^b)) \lesssim \inf_{h \in \mathcal{H}_1^N} \mathcal{R}_{(t_0, x_0)}(\widehat{\mu}_h^N, \mu^b) + \inf_{\mathbf{h} \in \mathcal{H}_2^N} \mathcal{R}_{(t_0, x_0)}(\widehat{\pi}_{\mathbf{h}}^N, \pi^b),$$

et, pour N assez grand,

$$\begin{aligned} \mathcal{R}_{(t_0, x_0)}(\widehat{b}_{\widehat{h}_{\beta, o}^N, \widehat{\mathbf{h}}_{\alpha, \beta, o_H, o_K}^N}^N, b(\cdot, \cdot, \mu^b)) &\lesssim \mathcal{R}_{(t_0, x_0)}(\widehat{\mu}_{\widehat{h}_{\beta, o}^N}^N, \mu^b) + \mathcal{R}_{(t_0, x_0)}(\widehat{\pi}_{\widehat{\mathbf{h}}_{\alpha, \beta, o_H, o_K}^N}^N, \pi^b) \\ &\lesssim N^{-\frac{2(\beta \wedge o)}{2(\beta \wedge o)+d}} + N^{-\frac{2\mathfrak{s}_d(\alpha, \beta, o_H, o_K)}{2\mathfrak{s}_d(\alpha, \beta, o_H, o_K)+1}}. \end{aligned}$$

Risque minimax sur les classes de Hölder

Soient deux réels $\alpha, \beta > 0$. Nous utiliserons comme modèles de régularité pour π^b et $b(\cdot, \cdot, \mu^b)$, la classe

$$\Pi_{\mathbf{b}}^{\alpha, \beta}(t_0, x_0) := \mathcal{S}_{\mathbf{b}}^{\alpha, \beta}(t_0, x_0) \bigcap \mathcal{D}_{\mathbf{b}}^{\alpha, \beta}(t_0, x_0),$$

et, pour $L, L' > 0$, l'ellipsoïde

$$\Pi_{\mathbf{b}, L, L'}^{\alpha, \beta}(t_0, x_0) := \mathcal{S}_{\mathbf{b}, L}^{\alpha, \beta}(t_0, x_0) \bigcap \mathcal{D}_{\mathbf{b}, L'}^{\alpha, \beta}(t_0, x_0),$$

où nous avons adopté les notations suivantes :

$$\mathcal{D}_{\mathbf{b}}^{\alpha, \beta}(t_0, x_0) := \{(b, c, \mu_0) \in \mathfrak{T}(\mathbf{b}) ; b(\cdot, \cdot, \mu^b) \in \mathcal{H}^{\alpha, \beta}(t_0, x_0)\},$$

$$\mathcal{D}_{\mathbf{b}, L'}^{\alpha, \beta}(t_0, x_0) := \{(b, c, \mu_0) \in \mathfrak{T}(\mathbf{b}) ; |b(\cdot, \cdot, \mu^b)|_{\mathcal{H}^{\alpha, \beta}(t_0, x_0)} \leq L'\}.$$

Remarque 93. Etant donné que la condition $(b, c, \mu_0) \in \Pi_{\mathbf{b}}^{\alpha, \beta}(t_0, x_0)$ entraîne que μ^b et $b(\cdot, \cdot, \mu^b)$ sont dans $\mathcal{H}^{\alpha, \beta}(t_0, x_0)$, et étant donné que par définition $\pi^b(t, x) = b(t, x, \mu_t^b)\mu_t^b(x)$, cela implique également que $\pi^b \in \mathcal{H}^{\alpha, \beta}(t_0, x_0)$.

Risque minimax, Vitesse de convergence optimale & Estimateur oracle

Nous définissons le *risque quadratique sur la classe de régularité* $\Pi_{\mathbf{b}, L, L'}^{\alpha, \beta}(t_0, x_0)$ pour un estimateur $\tilde{b}^N$ de $b(\cdot, \cdot, \mu^b)$ construit à partir de $\mathbb{X}^N$ par

$$\mathcal{R}(\tilde{b}^N ; \Pi_{\mathbf{b}, L, L'}^{\alpha, \beta}(t_0, x_0)) := \sup_{(b, c, \mu_0) \in \Pi_{\mathbf{b}, L, L'}^{\alpha, \beta}(t_0, x_0)} \mathcal{R}_{(t_0, x_0)}(\tilde{b}^N, b(\cdot, \cdot, \mu^b)).$$

D'après la Proposition 92,

$$\mathcal{R}(\hat{b}_{\hat{h}_{\beta, o}^N, \hat{h}_{\alpha, \beta, o_H, o_K}^N}^N ; \Pi_{\mathbf{b}, L, L'}^{\alpha, \beta}(t_0, x_0)) \lesssim N^{-\frac{2(\beta \wedge o)}{2(\beta \wedge o) + d}} + N^{-\frac{2s_d(\alpha, \beta, o_H, o_K)}{2s_d(\alpha, \beta, o_H, o_K) + 1}},$$

ce qui signifie que la *vitesse de convergence* de l'estimateur $\hat{b}_{\hat{h}_{\beta, o}^N, \hat{h}_{\alpha, \beta, o_H, o_K}^N}^N$ est

$$v_{\alpha, \beta, o, o_H, o_K}^N := N^{-\frac{2(\beta \wedge o)}{2(\beta \wedge o) + d}} + N^{-\frac{2s_d(\alpha, \beta, o_H, o_K)}{2s_d(\alpha, \beta, o_H, o_K) + 1}},$$

et, si nous choisissons un noyau d'ordre $o + 1 = \lfloor \beta \rfloor + 2$ pour l'estimation de μ^b , et des noyaux d'ordres respectifs $o_H + 1 = \lfloor \alpha \rfloor + 2$ et $o_K + 1 = \lfloor \beta \rfloor + 2$ pour l'estimation de π^b , nous obtenons alors, en notant

$$\hat{h}_{\beta}^N := \hat{h}_{\beta, \lfloor \beta \rfloor + 1}^N, \quad \hat{h}_{\alpha, \beta}^N := \hat{h}_{\alpha, \beta, \lfloor \alpha \rfloor + 1, \lfloor \beta \rfloor + 1}^N,$$

que

$$\mathcal{R}(\hat{b}_{\hat{h}_{\beta}^N, \hat{h}_{\alpha, \beta}^N}^N ; \Pi_{\mathbf{b}, L, L'}^{\alpha, \beta}(t_0, x_0)) \lesssim N^{-\frac{2\beta}{2\beta + d}} + N^{-\frac{2s_d(\alpha, \beta)}{2s_d(\alpha, \beta) + 1}},$$

où $s_d(\alpha, \beta) := s_d(\alpha, \beta, \lfloor \alpha \rfloor + 1, \lfloor \beta \rfloor + 1) = (\frac{1}{\alpha} + \frac{d}{\beta})^{-1}$ est la *régularité effective anisotrope*. Or $\frac{2\beta}{2\beta + d} = \frac{2(\frac{d}{\beta})^{-1}}{2(\frac{d}{\beta})^{-1} + 1} \geq \frac{2s_d(\alpha, \beta)}{2s_d(\alpha, \beta) + 1}$, donc

$$\mathcal{R}(\hat{b}_{\hat{h}_{\beta}^N, \hat{h}_{\alpha, \beta}^N}^N ; \Pi_{\mathbf{b}, L, L'}^{\alpha, \beta}(t_0, x_0)) \lesssim N^{-\frac{2s_d(\alpha, \beta)}{2s_d(\alpha, \beta) + 1}},$$

et la *vitesse de convergence de l'estimateur* $\hat{b}_{\hat{h}_{\beta}^N, \hat{h}_{\alpha, \beta}^N}^N$ est donc

$$v_{\alpha, \beta}^N := v_{\alpha, \beta, \lfloor \beta \rfloor + 1, \lfloor \alpha \rfloor + 1, \lfloor \beta \rfloor + 1}^N = N^{-\frac{2s_d(\alpha, \beta)}{2s_d(\alpha, \beta) + 1}}.$$

Comme lors de l'estimation de μ^b , ce résultat soulève plusieurs questions. Peut-on améliorer ce taux de convergence en prenant un autre estimateur que $\widehat{b}_{\widehat{h}_\beta^N, \widehat{h}_{\alpha, \beta}^N}^N$? Quel est la meilleure vitesse de convergence possible ? Comment choisir un estimateur si nous ne connaissons pas (α, β) ?

Pour répondre à ces questions, nous introduisons, comme lors de l'estimation de μ^b , la notion de *risque minimax sur la classe de régularité* $\Pi_{\mathfrak{b}, L, L'}^{\alpha, \beta}(t_0, x_0)$:

$$\mathcal{R}^*(\Pi_{\mathfrak{b}, L, L'}^{\alpha, \beta}(t_0, x_0)) := \inf_{\widetilde{b}^N} \mathcal{R}(\widetilde{b}^N; \Pi_{\mathfrak{b}, L, L'}^{\alpha, \beta}(t_0, x_0)) ,$$

où l'infimum est pris sur les estimateurs de $b(\cdot, \cdot, \mu^b)$ construits à partir de $\mathbb{X}^N = (\mathbb{X}_t^N)_{t \in [0, T]}$. L'estimateur oracle, $\widetilde{b}^{N, \text{oracle}}$, sur $\Pi_{\mathfrak{b}, L, L'}^{\alpha, \beta}(t_0, x_0)$, est défini par :

$$\mathcal{R}(\widetilde{b}^{N, \text{oracle}}; \Pi_{\mathfrak{b}, L, L'}^{\alpha, \beta}(t_0, x_0)) = \mathcal{R}^*(\Pi_{\mathfrak{b}, L, L'}^{\alpha, \beta}(t_0, x_0)) .$$

Dans la Section 7.4 du Chapitre 2, nous avons établi la *borne inférieure* suivante sur le risque minimax.

Théorème 94. *Pour tous $\alpha, \beta, L, L' > 0, \mathfrak{b} \in \mathfrak{E}$, il existe une constante $C_{\alpha, \beta, L, L', \mathfrak{b}} > 0$ telle que, pour N assez grand,*

$$\mathcal{R}^*(\Pi_{\mathfrak{b}, L, L'}^{\alpha, \beta}(t_0, x_0)) \geq C_{\alpha, \beta, L, L', \mathfrak{b}} N^{-\frac{2s_d(\alpha, \beta)}{2s_d(\alpha, \beta)+1}} .$$

Cela implique que l'estimateur $\widehat{b}_{\widehat{h}_\beta^N, \widehat{h}_{\alpha, \beta}^N}^N$ atteint bien la *vitesse de convergence optimale minimax* $v_{\alpha, \beta}^{N, \star} := N^{-\frac{s_d(\alpha, \beta)}{2s_d(\alpha, \beta)+1}}$ associée à la classe $\Pi_{\mathfrak{b}, L, L'}^{\alpha, \beta}(t_0, x_0)$. Nous pouvons donc dire (qu'à une constante près) $\widehat{b}_{\widehat{h}_\beta^N, \widehat{h}_{\alpha, \beta}^N}^N$ se comporte comme l'estimateur oracle. Cependant, en toute généralité, la régularité (α, β) de la fonction $b(\cdot, \cdot, \mu^b)$ est inconnue, et donc l'estimateur oracle est lui-même inconnu. Nous pouvons cependant espérer construire un estimateur $\widetilde{b}^N$ qui se comporte asymptotiquement presque de la même manière que lui, au sens où :

$$\mathcal{R}(\widetilde{b}^N; \Pi_{\mathfrak{b}, L, L'}^{\alpha, \beta}(t_0, x_0)) = \mathcal{R}^*(\Pi_{\mathfrak{b}, L, L'}^{\alpha, \beta}(t_0, x_0))(1 + o(1)) ,$$

et ceci est l'objet de la section suivante.

Estimateur minimax adaptatif & Méthode de Goldenshluger-Lepski

Définition 95. *Un estimateur $\widetilde{b}^N$, construit à partir de $\mathbb{X}^N$, est dit adaptatif au sens minimax sur $\{\Pi_{\mathfrak{b}, L, L'}^{\alpha, \beta}(t_0, x_0); \alpha, \beta, L, L' > 0, \mathfrak{b} \in \mathfrak{E}\}$ si pour tous $\alpha, \beta, L, L' > 0, \mathfrak{b} \in \mathfrak{E}$, il existe une constante $C_{\alpha, \beta, \mathfrak{b}, L, L'} > 0$ telle que, pour N assez grand,*

$$\mathcal{R}(\widetilde{b}^N; \Pi_{\mathfrak{b}, L, L'}^{\alpha, \beta}(t_0, x_0)) \leq C_{\alpha, \beta, \mathfrak{b}, L, L'} N^{-\frac{2s_d(\alpha, \beta)}{2s_d(\alpha, \beta)+1}} .$$

Pour construire un (presque) tel estimateur nous allons utiliser, pour le choix d'une fenêtre optimale dans le cadre de l'estimation de $\pi^b(t_0, x_0)$, la même variante de la *méthode de Goldenshluger-Lepski* pour l'estimation ponctuelle que celle utilisée précédemment pour estimer $\mu_{t_0}^b(x_0)$: cette

méthode *adaptive* va nous permettre de choisir une fenêtre optimale, et ce uniquement à partir des données $\mathbb{X}^N$, sans connaissance préalable sur la régularité de π^b . Nous allons également réutiliser l'estimateur $\hat{\mu}_{\text{GL}}^N(t_0, x_0)$ de $\mu_{t_0}^b(x_0)$.

Dans toute la suite, nous supposons que les noyaux H et K utilisés dans la définition de $\hat{\pi}_{\mathbf{h}}^N$ vérifient l'Hypothèse 83 et que $\mathcal{H}_2^N$ vérifie l'Hypothèse 87. De plus, nous supposons que

Hypothèse 96. $\mathcal{H}_2^N$ est muni d'une relation d'ordre $\preceq$ tel que pour tous $\mathbf{h}, \mathbf{h}' \in \mathcal{H}_2^N$, ou bien $\mathbf{h} \preceq \mathbf{h}'$, ou bien $\mathbf{h}' \preceq \mathbf{h}$.

Définissons tout d'abord le *biais* à l'échelle $\mathbf{h} \in \mathcal{H}_2^N$ de π^b au point (t_0, x_0) par :

$$\mathcal{B}_{\mathbf{h}}^N(\pi^b)(t_0, x_0) := \sup_{\mathbf{h}' \preceq \mathbf{h}, \mathbf{h}' \in \mathcal{H}_2^N} \left| \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) \pi^b(t, x) dx dt - \pi^b(t_0, x_0) \right|. \quad (34)$$

Nous avons alors

$$\left| \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) \pi^b(t, x) dx dt - \pi^b(t_0, x_0) \right|^2 \leq \mathcal{B}_{\mathbf{h}}^N(\pi^b)(t_0, x_0).$$

Définissons à présent le *proxy* suivant pour la variance de notre estimateur

$$\mathbf{V}_{\mathbf{h}}^N = \varpi_2 |H \otimes K|_2^2 (\log N) (N h_1 h_2^d)^{-1}, \quad (35)$$

où $\varpi_2 > 0$ est un *hyperparamètre* à fixer de manière préliminaire. Nous avons alors, d'après la Proposition 89,

$$\mathbb{E}_{\mathbb{P}_b^N} \left[\left| \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) (\pi^N(dt, dx) - \pi^b(t, x) dx dt) \right|^2 \right] \lesssim \mathbf{V}_{\mathbf{h}}^N,$$

et finalement, pour tout $\mathbf{h} \in \mathcal{H}_2^N$,

$$\mathcal{R}_{(t_0, x_0)}(\hat{\pi}_{\mathbf{h}}^N, \pi^b) \lesssim \mathcal{B}_{\mathbf{h}}^N(\pi^b)(t_0, x_0)^2 + \mathbf{V}_{\mathbf{h}}^N,$$

ce qui entraîne également globalement que, pour tout $(h, \mathbf{h}) \in \mathcal{H}_1^N \times \mathcal{H}_2^N$,

$$\mathcal{R}_{(t_0, x_0)}(\hat{b}_{h, \mathbf{h}}^N, b(\cdot, \cdot, \mu^b)) \lesssim (\mathcal{B}_h^N(\mu^b)(t_0, x_0)^2 + \mathbf{V}_h^N) + (\mathcal{B}_{\mathbf{h}}^N(\pi^b)(t_0, x_0)^2 + \mathbf{V}_{\mathbf{h}}^N).$$

Il nous reste maintenant à choisir une fenêtre *optimale* $\hat{\mathbf{h}}^N \in \mathcal{H}_2^N$, au sens où elle vérifie une *inégalité oracle*, c'est-à-dire que notre fenêtre *fait toujours aussi bien que l'oracle* au sens :

$$\mathcal{R}_{(t_0, x_0)}(\hat{\pi}_{\hat{\mathbf{h}}^N}^N, \pi^b) \lesssim \inf_{\mathbf{h} \in \mathcal{H}_2^N} \mathcal{R}_{(t_0, x_0)}(\hat{\pi}_{\mathbf{h}}^N, \pi^b).$$

En réalité, notre estimateur va vérifier l'*inégalité oracle approximative* suivante

$$\mathcal{R}_{(t_0, x_0)}(\hat{\pi}_{\hat{\mathbf{h}}^N}^N, \pi^b) \lesssim \inf_{\mathbf{h} \in \mathcal{H}_2^N} (\mathcal{B}_{\mathbf{h}}^N(\pi)(t_0, x_0)^2 + \mathbf{V}_{\mathbf{h}}^N).$$

Nous choisissons notre fenêtre optimale uniquement à partir des données $\mathbb{X}^N$ (et donc sans rien savoir de la régularité de π^b) de la manière suivante :

$$\hat{\mathbf{h}}^N \in \operatorname{argmin}_{\mathbf{h} \in \mathcal{H}_2^N} (\mathbf{A}_{\mathbf{h}}^N + \mathbf{V}_{\mathbf{h}}^N), \quad (36)$$

où $A_{\mathbf{h}}^N$ est une approximation du biais au carré de l'estimateur, et est définie par

$$A_{\mathbf{h}}^N := \max_{\mathbf{h}' \preceq \mathbf{h}, \mathbf{h}' \in \mathcal{H}_2^N} \left\{ \left| \hat{\pi}_{\mathbf{h}}^N(t_0, x_0) - \hat{\pi}_{\mathbf{h}'}^N(t_0, x_0) \right|^2 - (V_{\mathbf{h}}^N + V_{\mathbf{h}'}^N) \right\}_+,$$

ce qui revient donc à comparer deux-à-deux les estimateurs possibles $\{\hat{\pi}_{\mathbf{h}}^N(t_0, x_0); \mathbf{h} \in \mathcal{H}_2^N\}$ de $\pi^b(t_0, x_0)$.

Notre *estimateur de Goldenshluger-Lepski* de $\pi^b(t_0, x_0)$ est alors

$$\hat{\pi}_{\text{GL}}^N(t_0, x_0) := \hat{\pi}_{\hat{\mathbf{h}}^N}^N(t_0, x_0).$$

Remarque 97. Cet estimateur dépend des noyaux H et K et de l'hyperparamètre ϖ_2 . En pratique, comme l'hyperparamètre ϖ_1 pour l'estimation de μ^b , l'hyperparamètre ϖ_2 doit être réglé de manière préliminaire par d'autres méthodes.

Nous obtenons, comme annoncé ci-dessus, l'inégalité oracle approximative suivante, qui nous permet donc d'évaluer la performance de notre estimateur de $\pi^b(t_0, x_0)$.

Théorème 98. *Sous les Hypothèses 83, 87 et 96, si $\hat{\pi}_{\text{GL}}^N(t_0, x_0)$ est calibré avec*

$$\varpi_2 \geq 12d\kappa_3 \max(12T\kappa_2^{-1}\kappa_5^2, 25|\text{Tr}(c)|_\infty),$$

il existe une constante $C_{\mathbf{b}, H, K, (t_0, x_0)} > 0$ telle que, pour N assez grand, nous avons

$$\mathcal{R}_{(t_0, x_0)}(\hat{\pi}_{\text{GL}}^N, \pi) \leq C_{\mathbf{b}, H, K, (t_0, x_0)} \min_{\mathbf{h} \in \mathcal{H}_2^N} (\mathcal{B}_{\mathbf{h}}^N(\pi)(t_0, x_0)^2 + V_{\mathbf{h}}^N).$$

Enfin, l'estimateur de Goldenshluger-Lepski de $b(t_0, x_0, \mu_{t_0})$ construit à partir des données $\mathbb{X}^N$ est alors

$$\hat{b}_{\text{GL}}^N(t_0, x_0) := \hat{b}_{\hat{\mathbf{h}}^N, \hat{\mathbf{h}}^N}^N(t_0, x_0)_\varpi = \frac{\hat{\pi}_{\hat{\mathbf{h}}^N}^N(t_0, x_0)}{\hat{\mu}_{\hat{\mathbf{h}}^N}^N(t_0, x_0) \vee \varpi}, \quad (37)$$

où le seuil $\varpi > 0$ est un hyperparamètre à calibrer. Et nous obtenons l'inégalité oracle approximative suivante, qui nous permet d'évaluer la qualité de notre estimateur de $b(t_0, x_0, \mu_{t_0}^b)$.

Théorème 99. *Sous les Hypothèses 64, 67, 83, 87 et 96, si $\hat{\mu}_{\text{GL}}^N(t_0, x_0)$ et $\hat{\pi}_{\text{GL}}^N(t_0, x_0)$ sont calibrés respectivement avec*

$$\varpi_1 \geq 16\kappa_2^{-1}\kappa_3, \quad \varpi_2 \geq 12d\kappa_3 \max(12T\kappa_2^{-1}\kappa_5^2, 25|\text{Tr}(c)|_\infty),$$

et si de plus $\varpi \leq \kappa_4$, alors il existe une constante $C_{\mathbf{b}, K_0, H, K, (t_0, x_0)} > 0$, où K_0 est le noyau utilisé pour construire $\hat{\mu}_{\text{GL}}^N(t_0, x_0)$, telle que, pour N assez grand, nous avons

$$\begin{aligned} \mathcal{R}_{(t_0, x_0)}(\hat{b}_{\text{GL}}^N, b(\cdot, \cdot, \mu^b)) &\lesssim \mathcal{R}_{(t_0, x_0)}(\hat{\mu}_{\text{GL}}^N, \mu^b) + \mathcal{R}_{(t_0, x_0)}(\hat{\pi}_{\text{GL}}^N, \pi^b) \\ &\leq C_{\mathbf{b}, K_0, H, K, (t_0, x_0)} \\ &\quad \times \left(\min_{\mathbf{h} \in \mathcal{H}_1^N} (\mathcal{B}_{\mathbf{h}}^N(\mu)(t_0, x_0)^2 + V_{\mathbf{h}}^N) + \min_{\mathbf{h} \in \mathcal{H}_2^N} (\mathcal{B}_{\mathbf{h}}^N(\pi)(t_0, x_0)^2 + V_{\mathbf{h}}^N) \right). \end{aligned}$$

Enfin, pour terminer notre étude de l'estimation non-paramétrique de π^b et $b(\cdot, \cdot, \mu^b)$, nous nous intéressons à leur *estimation minimax adaptative dans les espaces de Hölder*.

Dans cette dernière partie, nous ferons l'hypothèse suivante sur (α, β) .

Hypothèse 100. α est une fonction croissante de β , telle que la fonction $\beta \mapsto \frac{\alpha(\beta)}{\beta}$ est décroissante sur $\mathbb{R}_+^*$.

Remarque 101. Ceci est bien le cas lorsque α dépend affinement de β , par exemple lorsque $\alpha = \frac{\beta}{2}$.

Posons $\delta^N : k \in \mathbb{N}^* \mapsto \frac{g^{-1}(k/\log N)}{\alpha(g^{-1}(k/\log N))}k$, et $\beta_k := g^{-1}(k/\log N)$, où $g : \beta \mapsto \frac{1}{\beta} \frac{s_d(\alpha(\beta), \beta)}{2s_d(\alpha(\beta), \beta)+1}$, et prenons comme espace des fenêtres

$$\mathcal{H}_2^N := \left\{ (e^{-k_1}, e^{-k_2}) ; k_1 = \delta^N(k_2), 1 \leq k_2 \leq \frac{\log(N/(\log N)^2)}{d+1}, \right. \\ \left. 2 \log \log N \leq \delta^N(k_2) \leq \frac{\log(N/(\log N)^2)}{d+1} \right\}.$$

Nous définissons un ordre $\preceq$ sur $\mathcal{H}_2^N$ par :

$$\mathbf{h} = (h_1, h_2) \preceq \mathbf{h}' = (h'_1, h'_2) \text{ si } h_1(k_2)^{\alpha(\beta_{k_2})} + h_2(k_2)^{\beta_{k_2}} \leq h'_1(k'_2)^{\alpha(\beta_{k'_2})} + h'_2(k'_2)^{\beta_{k'_2}}, \quad (38)$$

où $\mathbf{h} = (h_1, h_2) = (h_1(k_2), h_2(k_2)) = (e^{-\delta^N(k_2)}, e^{-k_2})$ et de même pour $\mathbf{h}'$. Le fait que $\mathbf{h} \preceq \mathbf{h}'$ se ramène alors à la seule condition $k_2 \geq k'_2$.

Nous obtenons la *borne supérieure adaptative minimax* suivante pour l'estimation de $\pi^b(t_0, x_0)$.

Théorème 102. Pour tout $\mathbf{b} \in \mathfrak{E}$, si $\hat{\pi}_{\text{GL}}^N(t_0, x_0)$ est construit à partir de noyaux H et K d'ordres respectifs $o_H + 1, o_K + 1 \geq 1$ et vérifiant l'Hypothèse 83, et est calibré avec

$$\varpi_2 \geq 12d\kappa_3 \max(12T\kappa_2^{-1}\kappa_5^2, 25|\text{Tr}(c)|_\infty),$$

alors pour tous $\alpha, \beta, L, L' > 0$, il existe une constante $C_{\mathbf{b}, H, K, \alpha, \beta, L, L', (t_0, x_0)} > 0$ telle que, pour N assez grand, nous avons

$$\mathcal{R}(\hat{\pi}_{\text{GL}}^N(t_0, x_0); \Pi_{\mathbf{b}, L, L'}^{\alpha, \beta}(t_0, x_0)) \leq C_{\mathbf{b}, H, K, \alpha, \beta, L, L', (t_0, x_0)} \left(\frac{\log N}{N} \right)^{\frac{2s_d(\alpha, \beta, o_H, o_K)}{2s_d(\alpha, \beta, o_H, o_K)+1}},$$

$$\text{où } s_d(\alpha, \beta, o_H, o_K) = \left(\frac{1}{\alpha \wedge o_H} + \frac{d}{\beta \wedge o_K} \right)^{-1}.$$

Enfin, nous obtenons la *borne supérieure adaptative minimax* suivante pour l'estimation de $b(t_0, x_0, \mu_{t_0}^b)$.

Théorème 103. Prenons $\mathcal{H}_1^N$ et $\mathcal{H}_2^N$ des espaces de fenêtres définis comme lors de la construction de $\hat{\mu}_{\text{GL}}^N(t_0, x_0)$ et $\hat{\pi}_{\text{GL}}^N(t_0, x_0)$. Pour tout $\mathbf{b} \in \mathfrak{E}$, si $\hat{\mu}_{\text{GL}}^N(t_0, x_0)$ (resp. $\hat{\pi}_{\text{GL}}^N(t_0, x_0)$) est construit à partir d'un noyau K_0 d'ordre $o + 1 \geq 1$ vérifiant l'Hypothèse 64 (resp. de noyaux H et K d'ordres $o_H + 1, o_K + 1 \geq 1$ vérifiant l'Hypothèse 83) et est calibré avec

$$\varpi_1 \geq 16\kappa_2^{-1}\kappa_3 \quad (\text{resp. } \varpi_2 \geq 12d\kappa_3 \max(12T\kappa_2^{-1}\kappa_5^2, 25|\text{Tr}(c)|_\infty)),$$

et si de plus $\varpi \leq \kappa_4$, alors, pour tous $\alpha, \beta, L, L' > 0$, il existe une constante

$$C_{\mathbf{b}, K_0, H, K, \alpha, \beta, L, L', (t_0, x_0)} > 0$$

telle que, pour N assez grand, nous avons

$$\begin{aligned} \mathcal{R}(\widehat{b}_{\text{GL}}^N(t_0, x_0); \Pi_{\mathbf{b}, L, L'}^{\alpha, \beta}(t_0, x_0)) &\leq C_{\mathbf{b}, K_0, H, K, \alpha, \beta, L, L', (t_0, x_0)} \\ &\times \left(\left(\frac{\log N}{N} \right)^{2(\beta \wedge o)/(2(\beta \wedge o) + d)} + \left(\frac{\log N}{N} \right)^{\frac{2s_d(\alpha, \beta, o_H, o_K)}{2s_d(\alpha, \beta, o_H, o_K) + 1}} \right), \end{aligned}$$

ce qui entraîne, si tous les noyaux sont du même ordre $o + 1 \geq 1$, que

$$\mathcal{R}(\widehat{b}_{\text{GL}}^N(t_0, x_0); \Pi_{\mathbf{b}, L, L'}^{\alpha, \beta}(t_0, x_0)) \leq C_{\mathbf{b}, K_0, H, K, \alpha, \beta, L, L', (t_0, x_0)} \left(\frac{\log N}{N} \right)^{\frac{2(\frac{1}{\alpha \wedge o} + \frac{d}{\beta \wedge o})^{-1}}{2(\frac{1}{\alpha \wedge o} + \frac{d}{\beta \wedge o})^{-1} + 1}},$$

et donc, pour $o_d := \frac{o}{d+1}$,

$$\mathcal{R}(\widehat{b}_{\text{GL}}^N(t_0, x_0); \Pi_{\mathbf{b}, L, L'}^{\alpha, \beta}(t_0, x_0)) \lesssim \left(\frac{\log N}{N} \right)^{\frac{2(s_d(\alpha, \beta) \wedge o_d)}{2(s_d(\alpha, \beta) \wedge o_d) + 1}}.$$

Ce dernier théorème montre bien qu'estimer $b(t_0, x_0, \mu_{t_0}^b)$ a la même complexité qu'estimer d fonctions de $1 + d$ variables en temps et espace : chaque composante $t \mapsto b^k(t, x, \mu_t)$ étant de régularité α , et chaque composante $x^\ell \mapsto b^k(t, (x^1, \dots, x^\ell, \dots, x^d), \mu_t)$ étant de régularité β pour $\ell = 1, \dots, d$, cela revient à estimer la fonction $(t, x) \mapsto b^k(t, x, \mu_t^b)$ de $1 + d$ variables de régularité anisotrope $(\alpha, \beta, \dots, \beta)$. Nous retrouvons donc bien la vitesse de convergence minimax anisotrope usuelle, pour la régularité effective $s_d(\alpha, \beta)$, qui est reliée à la moyenne harmonique de $(\alpha, \beta, \dots, \beta)$.

3.4 Comparaison : paramétrique ou non-paramétrique ?

Tout au long de cette thèse, nous avons suivi le fil conducteur suivant : comment, à partir de l'observation du système de particules en interaction $\mathbb{X}^N$, estimer $\mu^b = \mathcal{S}(b, c, \mu_0)$ et $b(\cdot, \cdot, \mu^b)$?

Or, si nous considérons que $\mu \in \{\mathcal{S}(b_0(\vartheta; \cdot), c, \mu_0); \vartheta \in \Theta\}$ où Θ est un sous-ensemble compact de $\mathbb{R}^p$, b_0 est connu à ϑ près, c et μ_0 sont connus, alors $\mu = \mu^{\vartheta_0} = \mathcal{S}(b_0(\vartheta_0; \cdot), c, \mu_0)$ pour un certain $\vartheta_0 \in \Theta$, et donc théoriquement, grâce aux résultats vus dans le cadre paramétrique, nous pourrions prendre comme estimateur de μ^b , au point (t_0, x_0) ,

$$\widehat{\mu}^{N, \text{mle}}(t_0, x_0) := \mathcal{S}(b_0(\widehat{\vartheta}^{N, \text{mle}}; \cdot), c, \mu_0)(t_0, x_0),$$

où $\widehat{\vartheta}^{N, \text{mle}}$ est l'estimateur du maximum de vraisemblance de ϑ_0 , et donc $\widehat{\mu}^{N, \text{mle}}(t_0, x_0) = \mu_{t_0}^{\widehat{\vartheta}^{N, \text{mle}}}(x_0)$.

Remarque 104. Dans les faits, l'application $\mathcal{S}$ agit le plus souvent comme une application inconnue, une sorte de boîte noire, et nous ne pouvons alors pas construire de manière effective un estimateur ainsi. Cependant, il existe quelques cas particuliers où l'application $\mathcal{S}$ est explicite, et ce sont ces cas particuliers que nous considérons ici.

Nous pouvons nous demander s'il serait ainsi possible d'améliorer la précision de notre estimation de $\mu_{t_0}^b(x_0)$ par rapport au taux de convergence obtenu à l'aide de notre estimateur non-paramétrique $\widehat{\mu}_{h^N}^N(t_0, x_0)$. Etudions pour cela le risque quadratique de $\widehat{\mu}^{N, \text{mle}}(t_0, x_0)$.

Dans toute cette section, nous nous plaçons sous l'hypothèse suivante.

Hypothèse 105. $\vartheta \mapsto \mu_t^\vartheta(x)$ est $\mathcal{C}^1$ sur l'intérieur de Θ , uniformément par rapport à (t, x) .

Nous avons

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} \left[\left| \hat{\mu}^{N, \text{mle}}(t_0, x_0) - \mu_{t_0}^{\vartheta_0}(x_0) \right|^2 \right] &= \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} \left[\left| \int_0^1 \nabla_{\vartheta} \mu_{t_0}^{\vartheta_0 + \lambda(\hat{\vartheta}^{N, \text{mle}} - \vartheta_0)}(x_0) \cdot (\hat{\vartheta}^{N, \text{mle}} - \vartheta_0) d\lambda \right|^2 \right] \\ &\leq \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} \left[\left| \hat{\vartheta}^{N, \text{mle}} - \vartheta_0 \right|^4 \right]^{1/2} \\ &\quad \times \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} \left[\sup_{\lambda \in [0, 1]} \left| \nabla_{\vartheta} \mu_{t_0}^{\vartheta_0 + \lambda(\hat{\vartheta}^{N, \text{mle}} - \vartheta_0)}(x_0) \right|^4 \right]^{1/2} \\ &\lesssim N^{-1} |I_{\mathcal{G}}(\vartheta_0)|^{-1} \sup_{\vartheta \in \Theta} |\nabla_{\vartheta} \mu_{t_0}^{\vartheta}(x_0)|, \end{aligned}$$

d'où

$$\mathcal{R}_{(t_0, x_0)}(\hat{\mu}^{N, \text{mle}}, \mu^{\vartheta_0}) = \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} \left[\left| \hat{\mu}^{N, \text{mle}}(t_0, x_0) - \mu_{t_0}^{\vartheta_0}(x_0) \right|^2 \right] \lesssim N^{-1},$$

et nous obtenons ainsi un taux de convergence de N^{-1} , ce qui améliore le taux $N^{-\frac{2\beta}{2\beta+d}}$ obtenu dans le cadre non-paramétrique (rappelons que β désigne ici la régularité hölderienne de $\mu_{t_0}^{\vartheta_0}$ en x dans un voisinage de x_0).

Passons maintenant à l'estimation de $b(t_0, x_0, \mu_{t_0}^b) = b_0(\vartheta_0; t_0, x_0, \mu_{t_0}^{\vartheta_0})$. Nous pouvons utiliser comme estimateur théorique

$$\hat{b}^{N, \text{mle}}(t_0, x_0) := b_0(\hat{\vartheta}^{N, \text{mle}}; t_0, x_0, \hat{\mu}^{N, \text{mle}}(t_0, \cdot)),$$

qui comporte le risque quadratique suivant :

$$\begin{aligned} \mathcal{R}_{(t_0, x_0)}(\hat{b}^{N, \text{mle}}(t_0, x_0), b(\cdot, \cdot, \mu^b)) &= \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} \left[\left| b_0(\hat{\vartheta}^{N, \text{mle}}; t_0, x_0, \hat{\mu}^{N, \text{mle}}(t_0, \cdot)) - b_0(\vartheta_0; t_0, x_0, \mu_{t_0}^{\vartheta_0}) \right|^2 \right] \\ &\lesssim \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} \left[\left| b_0(\hat{\vartheta}^{N, \text{mle}}; t_0, x_0, \hat{\mu}^{N, \text{mle}}(t_0, \cdot)) - b_0(\hat{\vartheta}^{N, \text{mle}}; t_0, x_0, \mu_{t_0}^{\vartheta_0}) \right|^2 \right] \\ &\quad + \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} \left[\left| b_0(\hat{\vartheta}^{N, \text{mle}}; t_0, x_0, \mu_{t_0}^{\vartheta_0}) - b_0(\vartheta_0; t_0, x_0, \mu_{t_0}^{\vartheta_0}) \right|^2 \right] \\ &\lesssim \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} [\mathcal{W}_1(\hat{\mu}^{N, \text{mle}}(t_0, \cdot), \mu_{t_0}^{\vartheta_0})^2] \\ &\quad + \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} \left[\left| \hat{\vartheta}^{N, \text{mle}} - \vartheta_0 \right|^2 \right] (1 + |x_0|^{2r_1} + \mathbf{m}_{2r_2}(\mu_{t_0}^{\vartheta_0})). \end{aligned}$$

Enfin, comme $\hat{\mu}^{N, \text{mle}}(t_0, \cdot) = \mu_{t_0}^{\hat{\vartheta}^{N, \text{mle}}}(\cdot)$, nous avons, d'après le Théorème 25,

$$\mathcal{W}_1(\hat{\mu}^{N, \text{mle}}(t_0, \cdot), \mu_{t_0}^{\vartheta_0}) \lesssim |\hat{\vartheta}^{N, \text{mle}} - \vartheta_0|,$$

d'où

$$\mathcal{R}_{(t_0, x_0)}(\hat{b}^{N, \text{mle}}(t_0, x_0), b(\cdot, \cdot, \mu^b)) \lesssim \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} \left[\left| \hat{\vartheta}^{N, \text{mle}} - \vartheta_0 \right|^2 \right] \lesssim N^{-1},$$

et nous obtenons ainsi un taux de convergence de N^{-1} , ce qui améliore le taux $N^{-\frac{2s_d(\alpha, \beta)}{2s_d(\alpha, \beta)+1}}$ obtenu dans le cadre non-paramétrique (rappelons que (α, β) désigne ici la régularité hölderienne de μ^{ϑ_0} en (t, x) dans un voisinage de (t_0, x_0) et que $\frac{1}{s_d(\alpha, \beta)} = \frac{1}{\alpha} + \frac{d}{\beta}$).

Dans une dernière section, nous présentons une méthode proposée dans l'article [65] pour estimer le noyau d'interaction dans le cadre non-paramétrique, lorsque la fonction de drift vérifie une relation de type Vlasov.

3.5 Estimation du noyau d'interaction dans le cadre non-paramétrique

Nous nous plaçons ici dans le cadre dit de *Vlasov* (voir [173]) avec un *drift homogène en temps* et une *diffusion constante* :

$$b(t, x, \nu) = b(x, \nu) = \int_{\mathbb{R}^d} \tilde{b}(x, y) \nu(dy), \text{ où } \tilde{b}(x, y) = F(x - y) + G(x), \quad x, y \in \mathbb{R}^d,$$

$$\sigma(t, x) = \sigma I_d, \sigma > 0.$$

Nous noterons ici $\mathbb{P}_b^N = \mathbb{P}_{(F,G)}^N$, μ à la place de μ^b , et nous ferons dans toute la suite de cette section l'hypothèse d'intégrabilité suivante :

Hypothèse 106. $F \in L^1(\mathbb{R}^d)$, $G \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$.

Comme la fonction de drift se réécrit dans ce cadre

$$\beta(t, x) = b(x, \mu_t) = G(x) + F \star \mu_t(x), \quad (39)$$

l'EDP (2) se réécrit

$$\begin{cases} \partial_t \mu_t = \frac{\sigma^2}{2} \Delta \mu_t - \operatorname{div}((G + F \star \mu_t) \mu_t), & (t, x) \in (0, T) \times \mathbb{R}^d, \\ \mu|_{t=0} = \mu_0. \end{cases}$$

et l'EDS (5) se réécrit

$$\begin{cases} dX_t^i = G(X_t^i)dt + N^{-1} \sum_{j=1}^N F(X_t^i - X_t^j)dt + \sigma dB_t^i, & 1 \leq i \leq N, \quad t \in [0, T], \\ \mathcal{L}(X_0^1, \dots, X_0^N) = \mu_0^{\otimes N}. \end{cases}$$

Nous nous intéressons au problème de l'identification de la fonction d'interaction $x \mapsto F(x)$ à partir des données $\mathbb{X}^N$, puis de l'identification de $x \mapsto G(x)$, que nous considérons cependant dans un premier temps plutôt comme un paramètre de nuisance : notre problème s'apparente donc à un *problème de déconvolution avec un bruit additif*, et la solution proposée s'inspire de Johannes [106].

Nous introduisons pour traiter ce problème une forme linéaire $\mathcal{L}$ qui agit sur les fonctions tests $\varphi : [0, T] \rightarrow \mathbb{C}$ de la manière suivante

$$\mathcal{L}\varphi := \int_{[0,T]} \varphi(t) w(t) \rho(dt), \quad (40)$$

où ρ est une certaine mesure de probabilité sur $[0, T]$ et $w : [0, T] \rightarrow \mathbb{R}$ est une fonction de *poids* vérifiant l'hypothèse

Hypothèse 107. w fonction bornée, à support compact, et telle que $\int_{[0,T]} w(t) \rho(dt) = 0$.

Nous noterons, pour $\Phi : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{C}$,

$$\text{pour tout } x \in \mathbb{R}^d, \mathcal{L}\Phi(\cdot, x) := \int_{[0,T]} \Phi(t, x) w(t) \rho(dt), \text{ et } \mathcal{L}\Phi : x \in \mathbb{R}^d \mapsto \mathcal{L}\Phi(\cdot, x).$$

En appliquant $\mathcal{L}$ de chaque côté de l'équation (39), nous obtenons

$$\mathcal{L}\beta(\cdot, x) = (F \star \mathcal{L}\mu)(x), \quad x \in \mathbb{R}^d, \quad (41)$$

ce qui implique en utilisant les propriétés de la transformée de Fourier $\mathcal{F}$, et en notant

$$\mathcal{F}(\mathcal{L}\beta)(\xi) := \int_{\mathbb{R}^d} e^{-2i\pi\xi^\top x} \mathcal{L}\beta(\cdot, x) dx ,$$

que

$$\mathcal{F}(\mathcal{L}\beta) = \mathcal{F}(F) \cdot \mathcal{F}(\mathcal{L}\mu) ,$$

i.e. que

$$\mathcal{F}(\mathcal{L}b(\cdot, \mu)) = \mathcal{F}(F) \cdot \mathcal{F}(\mathcal{L}\mu) ,$$

et ce qui nous donne finalement la décomposition formelle

$$\mathcal{F}(F) = \frac{\mathcal{F}(\mathcal{L}b(\cdot, \mu))}{\mathcal{F}(\mathcal{L}\mu)} = \frac{\mathcal{F}(\mathcal{L}b(\cdot, \mu)) \cdot \overline{\mathcal{F}(\mathcal{L}\mu)}}{|\mathcal{F}(\mathcal{L}\mu)|^2} \quad (42)$$

dès que ce quotient est bien défini.

Nous prenons comme estimateur de $\mathcal{F}(\mathcal{L}\mu)(\xi) = \mathcal{L}\mathcal{F}(\mu)(\xi)$ le *périodogramme*

$$\mathcal{L}\left(\int_{\mathbb{R}^d} e^{-2i\pi\xi^\top x} \mu^N(dx)\right) = \int_{\mathbb{R}^d} e^{-2i\pi\xi^\top x} \mathcal{L}\mu^N(dx) = \mathcal{F}(\mathcal{L}\mu^N)(\xi)$$

pour lequel nous n'avons pas besoin d'ajuster de fenêtre, et nous obtenons donc un estimateur $\widehat{F}_{(\varpi, \varpi', h, \mathbf{h}, r)}^N$ de F par la formule

$$\mathcal{F}(\widehat{F}_{(\varpi, \varpi', h, \mathbf{h}, r)}^N) := \frac{\mathcal{F}(\mathcal{L}(\widehat{b}_{h, \mathbf{h}}^N(\cdot, \cdot)_{\varpi'}^r)) \cdot \overline{\mathcal{F}(\mathcal{L}\mu^N)}}{|\mathcal{F}(\mathcal{L}\mu^N)|^2} \mathbf{1}_{\{|\mathcal{F}(\mathcal{L}\mu^N)|^2 \geq \varpi\}} \in \mathbb{R}^d , \quad (43)$$

avec

$$\widehat{b}_{h, \mathbf{h}}^N(t_0, x_0)_{\varpi'}^r := \begin{cases} \widehat{b}_{h, \mathbf{h}}^N(t_0, x_0)_{\varpi'} & \text{si } |x_0| \leq r \\ 0 & \text{sinon,} \end{cases}$$

comme estimateur *tronqué* de $b(t_0, x_0, \mu_{t_0}^b)$, construit à partir de l'estimateur $\widehat{b}_{h, \mathbf{h}}^N(t_0, x_0)_{\varpi'}$ de la Section 3.3.

Notre estimateur de F dépend donc de cinq paramètres : d'une part deux seuils $\varpi, r > 0$ qui lui sont propres, d'autre part $\varpi' > 0$ et $(h, \mathbf{h})$ qui sont respectivement le seuil et les fenêtres dans la définition de l'estimateur de $b(t_0, x_0, \mu_{t_0}^b)$.

Nous obtenons finalement le théorème de consistance suivant.

Théorème 108. *Si $|\mathcal{F}(\mathcal{L}\mu)(\xi)| > 0$ $d\xi$ -presque partout, il existe un choix de $(\varpi, \varpi', h, \mathbf{h}) = (\varpi_N, \varpi'_N, h_N, \mathbf{h}_N) \rightarrow 0$ et $r = r_N \rightarrow \infty$ tel que*

$$\mathbb{E}_{\mathbb{P}_{(F, G)}^N} [|\widehat{F}_{(\varpi, \varpi', h, \mathbf{h}, r)}^N - F|_2^2] \rightarrow 0 \text{ quand } N \rightarrow \infty.$$

Nous pouvons donc bien reconstruire la force d'interaction F à partir des données $\mathbb{X}^N$. Nous en déduisons également l'estimateur de G suivant :

$$\widehat{G}_{(\varpi, \varpi', h, \mathbf{h}, r)}^N := -\widehat{b}_{h, \mathbf{h}}^N(t_0, \cdot)_{\varpi'}^r - \widehat{F}_{(\varpi, \varpi', h, \mathbf{h}, r)}^N \star \widehat{\mu}_h^N(t_0, \cdot),$$

pour $t_0 \in (0, T)$ quelconque, et nous pouvons obtenir un résultat de consistance sur cet estimateur de la même manière que pour l'estimateur de F .

Enfin, nous consacrons la section suivante à la présentation d'une nouvelle inégalité de concentration de type Bernstein que nous avons établie au cours de notre étude statistique du modèle de McKean-Vlasov et de son système de particules en interaction, ainsi qu'aux résultats intermédiaires de type probabiliste qui nous ont permis d'y parvenir.

3.6 Inégalité de concentration de type Bernstein pour $\mathbb{P}_b^N$

Nous avons vu que les estimations non-paramétriques de μ^b et $b(\cdot, \cdot, \mu^b)$ reposent en grande partie sur l'inégalité de déviation de type Bernstein suivante que nous avons prouvée dans [65].

Théorème 109. *Il existe $\kappa_1 = \kappa_1(\mathbf{b}), \kappa_2 = \kappa_2(\mathbf{b}) > 0$, tels que, pour toute mesure de probabilité $\rho(dt)$ sur $[0, T]$, pour toute fonction bornée $\Phi : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$,*

$$\mathbb{P}_b^N \left(\int_{[0, T] \times \mathbb{R}^d} \Phi(t, x) (\mu_t^N - \mu_t^b)(dy) \otimes \rho(dt) \geq z \right) \leq \kappa_1 \exp \left(-\kappa_2 \frac{Nz^2}{|\Phi|_{L^2(\mu_t^b(dy) \otimes \rho(dt))}^2 + |\Phi|_\infty z} \right), \quad z \geq 0.$$

Cela entraîne que, pour tout $t_0 \in (0, T)$, pour toute fonction bornée $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$,

$$\mathbb{P}_b^N \left(\int_{\mathbb{R}^d} \phi(x) (\mu_{t_0}^N - \mu_{t_0}^b)(dy) \geq z \right) \leq \kappa_1 \exp \left(-\kappa_2 \frac{Nz^2}{|\phi|_{L^2(\mu_{t_0}^b)}^2 + |\phi|_\infty z} \right), \quad z \geq 0.$$

Nous donnons dans la suite la trame de la démonstration de ce résultat, ainsi que des résultats intermédiaires que nous avons établis (ou utilisés concernant le Théorème 2 de [74]).

Système limite de particules indépendantes

Rappelons que la probabilité N -tensorielle $\overline{\mathbb{P}}_b^{\otimes N}$ est l'unique mesure de probabilité sur $(\mathcal{C}^N, \mathcal{F}_T^N)$ sous laquelle $\mathcal{L}(\mathbb{X}_0^N) = \mu_0^{\otimes N}$ et le processus suivant est un mouvement brownien standard sur $(\mathbb{R}^d)^N$:

$$\overline{\mathbb{B}}^{N,b} = (\overline{B}_t^{i,b})_{1 \leq i \leq N, t \in [0, T]} := \left(\int_0^t c^{-1/2}(s, X_s^i) (dX_s^i - b(s, X_s^i, \mu_s^b) ds) \right)_{1 \leq i \leq N, t \in [0, T]}.$$

Sous $\overline{\mathbb{P}}_b^{\otimes N}$, le processus canonique $\mathbb{X}^N = (X^1, \dots, X^N)$ est donc solution de

$$\begin{cases} dX_t^i = b(t, X_t^i, \mu_t^b) dt + \sigma(t, X_t^i) d\overline{B}_t^{i,b}, & 1 \leq i \leq N, \quad t \in [0, T], \\ \mathcal{L}(X_0^1, \dots, X_0^N) = \mu_0^{\otimes N}. \end{cases} \quad (44)$$

Le processus réel

$$\overline{M}_t^{N,b} := \sum_{i=1}^N \int_0^t ((c^{-1/2}b)(s, X_s^i, \mu_s^N) - (c^{-1/2}b)(s, X_s^i, \mu_s^b))^\top d\overline{B}_s^{i,b}, \quad (45)$$

est quant à lui une $\overline{\mathbb{P}}_b^{\otimes N}$ -martingale.

Soit $\mathcal{E}_t(\overline{M}^{N,b}) = \exp(\overline{M}_t^{N,b} - \frac{1}{2}\langle \overline{M}^{N,b} \rangle_t)$ l'exponentielle de Doléans-Dade de $\overline{M}_t^{N,b}$ et $\langle \overline{M}^{N,b} \rangle_t$ son compensateur prédictible. D'après le Théorème de Girsanov, la $\mathbb{P}_b^{\otimes N}$ -martingale locale $\mathcal{E}_t(\overline{M}^{N,b})$ est en fait une vraie $\mathbb{P}_b^{\otimes N}$ -martingale, et $\mathbb{P}_b^N$ et $\mathbb{P}_b^{\otimes N}$ sont absolument continues l'une par rapport à l'autre, avec

$$\frac{d\mathbb{P}_b^N}{d\mathbb{P}_b^{\otimes N}} = \mathcal{E}_T(\overline{M}^{N,b}).$$

La proposition suivante (qui constitue en fait un analogue du *critère de Novikov* du Théorème de Girsanov, dans sa version *discrétisée* développée dans le livre [?], Lemme 5.14, p.198) permet de contrôler uniformément en N le changement de probabilité entre $\mathbb{P}_b^{\otimes N}$ et $\mathbb{P}_b^N$.

Proposition 110. *Pour tout $\tau > 0$, il existe $\delta_0 = \delta_0(\tau, \mathbf{b}) > 0$ tel que*

$$\sup_{N \geq 1} \sup_{t \in [0, T-\delta]} \mathbb{E}_{\mathbb{P}_b^{\otimes N}} \left[\exp(\tau(\langle \overline{M}^{N,b} \rangle_{t+\delta} - \langle \overline{M}^{N,b} \rangle_t)) \right] \leq C_1,$$

pour tout $0 \leq \delta \leq \delta_0$ et un certain $C_1 = C_1(\tau, \mathbf{b}) > 0$.

Soit $\mathcal{A}^{N,b} \in \mathcal{F}_T^N$. Comme $\mathbb{P}_b^N$ et $\mathbb{P}_b^{\otimes N}$ coïncident sur $\mathcal{F}_0^N$, nous avons

$$\mathbb{P}_b^N(\mathcal{A}^{N,b}) = \mathbb{E}_{\mathbb{P}_b^N}[\mathbb{P}_b^N(\mathcal{A}^{N,b} | \mathcal{F}_0)] = \mathbb{E}_{\mathbb{P}_b^{\otimes N}}[\mathbb{P}_b^N(\mathcal{A}^{N,b} | \mathcal{F}_0)].$$

Pour toute subdivision $0 = t_0 < t_1 < \dots < t_K \leq T$ et tout évènement $\mathcal{F}_T^N$ -mesurable $\mathcal{A}^{N,b}$, nous obtenons par des techniques de type *grandes déviations* (voir par exemple Gärtner [78] ou le Théorème 2.6 de Lacker [116]) que

$$\mathbb{E}_{\mathbb{P}_b^N}[\mathbb{P}_b^N(\mathcal{A}^{N,b} | \mathcal{F}_0)] \leq \mathbb{E}_{\mathbb{P}_b^{\otimes N}}[\mathbb{P}_b^N(\mathcal{A}^{N,b} | \mathcal{F}_T)]^{1/4^K} \prod_{j=1}^K \mathbb{E}_{\mathbb{P}_b^{\otimes N}} \left[\exp(2(\langle \overline{M}^{N,b} \rangle_{t_j} - \langle \overline{M}^{N,b} \rangle_{t_{j-1}})) \right]^{j/4},$$

ce qui entraîne, pour $\tau = 2$, $t_j = jT/K$ et K assez grand pour que $t_j - t_{j-1} \leq \delta_0$, que

$$\begin{aligned} \mathbb{P}_b^N(\mathcal{A}^{N,b}) &\leq \mathbb{P}_b^{\otimes N}(\mathcal{A}^{N,b})^{1/4^K} \sup_{N \geq 1} \sup_{t \in [0, T-\delta_0]} \left(\mathbb{E}_{\mathbb{P}_b^{\otimes N}} \left[\exp(2(\langle \overline{M}^{N,b} \rangle_{t+\delta_0} - \langle \overline{M}^{N,b} \rangle_t)) \right] \right)^{K(K+1)/8} \\ &\leq C_1^{K(K+1)/8} \mathbb{P}_b^{\otimes N}(\mathcal{A}^{N,b})^{1/4^K}. \end{aligned}$$

Enfin, en prenant

$$\mathcal{A}^{N,b} := \left\{ \int_{[0,T] \times \mathbb{R}^d} \Phi(t, x) (\mu_t^N - \mu_t^b)(dx) \otimes \rho(dt) \geq z \right\},$$

nous obtenons le résultat voulu grâce à l'*inégalité de Bernstein* usuelle suivante (voir [145]) : pour $Z_1, \dots, Z_N$ des v.a. réelles centrées, indépendantes et bornées par $Q > 0$,

$$\mathbb{P}\left(\sum_{i=1}^N Z_i \geq y\right) \leq \exp\left(-\frac{y^2}{2(\sum_{i=1}^N \mathbb{E}[Z_i^2] + \frac{Qy}{3})}\right) \text{ pour tout } y \geq 0.$$

Dans la suite de cette section, nous présentons les résultats intermédiaires essentiels à la preuve de la Proposition 110, et donc de cette inégalité de concentration.

Contrôle de la distance entre $b(s, X_s^i, \mu_s^N)$ et $b(s, X_s^i, \mu_s^b)$

C'est en effet l'élément essentiel de la preuve de la Proposition 110, puisque par définition

$$\langle \overline{M}^{N,b} \rangle_{t+\delta} - \langle \overline{M}^{N,b} \rangle_t \leq \sigma_-^{-2} \sum_{i=1}^N \int_t^{t+\delta} |b(s, X_s^i, \mu_s^N) - b(s, X_s^i, \mu_s^b)| ds.$$

Lemme 111. Notons $\xi_s^N(x) = b(s, x, \mu_s^N) - b(s, x, \mu_s^b)$. Nous avons alors, sous l'Hypothèse 16,

$$|\xi_s^N(X_s^i)|^2 \leq C \begin{cases} \mathcal{W}_1(\mu_s^N, \mu_s^b)^2 & \text{dans le cas (i),} \\ \sum_{\ell=1}^{k_b^*-1} \frac{1}{\ell!} \left| \int_{(\mathbb{R}^d)^\ell} \delta_\mu^\ell b(s, X_s^i, y^\ell, \mu_s^b) (\mu_s^N - \mu_s^b)^{\otimes \ell}(dy^\ell) \right|^2 \\ + \mathcal{W}_1(\mu_s^N, \mu_s^b)^2 \wedge \mathcal{W}_1(\mu_s^N, \mu_s^b)^{2k_b^*} & \text{dans le cas (ii),} \\ \sum_{\ell=1}^{k_b^*-1} \frac{1}{\ell!} \left| \int_{(\mathbb{R}^d)^\ell} \delta_\mu^\ell b(s, X_s^i, y^\ell, \mu_s^b) (\mu_s^N - \mu_s^b)^{\otimes \ell}(dy^\ell) \right|^2 & \text{dans le cas (iii),} \end{cases}$$

pour un certain $C \geq 1$ explicite ne dépendant que de $\mathfrak{b}$.

Le second élément essentiel à la preuve de la Proposition 110 est l'inégalité de déviation suivante provenant du Théorème 2 de Fournier et Guillin [74] :

$$\sup_{0 \leq s \leq T} \overline{\mathbb{P}}_b^{\otimes N}(\mathcal{W}_1(\mu_s^N, \mu_s^b) \geq x) \lesssim \varepsilon_N(x),$$

$$\text{avec } \varepsilon_N(x) = \begin{cases} \exp(-CNx^2) & \text{si } d = 1, \\ \exp(-CN \frac{x^2}{(\log(2+1/x))^2}) \mathbf{1}_{\{x \leq 1\}} + \exp(-CNx^2) \mathbf{1}_{\{x > 1\}} & \text{si } d = 2, \\ \exp(-CNx^d) \mathbf{1}_{\{x \leq 1\}} + \exp(-CNx^2) \mathbf{1}_{\{x > 1\}} & \text{si } d \geq 3, \end{cases}$$

où $C = C(d, \mathfrak{c}_5) > 0$, ce qui garantit que la Condition (1) du Théorème 2 de [74] est satisfaite, d'où l'uniformité en $s \in [0, T]$.

Enfin, le troisième et dernier élément essentiel à la preuve de la Proposition 110 est le résultat suivant, établi dans [65] (cf Chapitre 2), que l'on peut relier à certaines bornes sur les *moments des U-statistiques* :

Lemme 112. Soit $p \geq 1$. Pour $1 \leq \ell \leq k < N$, pour toute fonction $f : [0, T] \times \mathbb{R}^d \times (\mathbb{R}^d)^\ell \rightarrow \mathbb{R}^d$ lipschitzienne en les variables d'espace, nous avons

$$\mathbb{E}_{\overline{\mathbb{P}}_b^{\otimes N}} \left[\left| \int_{(\mathbb{R}^d)^\ell} f(s, X_s^N, y^\ell) (\mu_s^N - \mu_s^b)^{\otimes \ell}(dy^\ell) \right|^{2p} \right] \leq \frac{p! K_\ell^p}{(N-k)^p} |f(s, \cdot)|_{\text{Lip}}^{2p}, \quad (46)$$

pour une constante $K_\ell = K_\ell(\mathfrak{b}) > 0$ explicite.

Ces résultats extraits de l'article [65] (cf. Chapitre 2), nous ont également été essentiels pour prouver la Proposition 32 (i.e. contrôler la distance en variation totale entre $\overline{\mathbb{P}}_b^{\otimes N}$ et $\mathbb{P}_b^N$).

4 Composition de la thèse

La suite de cette thèse est divisée en trois parties :

- le [*Chapitre I*] est consacré à l'article d'estimation paramétrique

The LAN property for McKean-Vlasov models in a mean-field regime ,

co-écrit avec le Pr. Marc Hoffmann, accepté pour publication, voir [64] ;

- le [*Chapitre II*] est consacré à l'article d'estimation non-paramétrique

Nonparametric estimation for interacting particle systems : McKean-Vlasov models ,

co-écrit avec le Pr. Marc Hoffmann, publié, voir [65] ;

- le [*Chapitre III*] est consacré aux Annexes : nous y démontrons les résultats, énoncés dans l'Introduction, supplémentaires par rapport aux deux articles pré-cités.

CHAPITRE 1

THE LAN PROPERTY FOR MCKEAN-VLASOV MODELS IN A MEAN-FIELD REGIME

1 Introduction

1.1 Motivation

Collective dynamics models are becoming increasingly popular in modelling complex stochastic systems, with a versatiliy of applications, ranging from mathematical biology (neurosciences, Bala-dron *et al.* [6], structured models in population dynamics, Mogilner *et al.* [154], Burger *et al.* [32]) to social sciences (opinion dynamics, Chazelle *et al.* [43], cooperative behaviours, Canuto *et al.* [33]) and finance (systemic risk, Fouque and Sun [73]), or more recently, mean-field games (Cardaliaguet *et al.* [35], Cardaliaguet and Lehalle [36]). Whereas stochastic systems of interacting particles and associated nonlinear Markov processes in the sense of McKean [146] date back to the 1960's and have been studied extensively over more than half a century, see *e.g.* [21, 173, 172, 148, 179] among a myriad of references, the development of statistical inference in this setting is only emerging, (with some notable exceptions like Löcherbach [134] in large time or *e.g.* Kasonga [111] or Bishwal [16]) in a mean-field limit. Recently, Giesecke *et al.* [84] and Sharrock, Kantas, Parpas and Grigo-rios [169] revisit the work of Kasonga and consider a parametric framework where convergent and asymptotically normal contrast estimators are constructed. Several other parametric frameworks (that consider various observation schemes and asymptotic frameworks) have also been recently considered, like [44, 132, 189] or Genon-Catalot and Laredo [81, 82]. There also exist recent results in nonparametric inference : we mention our work [65] and Belometsny *et al.* [8], together with studies in identification like [120, 129, 130] or learning [119, 137, 136].

The present paper, close in spirit to [16, 84, 111] and [169] (in their so-called offline case) considers a parametric framework in a mean-field regime over a fixed time horizon. We take a deeper look at the asymptotic structure of the associated statistical experiment, in the sense of local asymptotic normality or LAN, in order to derive strong results for the maximum likelihood, both in asymptotic distribution and in an asymptotic minimax sense (up to constants) for various loss functions. For simplicity, we keep-up with continuous observations, but we briefly explain how

to move to a discrete data setting. Also, we look for simple and explicit criteria that enable us to verify identifiability and non-degeneracy of the model. This is a non-trivial issue in the context of nonlinear McKean-Vlasov models that is usually a bit overlooked in the literature.

1.2 Setting

We have a parameter of interest ϑ lying in a compact set $\Theta \subset \mathbb{R}^p$ (with non empty interior), for some fixed $p \geq 1$. For some fixed time horizon $T > 0$, we continuously observe a stochastic system of N interacting particles

$$X^{(N)} = (X_t^1, \dots, X_t^N)_{t \in [0, T]}, \quad (1.1)$$

evolving in an Euclidean ambient space $\mathbb{R}^d$, that solves

$$\begin{cases} dX_t^i = b(\vartheta; t, X_t^i, \mu_t^{(N)})dt + \sigma(t, X_t^i)dB_t^i, & 1 \leq i \leq N, t \in [0, T], \\ \mathcal{L}(X_0^1, \dots, X_0^N) = \mu_0^{\otimes N}, \end{cases} \quad (1.2)$$

where $\mu_t^{(N)} = N^{-1} \sum_{i=1}^N \delta_{X_t^i}$ is the empirical measure of the system. The $(B_t^i)_{t \in [0, T]}$ are independent $\mathbb{R}^d$ -valued Brownian motions. The initial condition μ_0 , the drift b and the diffusion coefficient σ are at least sufficiently regular so that

$$\mu^{(N)} = (\mu_t^{(N)})_{t \in [0, T]} \rightarrow \mu = (\mu_t)_{t \in [0, T]}$$

in distribution as $N \rightarrow \infty$, where μ is a family of probability measures that solves (in a weak sense) the parabolic nonlinear equation

$$\begin{cases} \partial_t \mu + \operatorname{div}(b(\vartheta; \cdot, \mu)\mu) = \frac{1}{2} \sum_{k, k'=1}^d \partial_{kk'}^2 (c_{kk'} \mu), & t \in [0, T], \\ \mu_{t=0} = \mu_0, \end{cases} \quad (1.3)$$

with $c = \sigma\sigma^\top$. We will write $\mu^\vartheta = (\mu_t^\vartheta)_{t \in [0, T]}$ to emphasise the dependence in ϑ . In this context, we are interested in estimating from data (1.1) the parameter $\vartheta \in \Theta$ of the function $(\vartheta; t, x, \nu) \mapsto b(\vartheta; t, x, \nu) \in \mathbb{R}^d$. Asymptotics are taken as $N \rightarrow \infty$.

A particular case of interest that covers many examples is when the dependence in the measure variable for b is linear : we then have

$$b(\vartheta; t, X_t^i, \mu_t^{(N)}) = \int_{\mathbb{R}^d} \tilde{b}(\vartheta; X_t^i, y) \mu_t^{(N)}(dy) = N^{-1} \sum_{j=1}^N \tilde{b}(\vartheta; X_t^i, X_t^j), \quad (1.4)$$

for some function $\tilde{b} : \Theta \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$. A typical form is $\tilde{b}(\vartheta; t, x, y) = G_\vartheta(x) + F_\vartheta(x - y)$ where $G_\vartheta, F_\vartheta : \mathbb{R}^d \rightarrow \mathbb{R}^d$ play the role of a common external force to the system and an interaction force respectively.

1.3 Results and organisation of the paper

In Section 2, we rigorously construct the (sequence of) statistical experiment(s) generated by the observation (1.1) under the dynamics (1.2) that we denote $(\mathcal{E}^N)_{N \geq 1}$. It is well defined and regular in the classical sense of Ibragimov and Hasminski [98] under strong integrability of the initial condition μ_0 and standard smoothness assumptions on the drift $\vartheta \mapsto b(\vartheta; \cdot)$ and the diffusion

matrix $c = \sigma\sigma^\top$, see Assumptions 1.1, 1.2, 1.3 and 1.4 and Proposition 1.6. The deep study of the asymptotic identifiability of $\mathcal{E}^N$ and the asymptotic non-degeneracy of its information matrix $\mathbb{I}_{\mathcal{E}^N}(\vartheta)$ is simplified via the two following accompanying experiments. The first one is $\mathcal{H}^{\otimes N}$, where $\mathcal{H}$ is generated by the continuous observation of a solution to the McKean-Vlasov equation

$$\begin{cases} dX_t = b(\vartheta; t, X_t, \mu_t^\vartheta)dt + \sigma(t, X_t)dB_t, & t \in [0, T], \\ \mathcal{L}(X_0) = \mu_0, \end{cases}$$

for a standard Brownian motion $(B_t)_{t \in [0, T]}$ on $\mathbb{R}^d$ and where μ_t^ϑ is the marginal distribution of the solution at time t . The second one is the family of experiments $\mathcal{G}_{\vartheta_0}^{\otimes N}$, $\vartheta_0 \in \Theta$, where for every $\vartheta_0 \in \Theta$ we denote by $\mathcal{G}_{\vartheta_0}$ the experiment generated by the continuous observation of a solution to the linear (in the McKean sense) equation

$$\begin{cases} dX_t = b(\vartheta; t, X_t, \mu_t^{\vartheta_0})dt + \sigma(t, X_t)dB_t, & t \in [0, T], \\ \mathcal{L}(X_0) = \mu_0, \end{cases}$$

for a standard Brownian motion $(B_t)_{t \in [0, T]}$ on $\mathbb{R}^d$. In particular, in the case of representation 1.4 we have that $\mathcal{E}^N$ and $\mathcal{H}^{\otimes N}$ do not separate asymptotically by a simple entropy argument, see Proposition 1.10. We also have the convergence of the normalized Fisher information matrices :

$$N^{-1}\mathbb{I}_{\mathcal{E}^N}(\vartheta) \rightarrow \mathbb{I}(\vartheta)$$

to some kind of a *proxy Fisher information matrix* $\mathbb{I}(\vartheta)$ in a mean-field limit $N \rightarrow \infty$, as established in Proposition 1.11. Actually, if, for every $\vartheta_0 \in \Theta$, $\mathbb{I}_{\mathcal{G}_{\vartheta_0}}(\vartheta)$ denotes the Fisher information matrix of the experiment $\mathcal{G}_{\vartheta_0}$ (if it exists), $\mathbb{I}(\vartheta)$ can be seen as the limit of $\mathbb{I}_{\mathcal{G}_{\vartheta_0}}(\vartheta)$ as $\vartheta_0 \rightarrow \vartheta$. Nonetheless, if we denote by $\mathbb{I}_{\mathcal{H}}(\vartheta)$ the Fisher information of the experiment $\mathcal{H}$ (if it exists), we do not have that $\mathbb{I}(\vartheta)$ and $\mathbb{I}_{\mathcal{H}}(\vartheta)$ coincide in general (yet, this would be of course the case in the "degenerate" case where b does not depend on the measure parameter). Despite this fact, this approximation is still the gateway to obtain explicit identifiability and non-degeneracy criteria, as detailed in Section 2.4. In particular, under additional regularity assumptions, we obtain a quite simple criterion for $\mathbb{I}(\vartheta)$ to be non-degenerate in Proposition 1.15, namely the property that one of the functions

$$x \mapsto \nabla_{\vartheta}(c^{-1/2}b)^j(\vartheta; 0, x, \mu_0)^\top z, \quad j = 1, \dots, d \quad (1.5)$$

is not identically vanishing, for every $z \in \mathbb{R}^p$ with $|z| = 1$, with $c^{-1/2}$ a square root of $c = \sigma\sigma^\top$. We use the notation $f = (f^j)_{1 \leq j \leq d}$ componentwise, the f^j being real-valued functions. In particular, (1.5) has the advantage to only relate to the initial condition μ_0 in the measure argument and not the whole $(\mu_t^\vartheta)_{t \in [0, T]}$ which is (almost) never explicit. Having a simple criterion to achieve the non-degeneracy of the Fisher information seems to have been a bit overlooked in the literature (where it is usually simply assumed to hold true) and our result is thus of interest for other studies.

In Section 3, we state the main results of the paper, Theorem 1.17, where we establish the LAN property : if we reparametrise the experiments via $\vartheta = \vartheta_0 + N^{-1/2}u$ locally around a fixed point ϑ_0 , with $u \in \mathbb{R}^p$ being now the unknown parameter, then $\mathcal{E}^N$ looks like a Gaussian shift : we observe

$$Y^N = u + \mathbb{I}(\vartheta_0)^{-1/2}\xi$$

where ξ is a standard Gaussian random vector in $\mathbb{R}^p$. This has important consequences in terms of existence and properties of optimal procedures : we have Hájek's convolution theorem (Corollary

1.18), namely for any estimator $\widehat{\vartheta}_N$,

$$\liminf_{N \rightarrow \infty} \sup_{|\vartheta' - \vartheta| \leq \delta} \mathbb{E}_{\mathbb{P}_{\vartheta'}^N} [w(N^{1/2} \mathbb{I}(\vartheta)^{1/2} (\widehat{\vartheta}_N - \vartheta'))] \geq \mathbb{E}[w(\xi)], \quad (1.6)$$

for small enough $\delta > 0$, where $\mathbb{P}_{\vartheta'}^N$ is the distribution of the data when the parameter is ϑ' and w is an arbitrary loss function satisfying some regularity properties. The bound (1.6) is achieved by the maximum likelihood estimator $\widehat{\vartheta}_N^{\text{MLE}}$ obtained by maximising the contrast

$$\vartheta \mapsto \ell^N(\vartheta; X^{(N)}) = \sum_{i=1}^N \int_0^T \left((c^{-1}b)(\vartheta; t, X_t^i, \mu_t^{(N)})^\top dX_t^i - \frac{1}{2} |(c^{-1/2}b)(\vartheta; t, X_t^i, \mu_t^{(N)})|^2 dt \right). \quad (1.7)$$

This implies in particular the convergence

$$\sqrt{N}(\widehat{\vartheta}_N^{\text{MLE}} - \vartheta) \rightarrow \mathcal{N}(0, \mathbb{I}(\vartheta)^{-1}) \quad (1.8)$$

in distribution. Moreover, we have in Theorem 1.19 the minimax asymptotic optimality of $\widehat{\vartheta}_N^{\text{MLE}}$, in the sense that

$$\mathcal{R}_w^N(\widehat{\vartheta}_N^{\text{MLE}}; \Theta) = \inf_{\widehat{\vartheta}_N} \mathcal{R}_w^N(\widehat{\vartheta}_N; \Theta)(1 + o(1))$$

where $\mathcal{R}_w^N(\widehat{\vartheta}_N; \Theta) = \sup_{\vartheta \in \Theta} \mathbb{E}_{\mathbb{P}_\vartheta^N} [w(N^{1/2} \mathbb{I}(\vartheta)^{1/2} (\widehat{\vartheta}_N - \vartheta))]$ is the classical minimax risk. Thus the LAN property enables us to obtain considerably stronger results than simply (1.8). In Section 4, we investigate several non-trivial examples that generalise the results of [111], and where our identifiability and non-degeneracy criteria easily apply. We treat in particular the case of a kinetic mean-field double layer potential that may serve as a representative model for swarming models, see in particular [19] and the references therein. The proofs are delayed until Sections 5 and 6, with an appendix (Section 7) that contains useful technical results.

In practice, maximising the function (1.7) is not feasible, since only discrete data are available. It is then reasonable to replace the ideal observation (1.1) by the more realistic

$$X^{(N,m)} = (X_t^1, \dots, X_t^N)_{t \in \{t_0^m, \dots, t_m^m\}},$$

where $(0 = t_0^m < t_1^m < \dots < t_m^m = T)$ is a subdivision of $[0, T]$ with mesh

$$\max_{1 \leq j \leq m} (t_j^m - t_{j-1}^m) \leq Cm^{-1}.$$

We thus have $(m+1) \times N$ data with values in $\mathbb{R}^d$. We may then replace (1.7) by

$$\begin{aligned} \vartheta \mapsto N^{-1} \sum_{i=1}^N \sum_{j=0}^m & \left((c^{-1}b)(\vartheta; t_j^m, X_{t_{j-1}^m}^i, \mu_{t_{j-1}^m}^{(N)})^\top (X_{t_j^m}^i - X_{t_{j-1}^m}^i) \right. \\ & \left. - \frac{1}{2} |(c^{-1/2}b)(\vartheta; t_{j-1}^m, X_{t_{j-1}^m}^i, \mu_{t_{j-1}^m}^{(N)})|^2 (t_j^m - t_{j-1}^m) \right). \end{aligned}$$

Assuming the function $(t, x) \mapsto (c^{-1/2}b)(\vartheta; t, x, \mu_t^{(N)})$ to be smooth, we may safely expect the discrete approximation to be close to its continuous counterpart up to an additional error of order $m^{-1/2}$, by standard high-frequency discretisation techniques, see the textbooks of Jacod and co-authors [2, 101, 102]. In particular, if $m \gg N$, the same results as for continuous observations

are likely to hold true. While this paper was under review, new results about contrast estimation from discrete data in the same setting (actually a bit more general with respect to the form of the diffusion coefficient) were proposed by Amorino, Heidari, Pilipauskaitė and Podolskij, see [3]. It is shown in particular that contrast estimators based on discrete data achieve optimality in the sense of (1.8), up to an $m^{-1/2}$ error term.

2 Construction and properties of the statistical model

2.1 Notation

The dimension $d \geq 1$ of the state space $\mathbb{R}^d$ and the dimension $p \geq 1$ of the parameter space Θ as well as the time horizon $T > 0$ are fixed once for all. We write $|\cdot|$ for the Euclidean distance on $\mathbb{R}^q$ ($q = p, d$ or any other integer, depending on the context) or for a matrix norm on $\mathbb{R}^p \otimes \mathbb{R}^p$ fixed throughout.

We consider functions that are mappings defined on products of metric spaces (typically $\Theta \times [0, T] \times \mathbb{R}^d \times \mathcal{P}_1$ or subsets of these) with values in $\mathbb{R}$ or $\mathbb{R}^d$. Here, $\mathcal{P}_1$ denotes the set of probability measures on $\mathbb{R}^d$ with a first moment, endowed with the Wasserstein 1-metric

$$\mathcal{W}_1(\mu, \nu) = \inf_{m \in \Gamma(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y| m(dx, dy) = \sup_{|\phi|_{\text{Lip}} \leq 1} \int_{\mathbb{R}^d} \phi d(\mu - \nu),$$

where $\Gamma(\mu, \nu)$ denotes the set of probability measures on the product space $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν . For a probability measure μ on $\mathbb{R}^d$, we also set

$$\mathbf{m}_r(\mu) = \int_{\mathbb{R}^d} |y|^r \mu(dy)$$

for its moment of order $r \geq 1$ and we say that $\mu \in \mathcal{P}_r$ if $\mathbf{m}_r(\mu)$ is finite. All the functions in the paper are implicitly measurable with respect to the Borel-sigma field induced by the product topology. A $\mathbb{R}^d$ -valued function f is written componentwise as $f = (f^k)_{1 \leq k \leq d}$ where the f^k are real-valued. We denote by ∂_{ϑ_k} , ∇_{ϑ} , $\partial_{\vartheta_k \vartheta_l}^2$ respectively the partial derivative of a function with respect to the k -th component ϑ_k , the gradient of a real-valued function with respect to ϑ , the second order partial derivative of a function with respect to the k -th and l -th components ϑ_k, ϑ_l .

Finally, we repeatedly use the notation C for a positive number that does not depend on N , nor ϑ , that may vary from line to line and that we call a constant, although it usually depends on some other (fixed) quantities of the model. In most cases, it is explicitly computable.

2.2 Model assumptions

Well-posedness of the model and its associated statistical experiment

We work under the following strong integrability property for the initial condition μ_0 .

Assumption 1.1. *For every $r \geq 1$, we have $\mu_0 \in \mathcal{P}_r$.*

As for the diffusion matrix $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$, we make the following strong ellipticity and Lipschitz smoothness assumption.

Assumption 1.2. *The diffusion matrix σ is measurable and for some $C \geq 0$, we have*

$$|\sigma(t, x') - \sigma(t, x)| \leq C|x' - x|.$$

Moreover, $c = \sigma\sigma^\top$ is such that $\sigma_-^2|y|^2 \leq (c(t, x)y)^\top y \leq \sigma_+^2|y|^2$ for some $\sigma_\pm > 0$.

As for the drift part $b : \Theta \times [0, T] \times \mathbb{R}^d \times \mathcal{P}_1 \rightarrow \mathbb{R}^d$, we work under usual Lipschitz smoothness assumptions.

Assumption 1.3. *The drift b is measurable and for some $C \geq 0$, we have*

$$\sup_{t \in [0, T], \vartheta \in \Theta} |b(\vartheta; t, x', \nu') - b(\vartheta; t, x, \nu)| \leq C(|x' - x| + \mathcal{W}_1(\nu', \nu)),$$

and

$$b_0 = \sup_{t \in [0, T], \vartheta \in \Theta} |b(\vartheta; t, 0, \delta_0)| < \infty.$$

We let $|b|_{\text{Lip}}$ denote the smallest $C \geq 0$ for which Assumption 1.3 holds.

Assumptions 1.1, 1.2, 1.3 together are sufficient to guarantee the well-posedness of the statistical model : there exists a unique weak solution to (1.2) for every $\vartheta \in \Theta$ hence the data $X^{(N)}$ of (1.1) is well-defined. More precisely, we let $\mathcal{C}^N = \mathcal{C}([0, T], (\mathbb{R}^d)^N)$ denote the space of continuous functions on $(\mathbb{R}^d)^N$, equipped with the natural filtration $(\mathcal{F}_t^N)_{t \in [0, T]}$ induced by the canonical mappings

$$X_t^{(N)}(\omega) = (X_t^1(\omega), \dots, X_t^N(\omega)) = \omega_t.$$

For $\mu_0 \in \mathcal{P}_1$ and $\vartheta \in \Theta$, the probability $\mathbb{P}_\vartheta^N$ on $(\mathcal{C}^N, \mathcal{F}^N)$ under which the canonical process $X^{(N)} = (X_t^{(N)})_{t \in [0, T]}$ is a solution of (1.2) for the initial condition $\mu_0^{\otimes N}$ is uniquely defined under Assumptions 1.1, 1.2 and 1.3. Recommended reference (that covers our set of assumptions) is the textbook by Carmona and Delarue [37] or the lectures notes of Lacker [115]. Moreover, for every $\vartheta \in \Theta$, the parabolic nonlinear equation (1.3) has a unique probability solution $\mu = (\mu_t^\vartheta)_{t \in [0, T]}$ and we have the convergence $\mu_t^{(N)} \rightarrow \mu_t^\vartheta$ under $\mathbb{P}_\vartheta^N$ in distribution, for every $\vartheta \in \Theta$.

We thus study under Assumptions 1.1, 1.2, 1.3 the (sequence of) statistical experiment(s) generated by the observation (1.1) under the dynamics (1.2) and that we realise as

$$(\mathcal{E}^N)_{N \geq 1} = \left(\mathcal{C}^N, \mathcal{F}^N, (\mathbb{P}_\vartheta^N, \vartheta \in \Theta) \right)_{N \geq 1}.$$

Note that at that stage, we do not impose any identifiability assumption *i.e.* we do not assume that the mapping $\vartheta \mapsto \mathbb{P}_\vartheta^N$ is one-to-one. We will discuss that matter together with the non-degeneracy of the model later in Section 2.4.

Regularity of the experiment $\mathcal{E}^N$

In order to study the regularity of the model, we need specific smoothness properties for the function $\vartheta \mapsto b(\vartheta, \cdot)$.

Assumption 1.4. *There exist $r_1, r_2 \geq 1$ and $C > 0$ such that for every point ϑ in the interior of Θ and every $(t, x, \nu) \in (0, T) \times \mathbb{R}^d \times \mathcal{P}_{\max(r_1, r_2)}$, the function $\vartheta \mapsto b(\vartheta; t, x, \nu)$ is twice differentiable and for every $1 \leq \ell, \ell' \leq p$,*

$$\sup_{t \in [0, T]} (|\partial_{\vartheta_\ell} b(\vartheta; t, x, \nu)| + |\partial_{\vartheta_\ell \vartheta_{\ell'}}^2 b(\vartheta; t, x, \nu)|) \leq C(1 + |x|^{r_1} + \mathbf{m}_{r_2}(\nu)),$$

$$\sup_{t \in [0, T]} |\partial_{\vartheta_\ell} b(\vartheta; t, x', \nu') - \partial_{\vartheta_\ell} b(\vartheta; t, x, \nu)| \leq C(|x' - x| + \mathcal{W}_1(\nu', \nu)).$$

The smoothness properties of the map $\vartheta \mapsto b(\vartheta; \cdot)$ granted by Assumption 1.4 enables us to explore further the regularity of the experiment $\mathcal{E}^N$. First, note that we have a log-likelihood by setting

$$\ell^N(\vartheta; X^{(N)}) = \sum_{i=1}^N \int_0^T (c^{-1}b)(\vartheta; t, X_t^i, \mu_t^{(N)})^\top dX_t^i - \frac{1}{2} \sum_{i=1}^N \int_0^T |(c^{-1/2}b)(\vartheta; t, X_t^i, \mu_t^{(N)})|^2 dt, \quad (1.9)$$

where $c^{-1/2}$ is fixed once for all. Indeed, by Girsanov's theorem again, the laws $\mathbb{P}_\vartheta^N$ are all absolutely continuous w.r.t. $\mathbb{W}^N$, defined as the unique probability on $(\mathcal{C}^N, \mathcal{F}^N)$ under which the processes

$$\left(\int_0^t c^{-1/2}(s, X_s^i) dX_s^i \right)_{t \in [0, T]}, \quad 1 \leq i \leq N$$

are independent standard Brownian motions on $\mathbb{R}^d$, together with $\mathcal{L}(X_0^1, \dots, X_0^N) = \mu_0^{\otimes N}$. In turn, for every $\vartheta \in \Theta$,

$$\frac{d\mathbb{P}_\vartheta^N}{d\mathbb{W}^N}(X^{(N)}) = \exp(\ell^N(\vartheta; X^{(N)}))$$

holds $\mathbb{W}^N$ -almost-surely. We further write $\mathcal{L}^N(\vartheta; X^{(N)}) = \exp(\ell^N(\vartheta; X^{(N)}))$ for the likelihood process, indexed by the parameter $\vartheta \in \Theta$. We recall one possible classical definition of a regular statistical experiment, following [98].

Definition 1.5. *The dominated (sequence of) experiment(s) $(\mathcal{E}^N)_{N \geq 1}$ is regular if*

- (i) $\vartheta \mapsto \mathcal{L}^N(\vartheta; X^{(N)})$ is differentiable for every ϑ in (the interior of) Θ , $\mathbb{W}^N$ -almost surely,
- (ii) $\vartheta \mapsto \nabla_\vartheta \mathcal{L}^N(\vartheta; X^{(N)})$ is continuous in quadratic $\mathbb{W}^N$ -mean, for every ϑ in (the interior of) Θ ,
- (iii) we have finite Fisher information

$$\mathbb{E}_{\mathbb{P}_\vartheta^N} [|\nabla_\vartheta \ell^N(\vartheta; X^{(N)})|^2] < \infty$$

for every ϑ in (the interior of) Θ .

Proposition 1.6. *Under Assumptions 1.1, 1.2, 1.3 and 1.4 the (sequence of) experiment(s) $(\mathcal{E}^N)_{N \geq 1}$ is regular.*

(Sketch of) Proof. By exchanging the order of the differentiation with respect to ϑ and the stochastic integral we have

$$\partial_{\vartheta_k} \ell^N(\vartheta; X^{(N)}) = \sum_{i=1}^N \int_0^T \partial_{\vartheta_k} (c^{-1}b)(\vartheta; t, X_t^i, \mu_t^{(N)})^\top dX_t^i$$

$$- \sum_{i=1}^N \int_0^T \partial_{\vartheta_k}(c^{-1/2}b)(\vartheta; t, X_t^i, \mu_t^{(N)})^\top (c^{-1/2}b)(\vartheta; t, X_t^i, \mu_t^{(N)}) dt.$$

We obtain the representation

$$\partial_{\vartheta_k} \ell^N(\vartheta; X^{(N)}) = \sum_{i=1}^N \int_0^T \partial_{\vartheta_k}(c^{-1/2}b)(\vartheta; t, X_t^i, \mu_t^{(N)})^\top dB_t^{i,N,\vartheta}, \quad (1.10)$$

where the

$$(B_t^{i,N,\vartheta})_{t \in [0,T]} = \left(\int_0^t c^{-1/2}(s, X_s^i)(dX_s^i - b(\vartheta; s, X_s^i, \mu_s^{(N)})ds) \right)_{t \in [0,T]}, \quad 1 \leq i \leq N$$

are independent Brownian motions on $\mathbb{R}^d$ under $\mathbb{P}_\vartheta^N$. The properties (i), (ii) and (iii) are then a simple consequence of Assumption 1.4 together with the following moment bound,

Lemma 1.7. *Under Assumptions 1.1, 1.2, 1.3, for every $r \geq 1$, we have*

$$\sup_{\vartheta \in \Theta, t \in [0,T], N \geq 1} \mathbb{E}_{\mathbb{P}_\vartheta^N} [|X_t^i|^r] < \infty.$$

Note that $\mathbb{E}_{\mathbb{P}_\vartheta^N} [|X_t^i|^r]$ does not depend on $1 \leq i \leq N$. The proof of Lemma 1.7 is given in Appendix 7.1. $\square$

Finally, we have a notion of Fisher information matrix by setting

$$\mathbb{I}_{\mathcal{E}^N}(\vartheta) = \mathbb{E}_{\mathbb{P}_\vartheta^N} [\nabla_\vartheta \ell^N(\vartheta; X^{(N)}) \nabla_\vartheta \ell^N(\vartheta; X^{(N)})^\top].$$

Thanks to (1.10), we also have

$$\mathbb{I}_{\mathcal{E}^N}(\vartheta) = \left(\sum_{i=1}^N \mathbb{E}_{\mathbb{P}_\vartheta^N} \left[\int_0^T \partial_{\vartheta_\ell}(c^{-1/2}b)(\vartheta; t, X_t^i, \mu_t^{(N)}) \partial_{\vartheta_{\ell'}}(c^{-1/2}b)(\vartheta; t, X_t^i, \mu_t^{(N)})^\top dt \right] \right)_{1 \leq \ell, \ell' \leq p}. \quad (1.11)$$

2.3 The companion McKean-Vlasov product experiment and its *linear* counterpart

We let $\mathcal{C} = \mathcal{C}([0, T], \mathbb{R}^d)$ denote the space of continuous functions on $\mathbb{R}^d$, equipped with the natural filtration $(\mathcal{F}_t)_{0 \leq t \leq T}$ induced by the canonical mapping $X_t(\omega) = \omega_t$. For every $\vartheta \in \Theta$, we let $\bar{\mathbb{P}}_\vartheta$ denote the unique law under which the process

$$(B_t^\vartheta)_{t \in [0,T]} = \left(\int_0^t c^{-1/2}(s, X_s)(dX_s - b(\vartheta; s, X_s, \mu_s^\vartheta)ds) \right)_{t \in [0,T]}$$

is a standard Brownian motion on $\mathbb{R}^d$, appended with the condition $\mathcal{L}(X_0) = \mu_0$, and $\mu^\vartheta = (\mu_t^\vartheta)_{t \in [0,T]}$ is a probability solution of (1.3). The family $(\bar{\mathbb{P}}_\vartheta)_{\vartheta \in \Theta}$ is well-defined under Assumptions 1.1, 1.2, 1.3. In particular, the canonical process X on $(\mathcal{C}, \mathcal{F}_T)$ is a solution to the McKean-Vlasov equation

$$\begin{cases} dX_t = b(\vartheta; t, X_t, \mu_t^\vartheta)dt + \sigma(t, X_t)dB_t^\vartheta, & t \in [0, T], \\ \mathcal{L}(X_0) = \mu_0. \end{cases} \quad (1.12)$$

The following result is the counterpart of Lemma 1.7. Note in particular that the marginals of $\bar{\mathbb{P}}_\vartheta$ coincide with the solution $\mu^\vartheta = (\mu_t^\vartheta)_{t \in [0,T]}$ of the Fokker-Planck equation (1.3).

Lemma 1.8. *Under Assumptions 1.1, 1.2, 1.3, for every $r \geq 1$, we have*

$$\sup_{\vartheta \in \Theta, t \in [0, T]} \mathbb{E}_{\mathbb{P}_\vartheta} [|X_t|^r] = \sup_{\vartheta \in \Theta, t \in [0, T]} \int_{\mathbb{R}^d} |x|^r \mu_t^\vartheta(dx) < \infty.$$

The proof is given in Section 7.2. We also have the following smoothness property in the parameter ϑ , proof of which is delayed until Section 6.1.

Proposition 1.9. *Under Assumption 1.1, 1.2, 1.3 and 1.4, the mapping $\vartheta \mapsto \mu_t^\vartheta$ is Lipschitz continuous in the Wasserstein-1 metric $\mathcal{W}_1$, uniformly in $t \in [0, T]$.*

We next consider the first one of our two limit experiments

$$\mathcal{H} = (\mathcal{C}, \mathcal{F}_T, (\bar{\mathbb{P}}_\vartheta)_{\vartheta \in \Theta})$$

and its N -fold counterpart

$$\mathcal{H}^{\otimes N} = (\mathcal{C}^N, \mathcal{F}_T^N, (\bar{\mathbb{P}}_\vartheta^{\otimes N})_{\vartheta \in \Theta})$$

that serves as an approximation in the mean-field limit for the experiment $\mathcal{E}^N$. Inspired by classical propagation of chaos techniques (see in particular [116]), we can easily show that the measures $\mathbb{P}_\vartheta^N$ and $\bar{\mathbb{P}}_\vartheta^{\otimes N}$ are indistinguishable when the drift is of the form

$$b(\vartheta; t, x, \nu) = \int_{\mathbb{R}^d} \tilde{b}(\vartheta; t, x, y) \nu(dy), \quad (1.13)$$

for some kernel $\tilde{b}(\vartheta; \cdot) : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that

$$\sup_{t \in [0, T], \vartheta \in \Theta} |\tilde{b}(\vartheta; t; x, y)| \leq C(1 + |x|^{r_1} + |y|^{r_2}) \quad (1.14)$$

for some $r_1, r_2 \geq 1$, a situation that covers most of our examples, see Section 4 below. More precisely, we have the following

Proposition 1.10. *Under Assumptions 1.1, 1.2, 1.3, if b has moreover the form (1.13)-(1.14), we have*

$$\limsup_{N \rightarrow \infty} \sup_{\vartheta \in \Theta} \mathbb{E}_{\bar{\mathbb{P}}_\vartheta^{\otimes N}} \left[\log \frac{d\bar{\mathbb{P}}_\vartheta^{\otimes N}}{d\mathbb{P}_\vartheta^N} \right] < \infty. \quad (1.15)$$

In particular, if

$$\sup_{\vartheta \in \Theta} \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} |\tilde{b}(\vartheta; t, x, y)|^2 (\mu_t^\vartheta \otimes \mu_t^\vartheta)(dx, dy) dt < 4\sigma_-^2,$$

then

$$\limsup_{N \rightarrow \infty} \sup_{\vartheta \in \Theta} \|\mathbb{P}_\vartheta^N - \bar{\mathbb{P}}_\vartheta^{\otimes N}\|_{TV} < 1, \quad (1.16)$$

where $\|\cdot\|_{TV}$ denotes total variation distance and σ_- is defined in Assumption 1.2).

The proof is given in Section 6.2. Some remarks are in order : **1)** The estimate (1.15) tells us that it is impossible to statistically discriminate between $\mathbb{P}_\vartheta^N$ and $\bar{\mathbb{P}}_\vartheta^{\otimes N}$ asymptotically. More precisely, inequality (1.16) shows in particular that provided $\tilde{b}$ is not too big or T not too large, then there exists no test of the null $H_0 : \mathbb{P}_\vartheta^N = \bar{\mathbb{P}}_\vartheta^{\otimes N}$ against the alternative $H_1 : \mathbb{P}_\vartheta^N \neq \bar{\mathbb{P}}_\vartheta^{\otimes N}$ with

asymptotically arbitrarily small first and second kind error in the limit $N \rightarrow \infty$ in the following sense : for fixed ϑ , any test which rejects the null on the event $\mathcal{R}_N$ has sum of the first and second kind errors equal to $\bar{\mathbb{P}}_{\vartheta}^{\otimes N}(\mathcal{R}_N) + \mathbb{P}_{\vartheta}^N(\mathcal{R}_N^c) \geq 1 - \|\mathbb{P}_{\vartheta}^N - \bar{\mathbb{P}}_{\vartheta}^{\otimes N}\|_{TV}$, which is asymptotically bounded away from zero by (1.16). **2)** Finally, (1.15) may hold in wider generality when the dependence in the measure variable in the drift is nonlinear, as soon as we have some differentiability in the following sense : there exists $\partial_{\nu}b(\vartheta; t, x, \cdot) : \mathbb{R}^d \times \mathcal{P}_1 \rightarrow \mathbb{R}^d$ such that

$$b(\vartheta; t, x, \nu) - b(\vartheta; t, x, \nu') = \int_0^1 \partial_{\nu}b(\vartheta; t, x, y, \lambda\nu + (1-\lambda)\nu')(y - \nu') dy$$

for every $\nu, \nu' \in \mathcal{P}_1$ and $\partial_{\nu}b(\vartheta; t, x, \cdot)$ satisfies additional smoothness properties. Iterating the operator ∂_{ν} , if $\partial_{\nu}^k b(\vartheta; t, x, \cdot) : (\mathbb{R}^d)^k \times \mathcal{P}_1 \rightarrow \mathbb{R}^d$ exists and satisfies some smoothness and integrability properties, we may expect (1.15) to hold as soon as $k \geq d/2$. We refer to Assumption 4 and Proposition 19 of [65] where this approach is developed.

We also have a log-likelihood in the experiment $\mathcal{H}^{\otimes N}$ by setting

$$\bar{\ell}^N(\vartheta; X^{(N)}) = \sum_{i=1}^N \int_0^T (c^{-1}b)(\vartheta; t, X_t^i, \mu_t^{\vartheta})^{\top} dX_t^i - \frac{1}{2} \sum_{i=1}^N \int_0^T |(c^{-1/2}b)(\vartheta; t, X_t^i, \mu_t^{\vartheta})|^2 dt. \quad (1.17)$$

This is the same argument as before : the laws $\bar{\mathbb{P}}_{\vartheta}^{\otimes N}$ are all absolutely continuous w.r.t. $\mathbb{W}^N$, and for every $\vartheta \in \Theta$,

$$\frac{d\bar{\mathbb{P}}_{\vartheta}^{\otimes N}}{d\mathbb{W}^N}(X^{(N)}) = \exp(\bar{\ell}^N(\vartheta; X^{(N)})) \quad (1.18)$$

holds $\mathbb{W}^N$ -almost-surely.

Finally under Assumptions 1.1, 1.2, 1.3 and 1.4, the (sequence of) experiment(s) $\mathcal{H}^{\otimes N}$ is also a regular model and its (normalised) Fisher information $\mathbb{I}_{\mathcal{H}}(\vartheta) = N^{-1}\mathbb{I}_{\mathcal{H}^{\otimes N}}(\vartheta)$ is given by

$$\begin{aligned} \mathbb{I}_{\mathcal{H}}(\vartheta) &= N^{-1} \mathbb{E}_{\bar{\mathbb{P}}_{\vartheta}^{\otimes N}} [\nabla_{\vartheta} \bar{\ell}^N(\vartheta; X^{(N)}) \nabla_{\vartheta} \bar{\ell}^N(\vartheta; X^{(N)})^{\top}] \\ &= \sum_{j=1}^d \int_0^T \int_{\mathbb{R}^d} \nabla_{\vartheta} [(c^{-1/2}b)^j(\vartheta; t, x, \mu_t^{\vartheta})] \nabla_{\vartheta} [(c^{-1/2}b)^j(\vartheta; t, x, \mu_t^{\vartheta})]^{\top} \mu_t^{\vartheta}(dx) dt \end{aligned}$$

as soon as $\vartheta \mapsto b(\vartheta; t, x, \mu_t^{\vartheta})$ is smooth enough, with appropriate integrability properties of its derivative w.r.t. ϑ .

We now consider our second limit experiment, or actually our family of limit experiments. For every $\vartheta_0, \vartheta \in \Theta$, we let $\bar{\mathbb{P}}_{\vartheta, \vartheta_0}$ denote the unique law under which the process

$$(B_t^{\vartheta, \vartheta_0})_{t \in [0, T]} = \left(\int_0^t c^{-1/2}(s, X_s) (dX_s - b(\vartheta; s, X_s, \mu_s^{\vartheta_0}) ds) \right)_{t \in [0, T]}$$

is a standard Brownian motion on $\mathbb{R}^d$, appended with the condition $\mathcal{L}(X_0) = \mu_0$. For every $\vartheta_0 \in \Theta$, the family $(\bar{\mathbb{P}}_{\vartheta, \vartheta_0})_{\vartheta \in \Theta}$ is well-defined under Assumptions 1.1, 1.2, 1.3. In particular, the canonical process X on $(\mathcal{C}, \mathcal{F}_T)$ is solution to the linear (in the McKean sense) equation

$$\begin{cases} dX_t = b(\vartheta; t, X_t, \mu_t^{\vartheta_0}) dt + \sigma(t, X_t) dB_t^{\vartheta, \vartheta_0}, & t \in [0, T], \\ \mathcal{L}(X_0) = \mu_0. \end{cases} \quad (1.19)$$

Our second limit experiment is then the family of experiments $\mathcal{G}_{\vartheta_0}, \vartheta_0 \in \Theta$, where for every fixed $\vartheta_0 \in \Theta$,

$$\mathcal{G}_{\vartheta_0} = (\mathcal{C}, \mathcal{F}_T, (\bar{\mathbb{P}}_{\vartheta, \vartheta_0})_{\vartheta \in \Theta})$$

and N -fold counterpart family of experiments

$$\mathcal{G}_{\vartheta_0}^{\otimes N} = (\mathcal{C}^N, \mathcal{F}_T^N, (\bar{\mathbb{P}}_{\vartheta, \vartheta_0}^{\otimes N})_{\vartheta \in \Theta}), \vartheta_0 \in \Theta.$$

We also have, for every $\vartheta_0 \in \Theta$, a log-likelihood in the experiment $\mathcal{G}_{\vartheta_0}^{\otimes N}$ by setting

$$\bar{\ell}_{\vartheta_0}^N(\vartheta; X^{(N)}) = \sum_{i=1}^N \int_0^T (c^{-1}b)(\vartheta; t, X_t^i, \mu_t^{\vartheta_0})^\top dX_t^i - \frac{1}{2} \sum_{i=1}^N \int_0^T |(c^{-1/2}b)(\vartheta; t, X_t^i, \mu_t^{\vartheta_0})|^2 dt. \quad (1.20)$$

This is the same argument as before : the laws $\bar{\mathbb{P}}_{\vartheta, \vartheta_0}^{\otimes N}$ are all absolutely continuous w.r.t. $\mathbb{W}^N$, and for every $\vartheta \in \Theta$,

$$\frac{d\bar{\mathbb{P}}_{\vartheta, \vartheta_0}^{\otimes N}}{d\mathbb{W}^N}(X^{(N)}) = \exp(\bar{\ell}_{\vartheta_0}^N(\vartheta; X^{(N)})) \quad (1.21)$$

holds $\mathbb{W}^N$ -almost-surely.

Finally under Assumptions 1.1, 1.2, 1.3 and 1.4, the (sequence of) experiment(s) $\mathcal{G}_{\vartheta_0}^{\otimes N}$ is also a regular model and its (normalised) Fisher information $\mathbb{I}_{\mathcal{G}_{\vartheta_0}}(\vartheta) = N^{-1} \mathbb{I}_{\mathcal{G}_{\vartheta_0}^{\otimes N}}(\vartheta)$ is given by

$$\begin{aligned} \mathbb{I}_{\mathcal{G}_{\vartheta_0}}(\vartheta) &= N^{-1} \mathbb{E}_{\bar{\mathbb{P}}_{\vartheta, \vartheta_0}^{\otimes N}} [\nabla_{\vartheta} \bar{\ell}_{\vartheta_0}^N(\vartheta; X^{(N)}) \nabla_{\vartheta} \bar{\ell}_{\vartheta_0}^N(\vartheta; X^{(N)})^\top] \\ &= \sum_{j=1}^d \int_0^T \int_{\mathbb{R}^d} \nabla_{\vartheta} (c^{-1/2}b)^j(\vartheta; t, x, \mu_t^{\vartheta_0}) \nabla_{\vartheta} (c^{-1/2}b)^j(\vartheta; t, x, \mu_t^{\vartheta_0})^\top \mu_t^{\vartheta, \vartheta_0}(dx) dt, \end{aligned}$$

where $\mu_t^{\vartheta, \vartheta_0}$ denotes the law of X_t^1 under $\bar{\mathbb{P}}_{\vartheta, \vartheta_0}$. The mapping $(\vartheta_0, \vartheta) \mapsto \mathbb{I}_{\mathcal{G}_{\vartheta_0}}(\vartheta)$ is smooth, and we can then define, for every $\vartheta \in \Theta$,

$$\mathbb{I}(\vartheta) := \lim_{\vartheta_0 \rightarrow \vartheta} \mathbb{I}_{\mathcal{G}_{\vartheta_0}}(\vartheta)$$

which means that

$$\mathbb{I}(\vartheta) = \sum_{j=1}^d \int_0^T \int_{\mathbb{R}^d} \nabla_{\vartheta} (c^{-1/2}b)^j(\vartheta; t, x, \mu_t^{\vartheta}) \nabla_{\vartheta} (c^{-1/2}b)^j(\vartheta; t, x, \mu_t^{\vartheta})^\top \mu_t^{\vartheta}(dx) dt.$$

The mapping $\vartheta \mapsto \mathbb{I}(\vartheta)$ is of course itself smooth and appears as the (normalised) asymptotic information of $\mathcal{E}^N$:

Proposition 1.11. *Under Assumptions 1.1, 1.2, 1.3 and 1.4, the mapping $\vartheta \mapsto \mathbb{I}(\vartheta)$ is Lipschitz continuous. Moreover, for every ϑ in (the interior of) Θ , we have*

$$N^{-1} \mathbb{I}_{\mathcal{E}^N}(\vartheta) \rightarrow \mathbb{I}(\vartheta)$$

as $N \rightarrow \infty$, where $\mathbb{I}_{\mathcal{E}^N}(\vartheta)$ is the Fisher information matrix of the experiment $\mathcal{E}^N$ defined in (1.11) above.

The proof is given in Section 6.3. We remark that, in the case of the limit experiment $\mathcal{H}$, because of the fact that we must take a derivative w.r.t. ϑ in $b(\vartheta; t, x, \mu_t^{\vartheta})$, the Fisher information $\mathbb{I}_{\mathcal{H}}(\vartheta)$ differs from the limiting information $\mathbb{I}(\vartheta)$ of the experiment $\mathcal{E}^N$.

2.4 Identifiability and non-degeneracy of the Fisher information

Motivation

In the preceding section, we have built $\mathcal{E}^N$, $\mathcal{H}^{\otimes N}$ (equivalently $\mathcal{H}$), and the family $\mathcal{G}_{\vartheta_0}^{\otimes N}(\vartheta_0 \in \Theta)$ (equivalently $\mathcal{G}_{\vartheta_0}(\vartheta_0 \in \Theta)$) as possibly redundant, in the sense that the mappings $\vartheta \mapsto \mathbb{P}_{\vartheta}^N$, $\vartheta \mapsto \bar{\mathbb{P}}_{\vartheta}$, $\vartheta \mapsto \bar{\mathbb{P}}_{\vartheta, \vartheta_0}(\vartheta_0 \in \Theta)$ are not necessarily one-to-one on Θ . Having a well-posed parametrisation is required since we wish to have at least consistent estimators.

Asymptotic identifiability of $\mathcal{E}^N$ is somehow linked to the non-degeneracy of the (normalised) asymptotic information matrix $\mathbb{I}$. The experiment $\mathcal{E}^N$ is called *locally asymptotically identifiable* at ϑ if the mapping $\vartheta' \mapsto (\mathbb{P}_{\vartheta'}^N)_{N \geq 1}$ is injective in a neighbourhood of ϑ . We have the following classical result (that goes back at least to Cramer [50]) :

Proposition 1.12 (Theorem 1 in [167]). *Let ϑ in the interior of Θ . If $\mathbb{I}(\vartheta)$ has full rank, then $\mathcal{E}^N$ is locally asymptotically identifiable at ϑ .*

Unfortunately, there is no hope to obtain a global result that links the two notions unless in very specific cases, see Proposition 1.16 below. We next give *ad-hoc* assumptions that give sufficient (and independent) condition for both identifiability and non-degeneracy of the Fisher information.

An identifiability assumption

We first have a relatively weak assumption that guarantees global identifiability in $\mathcal{E}^N$, $\mathcal{H}$ and $\mathcal{G}_{\vartheta_0}(\vartheta_0 \in \Theta)$:

Assumption 1.13. (*Criterion for $\mathcal{E}^N$*) : For all $\vartheta \in \Theta$, for $\mathbb{P}_{\vartheta}^N$ -almost all ω , for all $\vartheta' \neq \vartheta$, the two sequences of functions $t \mapsto (b(\vartheta; t, X_t^i(\omega), \mu_t^{(N)}(\omega)))_{1 \leq i \leq N}$ and $t \mapsto (b(\vartheta'; t, X_t^i(\omega), \mu_t^{(N)}(\omega)))_{1 \leq i \leq N}$ are not dt -a.e. equal.

(*Criterion for $\mathcal{H}$*) : For all $\vartheta \in \Theta$, for $\bar{\mathbb{P}}_{\vartheta}$ -almost all ω , for all $\vartheta' \neq \vartheta$, the functions $t \mapsto b(\vartheta; t, X_t(\omega), \mu_t^{\vartheta})$ and $t \mapsto b(\vartheta'; t, X_t(\omega), \mu_t^{\vartheta'})$ are not dt -a.e. equal.

(*Criterion for $\mathcal{G}_{\vartheta_0}(\vartheta_0 \in \Theta)$*) : For all $\vartheta \in \Theta$, for $\bar{\mathbb{P}}_{\vartheta}^{\vartheta_0}$ -almost all ω , for all $\vartheta' \neq \vartheta$, the functions $t \mapsto b(\vartheta; t, X_t(\omega), \mu_t^{\vartheta_0})$ and $t \mapsto b(\vartheta'; t, X_t(\omega), \mu_t^{\vartheta_0})$ are not dt -a.e. equal.

Assumption 1.13 is relatively standard in the literature of statistics of random processes and minimal (see e.g. [79] in a somewhat analogous context). Indeed, let us check this for instance for $\mathcal{H}$: by (1.18), for two different parameters $\vartheta, \vartheta' \in \Theta$, the laws $\bar{\mathbb{P}}_{\vartheta}$ and $\bar{\mathbb{P}}_{\vartheta'}$ are absolutely continuous and

$$\begin{aligned} \log \frac{d\bar{\mathbb{P}}_{\vartheta'}}{d\bar{\mathbb{P}}_{\vartheta}}(X) &= \int_0^T ((c^{-1}b)(\vartheta'; s, X_s, \mu_s^{\vartheta'}) - (c^{-1}b)(\vartheta; s, X_s, \mu_s^{\vartheta}))^{\top} dX_s \\ &\quad - \frac{1}{2} \int_0^T (|(c^{-1/2}b)(\vartheta'; s, X_s, \mu_s^{\vartheta'})|^2 - |(c^{-1/2}b)(\vartheta; s, X_s, \mu_s^{\vartheta})|^2) ds. \end{aligned}$$

Having Assumption 1.13 fail for some ϑ implies $\bar{\mathbb{P}}_{\vartheta'}(\frac{d\bar{\mathbb{P}}_{\vartheta'}}{d\bar{\mathbb{P}}_{\vartheta}}(X) = 1) = 1$, i.e. $\bar{\mathbb{P}}_{\vartheta} = \bar{\mathbb{P}}_{\vartheta'}$. Assumption 1.13 may be difficult to check in practice. Yet, it is satisfied as soon as the mapping $\vartheta \mapsto ((t, x, \nu) \mapsto b(\vartheta; t, x, \nu))$ is one-to-one. Also, for certain form of the likelihood, we have other criteria, see Proposition 1.16 below.

Non-degeneracy of the information

We need some notation. For any $\vartheta, \vartheta' \in \Theta$ such that the segment $[\vartheta, \vartheta'] = \{\vartheta + \lambda(\vartheta' - \vartheta), \lambda \in [0, 1]\} \subset \Theta$ and a function ϕ defined on Θ , we set

$$\phi([\vartheta, \vartheta']) = \int_0^1 \phi(\vartheta + \lambda(\vartheta' - \vartheta)) d\lambda.$$

Definition 1.14. *The family of statistical experiments $(\mathcal{E}^N)_{N \geq 1}$ is asymptotically non-degenerate if*

$$\inf_{[\vartheta, \vartheta'] \subset \Theta} \det \mathbb{I}([\vartheta, \vartheta']) > 0, \quad (1.22)$$

where $\det$ denotes the determinant.

Equivalently, we can rewrite (1.22) as

$$\inf \det \left(\sum_{j=1}^d \int_0^T \int_{\mathbb{R}^d} \nabla_{\vartheta} (c^{-1/2} b)^j([\vartheta, \vartheta']; t, x, \mu_t^{\vartheta}) \nabla_{\vartheta} (c^{-1/2} b)^j([\vartheta, \vartheta']; t, x, \mu_t^{\vartheta})^{\top} \mu_t^{\vartheta}(dx) dt \right) > 0,$$

where the infimum is taken over all segments $[\vartheta, \vartheta'] \subset \Theta$. Obviously, if $(\mathcal{E}^N)_{N \geq 1}$ is asymptotically non-degenerate, taking $\vartheta = \vartheta'$, Definition 1.14 boils down to

$$\inf_{\vartheta \in \Theta} \det \mathbb{I}(\vartheta) > 0 \quad (1.23)$$

i.e. $\vartheta \mapsto \mathbb{I}(\vartheta)$ has full rank uniformly in ϑ and we find back the usual non-degeneracy of the Fisher information. The somewhat stronger non-degeneracy criterion that we pick in Definition 1.14 enables us to check the assumptions of the theory of Ibragimov and Hasminski for obtaining sharp properties for the maximum likelihood estimator (see in particular Step 2 of the proof of Theorem 1.19 in Section 5.3 below). In explicit examples, proving (1.22) is no more difficult than proving (1.23), see Section 4 below.

Checking (1.22) or (1.23) in practice

The special difficulty for the statistical analysis of $\mathcal{E}^N$ (resp. $\mathcal{H}$, $\mathcal{G}_{\vartheta_0}(\vartheta_0 \in \Theta)$) lies of course in the presence of $(\mu_t^{(N)})_{0 \leq t \leq T}$ (resp. $(\mu_t^{\vartheta})_{0 \leq t \leq T}$, $(\mu_t^{\vartheta_0})_{0 \leq t \leq T}$) in the drift. Moreover, even if the study of $\mathbb{I}_{\mathcal{E}^N}(\vartheta)$ reduces asymptotically to the study of its asymptotic information matrix $\mathbb{I}(\vartheta)$, $(\mu_t^{\vartheta})_{0 \leq t \leq T}$ still appears in the form of $\mathbb{I}(\vartheta)$, and is never explicit, except in very special cases with a specific moment structure in the measure dependence, see Section 4 below and Belomestny *et al.* [9, 8].

Nevertheless, it is noteworthy that (1.22) can usually be tested in a simple way given an explicit parametrisation. Indeed, Definition 1.14 is equivalent to show that for every segment $[\vartheta, \vartheta'] \subset \Theta$,

$$\inf_{[\vartheta, \vartheta'] \subset \Theta} \min_{|z|=1} \sum_{j=1}^d \int_0^T \int_{\mathbb{R}^d} (\nabla_{\vartheta} (c^{-1/2} b)^j([\vartheta, \vartheta']; t, x, \mu_t^{\vartheta})^{\top} z)^2 \mu_t^{\vartheta}(dx) dt > 0.$$

Under Assumptions 1.1, 1.2, 1.3, we have that $\mu_t^{\vartheta}(dx) = \mu_t^{\vartheta}(x)dx$ is absolutely continuous on $\mathbb{R}^d$ for $t > 0$, and we may pick a version μ_t^{ϑ} of the density that is continuous and positive on $\mathbb{R}^d$. This

follows from classical Gaussian tail estimates for the solution of parabolic equations. We refer for example to Corollary 8.2.2 of [17]. By a simple continuity argument, it is then sufficient to show that there cannot exist a segment $[\vartheta, \vartheta'] \subset \Theta$ and some $|z| = 1$, such that the function

$$x \mapsto \int_0^T \sum_{j=1}^d (\nabla_{\vartheta}(c^{-1/2}b)^j([\vartheta, \vartheta']; t, x, \mu_t^{\vartheta})^{\top} z)^2 dt$$

vanishes asymptotically, or, as soon as we have continuity in t as $t \rightarrow 0$, if one of the functions

$$x \mapsto \nabla_{\vartheta}(c^{-1/2}b)^j([\vartheta, \vartheta']; 0, x, \mu_0)^{\top} z, \quad j = 1, \dots, d$$

does not identically vanishes. This last criterion has the advantage to avoid the term μ_t^{ϑ} for $t > 0$. We gather these observations in the following :

Proposition 1.15. *Work under Assumptions 1.1, 1.2, 1.3 and 1.4. Assume moreover that the functions*

$$t \mapsto \nabla_{\vartheta}(c^{-1/2}b)^j([\vartheta, \vartheta']; t, x, \mu_t^{\vartheta}), \quad j = 1, \dots, d$$

are all continuous at $t = 0$ for every $[\vartheta, \vartheta'] \subset \Theta$ and a.e.-almost $x \in \mathbb{R}^d$.

If, for every $[\vartheta, \vartheta'] \subset \Theta$ and any $z \in \mathbb{R}^p$ with $|z| = 1$, one of the functions

$$x \mapsto \nabla_{\vartheta}(c^{-1/2}b)^j([\vartheta, \vartheta']; 0, x, \mu_0)^{\top} z, \quad j = 1, \dots, d \tag{1.24}$$

does not identically vanish, then $(\mathcal{E}^N)_{N \geq 1}$ is asymptotically non-degenerate in the sense of Definition 1.14.

We specifically apply this criterion in the examples Section 4 and check that the criterion (1.24) is particularly simple to establish when the dependence in the measure argument of the function b is of the form (1.13).

A case of equivalence between global identifiability and non-degeneracy of the information

We revisit Theorem 3 in [167] to obtain the following criterion :

Proposition 1.16. *Work under Assumptions 1.1, 1.2, 1.3 and 1.4. Assume that the log-likelihood $\ell^N(\vartheta; X^{(N)})$ in $\mathcal{E}^N$ defined by (1.9) has the form*

$$\ell^N(\vartheta; X^{(N)}) = \vartheta^{\top} G^N(X^{(N)}) + \vartheta^{\top} H^N(X^{(N)}) \vartheta, \tag{1.25}$$

where G^N and H^N are functions of the trajectory $X^{(N)}$ with values in $\mathbb{R}^p$ and $\mathbb{R}^p \otimes \mathbb{R}^p$ respectively, and $(H^N)^{\top} = H^N$ is symmetric. If $\Theta_0 \subset \Theta$ is a convex set such that $\mathbb{I}(\vartheta)$ is non-singular for every $\vartheta \in \Theta_0$, then, $(\mathcal{E}^N)_{N \geq 1}$ is identifiable on Θ_0 .

By identifiability of the sequence of experiment $(\mathcal{E}^N)_{N \geq 1}$, we mean injectivity of the mapping $\vartheta \mapsto (\mathbb{P}_{\vartheta}^N)_{N \geq 1}$ (i.e. simultaneously for every $N \geq 1$). The proof is given in Section 6.4. In the specific case of McKean type models that date back to [146, 173, 178] and widely used in practice (see e.g. [37, 73] or [111] in statistics), we have in some instances a representation like (1.25) and explicit formulas for $\mathbb{I}(\vartheta)$, which gives global identifiability for free as soon as $\mathbb{I}(\vartheta)$ is non-degenerate. See the examples in Section 4.

3 Main results

3.1 The LAN property

The local asymptotic normality property of a statistical model characterises its regularity : it expresses the fact that the experiment locally resembles a Gaussian shift in an optimal scale driven by the Fisher information. It has powerful consequences in terms of properties of optimal procedures via the celebrated Hájek convolution theorem [90]. More precisely the sequence of experiments $(\mathcal{E}_N)_{N \geq 1}$ satisfies the LAN property at $\vartheta \in \Theta$ with information rate $N\mathbb{I}(\vartheta)$ if

$$\log \frac{d\mathbb{P}_\vartheta^N}{d\mathbb{P}_\vartheta^N} = u^\top \xi_\vartheta^N - \frac{1}{2}|u|^2 + r_N(\vartheta, u), \quad (1.26)$$

where ξ_ϑ^N converges in distribution under $\mathbb{P}_\vartheta^N$ to standard Gaussian variable in $\mathbb{R}^p$ and $r_N(\vartheta, u) \rightarrow 0$ in $\mathbb{P}_\vartheta^N$ -probability. Of course, the convergence (1.26) is meaningful only if $\vartheta + (N\mathbb{I}(\vartheta))^{-1/2}u \in \Theta$ and is well-defined, *i.e.* if $\det \mathbb{I}(\vartheta) > 0$. This is granted for instance for ϑ in the interior of Θ for large enough N and under (1.23).

Theorem 1.17. *Work under Assumptions 1.1, 1.2, 1.3, 1.4 and 1.13. Assume moreover that $\mathcal{E}^N$ is asymptotically non-degenerate according to Definition 1.14. For every ϑ in (the interior of) Θ , the sequence of experiments $(\mathcal{E}^N)_{N \geq 1}$ is then locally asymptotically normal at ϑ with information rate $N\mathbb{I}(\vartheta)$.*

Several remarks are in order : **1)** Theorem 1.17 is the most powerful result one can obtain about the structure of $(\mathcal{E}^N)_{N \geq 1}$: it tells us that around a given point ϑ_0 , if we parametrise locally the experiment via $\vartheta = \vartheta_0 + N^{-1/2}u$ with $u \in \mathbb{R}^p$ being the unknown parameter, then the experiments look like the simplest possible experiment, namely a Gaussian shift

$$Y^N = u + \mathbb{I}(\vartheta_0)^{-1/2}\xi + o(1)$$

where ξ is a standard normal $\mathcal{N}(0, \text{Id}_{\mathbb{R}^p})$ and $o(1)$ is a small term that vanishes in $\mathbb{P}_\vartheta^N$ probability, locally uniformly in u . **2)** The LAN property has several consequences in terms of strong properties of the maximum likelihood estimator, see Theorem 1.19 below. In particular, the first simple consequence is given in terms of exact asymptotic minimax lower bounds : call a centrally symmetric function $w : \mathbb{R}^p \rightarrow [0, \infty)$ such that the sets $\{w < c\}, c > 0$ are all convex a *polynomial loss function* if it admits a polynomial majorant.

Corollary 1.18. *In the setting of Theorem 1.17, let w be a polynomial loss function. Then, for any estimator $\hat{\vartheta}_N$ in $\mathcal{E}^N$ and any sufficiently small $\delta > 0$, for every ϑ in (the interior of) Θ for which $\det \mathbb{I}(\vartheta) > 0$, we have*

$$\liminf_{N \rightarrow \infty} \sup_{|\vartheta' - \vartheta| \leq \delta} \mathbb{E}_{\mathbb{P}_{\vartheta'}^N} [w(N^{1/2}\mathbb{I}(\vartheta)^{1/2}(\hat{\vartheta}_N - \vartheta'))] \geq (2\pi)^{-p/2} \int_{\mathbb{R}^p} w(x) \exp(-\frac{1}{2}|x|^2) dx.$$

Corollary 1.18 is a simple application of Hájek convolution theorem, given the LAN property of Theorem 1.17, see *e.g.* Theorem II.12.1 (and in particular Remark III.12.1) in [98]. It gives a sharp local asymptotically minimax bound. We shall see below that the maximum likelihood estimator achieves this bound.

3.2 Maximum likelihood estimation and properties

We elaborate on the properties of the maximum likelihood estimator by relying on (a uniform version of) the LAN property of Theorem 1.17. It implies several fine results that go beyond the usual asymptotic weak expansions given by an *ad-hoc* study of the form of the estimator, as is usually the case in the literature.

Theorem 1.19. *Work under Assumptions 1.1, 1.2, 1.3, 1.4 and 1.13. Then, for large enough N , the solution $\hat{\vartheta}_N^{\text{MLE}}$ to*

$$\mathcal{L}^N(\hat{\vartheta}_N^{\text{MLE}}; X^{(N)}) = \sup_{\vartheta \in \Theta} \mathcal{L}^N(\vartheta; X^{(N)}) \quad (1.27)$$

is well-defined. Moreover, the following asymptotic upper bounds are valid :

(i) *if $\mathcal{E}^N$ is asymptotically non-degenerate in the sense of Definition 1.14,*

$$\sqrt{N}(\hat{\vartheta}_N^{\text{MLE}} - \vartheta) \rightarrow \mathcal{N}(0, \mathbb{I}(\vartheta)^{-1})$$

in $\mathbb{P}_\vartheta^N$ -distribution as $N \rightarrow \infty$.

(ii) *For every polynomial loss function w and any ϑ in the interior of Θ , we have exact local asymptotic minimax optimality :*

$$\limsup_{N \rightarrow \infty} \sup_{|\vartheta' - \vartheta| \leq \delta} \mathbb{E}_{\mathbb{P}_{\vartheta'}^N} [w(N^{1/2} \mathbb{I}(\vartheta)^{1/2} (\hat{\vartheta}_N^{\text{MLE}} - \vartheta'))] \rightarrow (2\pi)^{-p/2} \int_{\mathbb{R}^p} w(x) \exp(-\frac{1}{2}|x|^2) dx$$

as $\delta \rightarrow 0$.

(iii) *For every polynomial loss function w and any (non empty) open set $\Theta_0 \subset \Theta$, we have global asymptotic minimax optimality :*

$$\mathcal{R}_w^N(\hat{\vartheta}_N^{\text{MLE}}; \Theta_0) = \inf_{\hat{\vartheta}_N} \mathcal{R}_w^N(\hat{\vartheta}_N; \Theta_0)(1 + o(1))$$

as $N \rightarrow \infty$, where

$$\mathcal{R}_w^N(\hat{\vartheta}_N; \Theta_0) = \sup_{\vartheta \in \Theta_0} \mathbb{E}_{\mathbb{P}_\vartheta^N} [w(N^{1/2} \mathbb{I}(\vartheta)^{1/2} (\hat{\vartheta}_N - \vartheta))].$$

Some further remarks : **1)** We find back the classical asymptotic properties (i) of the maximum likelihood estimator that are given in the literature, but the result is appended by a much stronger convergence in (ii), that matches in particular the lower bound of Corollary 1.18. **2)** We finally obtain global asymptotic minimax optimality by (iii), which is the parametric analog (in a much more precise way) of our minimax results of Section 4 in [65] in the nonparametric case. **3)** Theorem 1.19 is likely to remain valid if we replace the experiment $\mathcal{E}^N$ by $\mathcal{H}^{\otimes N}$ or $\mathcal{G}_{\vartheta_0}^{\otimes N}$ ($\vartheta_0 \in \Theta$), with appropriate modifications of the definition of the maximum likelihood and the information matrix : $\hat{\vartheta}_N^{\text{MLE}}$ would be now defined respectively via the log-likelihood $\bar{\ell}^N(\vartheta; X^{(N)})$ as defined in (1.18) and via the log-likelihood $\bar{\ell}_{\vartheta_0}^N(\vartheta; X^{(N)})$ as defined in (1.21), and $\mathbb{I}(\vartheta)$ by respectively $\mathbb{I}_{\mathcal{H}}(\vartheta)$ and $\mathbb{I}_{\mathcal{G}_{\vartheta_0}}(\vartheta)$. However, let us stress that this is not a feasible choice since these likelihoods explicitly depend respectively on μ_t^ϑ and $\mu_t^{\vartheta_0}$ which are intractable in practice (and that we need additional regularity property on the model in order to ensure that $\mathbb{I}_{\mathcal{H}}(\vartheta)$ is well defined).

4 Examples

In this section, we elaborate on specific examples that appear in the literature and in applications. We first revisit the linear McKean model studied at length in [111]. We slightly extend in Section 4.1 his example (1.3) from $p = 2$ to $p = 3$. In Section 4.2, we develop an example of a generalised linear form and show in particular how our identifiability and non-degeneracy criteria of Section 2.4 are easily implementable and avoid to use the machinery of [111]. In Section 4.3, we develop a non-trivial example of kinetic mean-field model with a double layer potential that may serve in many applications, like swarming models or more general individual based-models, see [19] and the references therein. We finally develop a genuinely non-linear example, *i.e.* when the measure argument is not linear like in (1.4), as for instance in the examples of [162]. Assumption 1.1 is in force throughout.

4.1 McKean-like models

In many applications, (1.2) takes the explicit form

$$dX_t^i = (\vartheta_1 X_t^i + \vartheta_2)dt - \vartheta_3 N^{-1} \sum_{j=1}^N (X_t^i - X_t^j)dt + dB_t^i, \quad i = 1, \dots, N \quad (1.28)$$

with $X_t^i \in \mathbb{R}$. The parameter is $\vartheta = (\vartheta_1 \ \vartheta_2 \ \vartheta_3)^\top$. In [111] the case $\vartheta_2 = 0$ is studied at length in particular. In our setting, we can encompass a more general situation with $X_t^i \in \mathbb{R}^d$ for some arbitrary $d \geq 1$ and replace ϑ_3 by a parameter in $\mathbb{R}^d \otimes \mathbb{R}^d$ as well as ϑ_2 by a parameter in $\mathbb{R}^d$. In this case, Assumptions 1.2, 1.3 and 1.4 are readily checked. Likewise, the identifiability and non-degeneracy assumptions can be obtained with some extra care on the initial condition. We elaborate on a specific case below.

Likelihood equations

To keep-up with notational simplicity, we detail the case $p = 3$ with $\vartheta = (\vartheta_1 \ \vartheta_2 \ \vartheta_3)^\top \in \Theta$ as a compact subset of $\mathbb{R}^3$ for an ambient dimension $d = 1$, with $\vartheta_1 \neq \vartheta_3$ and $\vartheta_1 \neq 0$. Introduce

$$A_t^N(x) = \begin{pmatrix} x^2 & x & -\langle \cdot - x, \mu_t^{(N)} \rangle^2 \\ x & 1 & 0 \\ -\langle \cdot - x, \mu_t^{(N)} \rangle^2 & 0 & \langle \cdot - x, \mu_t^{(N)} \rangle^2 \end{pmatrix}, \quad B_t^N(x) = \begin{pmatrix} x \\ 1 \\ \langle \cdot - x, \mu_t^{(N)} \rangle \end{pmatrix},$$

where we use the bracket notation $\langle \cdot, \nu \rangle$ to denote integration w.r.t. the measure ν . Define

$$A_T^N = \int_0^T \langle A_t^N(x), \mu_t^{(N)} \rangle dt \quad (1.29)$$

and

$$B_T^N = N^{-1} \sum_{i=1}^N \int_0^T B_t^N(X_t^i) dX_t^i. \quad (1.30)$$

Thanks to the linearity in ϑ of the drift $b(\vartheta; t, x, \nu) = \vartheta_1 x + \vartheta_2 - \vartheta_3 \int_{\mathbb{R}} (x - y) \nu(dy)$, the likelihood equations are explicit and the maximum likelihood estimator $\hat{\vartheta}_N^{\text{mle}}$ solves

$$A_T^N \hat{\vartheta}_N^{\text{mle}} = B_T^N. \quad (1.31)$$

Moreover, the asymptotic information matrix is given by

$$\mathbb{I}(\vartheta) = \int_0^T \langle A_t(\vartheta; x), \mu_t^\vartheta \rangle dt,$$

with

$$A_t(\vartheta; x) = \begin{pmatrix} x^2 & x & -\langle \cdot - x, \mu_t^\vartheta \rangle^2 \\ x & 1 & 0 \\ -\langle \cdot - x, \mu_t^\vartheta \rangle^2 & 0 & \langle \cdot - x, \mu_t^\vartheta \rangle^2 \end{pmatrix}.$$

Non-degeneracy and identifiability

The property $\det \mathbb{I}(\vartheta) > 0$ can be verified on the explicit form of its matrix :

$$\det \mathbb{I}(\vartheta) = \int_0^T \text{Var}(\mu_t^\vartheta) dt \left(T \int_0^T \bar{\mathbf{m}}_1(\mu_t^\vartheta)^2 dt - \left(\int_0^T \bar{\mathbf{m}}_1(\mu_t^\vartheta) dt \right)^2 \right)$$

with $\text{Var}(\nu) = \int_{\mathbb{R}} (x - \bar{\mathbf{m}}_1(\nu))^2 \nu(dx)$ and $\bar{\mathbf{m}}_1(\nu) = \int_{\mathbb{R}^d} x \nu(dx)$. Therefore $\det \mathbb{I}(\vartheta) > 0$ unless μ_t^ϑ is degenerate for all t or stationary. In the case of a linear equation of the type (1.28), the accompanying limiting measure μ_t^ϑ associated to the McKean-Vlasov equation

$$dX_t = (\vartheta_1 X_t + \vartheta_2) dt - \vartheta_3 (X_t - \mathbb{E}_{\mathbb{P}_\vartheta}[X_t]) dt + dB_t$$

is a Gaussian process that can be thought of as an inhomogeneous Ornstein-Uhlenbeck model for which we have closed-form moment formulas :

$$\bar{\mathbf{m}}_1(\mu_t^\vartheta) = -\vartheta_1^{-1} \vartheta_2 + (\bar{\mathbf{m}}_1(\mu_0) + \vartheta_1^{-1} \vartheta_2) \exp(\vartheta_1 t) \quad (1.32)$$

and

$$\mathbf{m}_2(\mu_t^\vartheta) = \exp(-2(\vartheta_1 - \vartheta_3)t) \text{Var}(\mu_0) + \frac{1 - \exp(-2(\vartheta_1 - \vartheta_3)t)}{2(\vartheta_1 - \vartheta_3)} + \bar{\mathbf{m}}_1(\mu_t^\vartheta)^2. \quad (1.33)$$

In particular, having

$$\bar{\mathbf{m}}_1(\mu_0) + \vartheta_1^{-1} \vartheta_2 \neq 0, \quad \vartheta_1 \neq 0, \quad \vartheta_1 \neq \vartheta_3 \quad (1.34)$$

yields the non-degeneracy of $\mathbb{I}(\vartheta)$ as well as the asymptotic non-degeneracy of $\mathcal{E}^N$ in the sense of Definition 1.14, since $\nabla_{\vartheta} b(\vartheta; t, x, \nu)$ does not depend on ϑ . Also, the convergence $A_T^N \rightarrow \mathbb{I}(\vartheta)$ in $\mathbb{P}_\vartheta^N$ -probability as $N \rightarrow \infty$ tells us that (1.37) has a well defined and unique solution with $\mathbb{P}_\vartheta^N$ probability tending to one as $N \rightarrow \infty$. This last statement can be quantified via the convergence Lemma 1.21 below. Writing $\phi(x, \nu) = (x \ 1 \ - \int_{\mathbb{R}} (x - y) \nu(dy))$, the log-likelihood

$$\ell^N(\vartheta; X^{(N)}) = \sum_{i=1}^N \int_0^T \vartheta^\top \phi(X_t^i, \mu_t^{(N)}) dX_t^i - \frac{1}{2} \sum_{i=1}^N \int_0^T (\vartheta^\top \phi(X_t^i, \mu_t^{(N)}))^2 dt \quad (1.35)$$

has representation (1.25) with

$$G^N(X^{(N)}) = \sum_{i=1}^N \int_0^T \phi(X_t^i, \mu_t^{(N)}) dX_t^i \text{ and } H^N(X^{(N)}) = -\frac{1}{2} \sum_{i=1}^N \int_0^T \phi(X_t^i, \mu_t^{(N)}) \phi(X_t^i, \mu_t^{(N)})^\top dt$$

and we obtain the asymptotic identifiability of $(\mathcal{E}^N)_{N \geq 1}$ over compact parameter sets $\Theta \subset \mathbb{R}^3$ that are moreover convex and satisfy the constraint (1.34) by Proposition 1.16. Finally, explicit formulas for $\mathbb{I}(\vartheta)$ and its inverse can be derived thanks to (1.32) and (1.33).

4.2 Generalised linear like models

We push further the preceding linear structure by considering the following model

$$dX_t^i = \vartheta_1 f(X_t^i) dt + \vartheta_2 N^{-1} \sum_{j=1}^N g(X_t^i - X_t^j) dt + dB_t^i, \quad 1 \leq i \leq N$$

where f and g are known and Lipschitz continuous real-valued functions and $X_t^i \in \mathbb{R}$ for simplicity. The parameter is $\vartheta = (\vartheta_1, \vartheta_2) \in \mathbb{R}^2$. Thus $p = 2$ which again yields simple and explicit formulas. Assumptions 1.2, 1.3 and 1.4 are readily checked. Writing $\phi(x, \nu) = (f(x) g \star \nu(x))^\top$, with $g \star \nu(x) = \int_{\mathbb{R}} g(x-y) \nu(dy)$ the log-likelihood function $\ell^N(\vartheta; X^{(N)})$ has again representation (1.35) hence (1.25) holds as well and we obtain the asymptotic identifiability of $(\mathcal{E}^N)_{N \geq 1}$ over compact parameter sets $\Theta \subset \mathbb{R}^2$ that are moreover convex as soon as $\det \mathbb{I}(\vartheta) > 0$ by Proposition 1.16.

Non-degeneracy

For proving asymptotic non-degeneracy (that implies in particular $\det \mathbb{I}(\vartheta) > 0$), we plan to apply Proposition 1.15. We first notice that $\nabla_{\vartheta} b(\vartheta; t, x, \nu) = \phi(x, \nu)$ does not depend on ϑ , hence $\nabla_{\vartheta} b([\vartheta, \vartheta']; t, x, \nu) = (f(x) g \star \nu(x))^\top$ does not depend on the segment $[\vartheta, \vartheta']$ either. The continuity of $t \mapsto (f(x) g \star \mu_t^\vartheta(x))^\top$ follows from the representation

$$(f(x) g \star \mu_t^\vartheta(x))^\top = (f(x) \mathbb{E}_{\mathbb{P}_\vartheta} [g(X_t)])^\top,$$

the fact that $t \mapsto X_t$ is continuous in probability at $t = 0$ under $\mathbb{P}_\vartheta$ and Lebesgue dominated convergence. Now, let $z = (z_1, z_2)$ with $z_1^2 + z_2^2 = 1$. In order to obtain non-degeneracy, it is sufficient by Proposition 1.15 to show that the function

$$x \mapsto f(x) z_1 + g \star \mu_0(x) z_2$$

is non identically zero. If f does not identically vanishes, we may assume that $z_2 \neq 0$. Then it is sufficient to have that

$$x \mapsto \lambda f(x) + g \star \mu_0(x) \text{ does not vanish identically for every } \lambda \neq 0. \quad (1.36)$$

It is then very easy to build families of functions and initial condition (f, g, μ_0) such that (1.36) is satisfied. For instance, if $\mu_0 = \delta_{x_0}$ for some arbitrary x_0 , then having f non-identically equal to a constant is sufficient.

Likelihood equations

Finally, we explicitly solve the likelihood equations. Again, they are of a simple form, and the maximum likelihood estimator $\hat{\vartheta}_N^{\text{MLE}}$ solves

$$\mathbf{A}_T^N \hat{\vartheta}_N^{\text{MLE}} = \mathbf{B}_T^N, \quad (1.37)$$

where $\mathbf{A}_T^N$ and $\mathbf{B}_T^N$ are defined via (1.29) and (1.30), with

$$\mathbf{A}_t^N(x) = \begin{pmatrix} f(x)^2 & f(x) \langle g(x - \cdot), \mu_t^{(N)} \rangle \\ f(x) \langle g(x - \cdot), \mu_t^{(N)} \rangle & \langle g(x - \cdot), \mu_t^{(N)} \rangle^2 \end{pmatrix}, \quad \mathbf{B}_t^N(x) = \begin{pmatrix} f(x) \\ \langle g(x - \cdot), \mu_t^{(N)} \rangle \end{pmatrix}.$$

Again, we have

$$\mathbb{I}(\vartheta) = \begin{pmatrix} \int_0^T \langle f^2, \mu_t^\vartheta \rangle dt & \int_0^T \langle f(g \star \mu_t^\vartheta), \mu_t^\vartheta \rangle dt \\ \int_0^T \langle f(g \star \mu_t^\vartheta), \mu_t^\vartheta \rangle dt & \int_0^T \langle (g \star \mu_t^\vartheta)^2, \mu_t^\vartheta \rangle dt \end{pmatrix},$$

by taking the limit in $\mathbb{P}_\vartheta^N$ -probability of $A_T^N = \int_0^T \langle A_t^N(x), \mu_t^{(N)} \rangle dt$. We also know that $\det \mathbb{I}(\vartheta) > 0$ by the non-degeneracy property established above thanks to Proposition 1.15. Hence A_T^N is invertible with $\mathbb{P}_\vartheta^N$ -probability that goes to one as $N \rightarrow \infty$ and $\widehat{\vartheta}_N^{\text{ml}\vartheta}$ is asymptotically well defined.

4.3 A double layer potential model

We depart from the structure (1.25) of the likelihood as in the two preceding linear-like models and study the model

$$dX_t^i = N^{-1} \sum_{j=1}^N \nabla U_\vartheta(X_t^i - X_t^j) dt + dB_t^i, \quad 1 \leq i \leq N,$$

with ambient state space $\mathbb{R}^d$ for $d \geq 1$ and where $U_\vartheta : \mathbb{R}^d \rightarrow \mathbb{R}$ is a family of pairwise potentials of the form

$$U_\vartheta(x) = -\vartheta_1 \exp(-\vartheta_2 |x|^2) + \vartheta_3 \exp(-\vartheta_4 |x|^2)$$

modelling short range repulsion and long range attraction, where ϑ_1, ϑ_3 and ϑ_2, ϑ_4 are respectively the strengths and the lengths of attraction and repulsion. The parameter is $\vartheta = (\vartheta_1, \dots, \vartheta_4) \in (0, \infty)^4$. As minimal identifiability condition, we impose $\vartheta_2 \neq \vartheta_4$. Such models are commonly used (in their kinetic version) in swarming modelling, see *e.g.* [19].

We have $b(\vartheta; t, x, \nu) = \nabla U_\vartheta \star \nu(x)$ and it is readily verified that Assumptions 1.2, 1.3 and 1.4 are met. Here, there is no hope to explicitly solve the likelihood equations, and a numerical scheme has to be implemented. We further investigate the identifiability and non-degeneracy of the model. Note that the assumptions on the drift and the initial condition ensure that for every $\vartheta \in \Theta$, μ_t^ϑ is absolutely continuous with a nowhere vanishing density for $t > 0$ (we refer to [17] and in particular Corollary 8.2.2).

Identifiability

We study the injectivity of the mapping $\vartheta \mapsto \overline{\mathbb{P}}_\vartheta$ via the injectivity of $\vartheta \mapsto ((t, x) \mapsto \nabla U_\vartheta \star \mu_t^\vartheta)$, see Assumption 1.13. If ϑ, ϑ' are such that $\overline{\mathbb{P}}_\vartheta = \overline{\mathbb{P}}_{\vartheta'}$, this also implies $\mu_t^\vartheta = \mu_t^{\vartheta'}$ for every $t \in [0, T]$. Hence the condition-

$$\nabla U_\vartheta \star \mu_t^\vartheta = \nabla U_{\vartheta'} \star \mu_t^{\vartheta'}$$

for almost every $t \in [0, T]$ becomes

$$\nabla U_\vartheta \star \mu_t^\vartheta = \nabla U_{\vartheta'} \star \mu_t^\vartheta \tag{1.38}$$

for almost every $t \in [0, T]$. Let $\mathcal{F}(\nu)(\xi) = \int_{\mathbb{R}^d} e^{i x^\top \xi} \nu(dx)$ denote a Fourier transform of ν (well defined if ν is a probability measure or an integrable function). Since $\xi \mapsto \mathcal{F}(\mu_t)(\xi)$ is continuous and $\mathcal{F}(\mu_t)(\xi) = 1$ (the μ_t are all probability measures on $\mathbb{R}^d$), there are infinitely many points $\xi \in \mathbb{R}^d$ such that $\mathcal{F}(\mu_t)(\xi) \neq 0$. Applying $\mathcal{F}$ on both side of (1.38), we obtain

$$\mathcal{F}(\nabla U_\vartheta)(\xi) = \mathcal{F}(\nabla U_{\vartheta'})(\xi)$$

for such points ξ . Moreover

$$\nabla U_{\vartheta}(x) = 2\vartheta_1\vartheta_2x \exp(-\vartheta_2|x|^2) - 2\vartheta_3\vartheta_4x \exp(-\vartheta_4|x|^2), \quad (1.39)$$

and

$$\mathcal{F}(\nabla U_{\vartheta})(\xi) = i\xi\pi^{d/2}(\vartheta_1\vartheta_2^{-d/2} \exp(-\frac{1}{4\vartheta_2}|\xi|^2) - \vartheta_3\vartheta_4^{-d/2} \exp(-\frac{1}{4\vartheta_4}|\xi|^2)).$$

It follows that the condition $\vartheta_2 \neq \vartheta_4$ is sufficient to achieve identifiability, *i.e.* $\vartheta = \vartheta'$. Henceforth, we may parametrise our model via any compact $\Theta \subset (0, \infty)^4$ such that $\vartheta_2 \neq \vartheta_4$.

Non-degeneracy

We plan to apply Proposition 1.15. From (1.39), we have

$$\nabla_{\vartheta} b(\vartheta; t, x, \nu)^j = \nabla_{\vartheta} (\nabla U_{\vartheta} \star \nu(x))^j = G(\vartheta; x)^j \star \nu(x),$$

with

$$G(\vartheta; x)^j = \begin{pmatrix} 2\vartheta_2x_j \exp(-\vartheta_2|x|^2) \\ 2\vartheta_1x_j(1 - \vartheta_2|x|^2) \exp(-\vartheta_2|x|^2) \\ -2\vartheta_4x_j \exp(-\vartheta_4|x|^2) \\ -2\vartheta_3x_j(1 - \vartheta_4|x|^2) \exp(-\vartheta_4|x|^2) \end{pmatrix}, \quad j = 1, \dots, d$$

The mapping

$$t \mapsto \nabla_{\vartheta} b([\vartheta, \vartheta']; t, x, \mu_t^{\vartheta})^j = \mathbb{E}_{\mathbb{P}_{\vartheta}} [G([\vartheta, \vartheta']; x - X_t)^j]$$

is continuous at $t = 0$, as a simple consequence of the fact that $t \mapsto X_t$ is continuous in $\overline{\mathbb{P}}_{\vartheta}$ -probability at $t = 0$. It is then sufficient to prove that for any $z \in \mathbb{R}^4$ with $z_1^2 + z_2^2 + z_3^2 + z_4^2 = 1$, one of the functions

$$x \mapsto (G([\vartheta, \vartheta']; \cdot)^j \star \mu_0(x))^{\top} z, \quad j = 1, \dots, d,$$

does not vanish identically. Assume on the contrary that $(G([\vartheta, \vartheta']; \cdot)^j \star \mu_0)^{\top} z$ is identically 0 for every $1 \leq j \leq d$. Then this is also the case for $\mathcal{F}((G([\vartheta, \vartheta']; \cdot)^j \star \mu_0)^{\top} z)$. Assume now that $\mathcal{F}(\mu_0)(\xi) \neq 0$ $d\xi$ -a.e. Then from

$$\mathcal{F}((G([\vartheta, \vartheta']; \cdot)^j \star \mu_0)^{\top} z)(\xi) = \mathcal{F}(G([\vartheta, \vartheta']; \cdot)^j(\xi))^{\top} z \cdot \mathcal{F}(\mu_0)(\xi)$$

we must have

$$\xi \mapsto \mathcal{F}(G([\vartheta, \vartheta']; \cdot)^j(\xi))^{\top} z = 0 \quad d\xi - a.e.$$

for every $1 \leq j \leq d$, or equivalently

$$x \mapsto (G([\vartheta, \vartheta']; x)^j)^{\top} z = 0 \quad dx - a.e.$$

This is not possible, as proved by an inspection of the equation

$$\begin{aligned} & \int_0^1 ([\vartheta_2, \vartheta'_2]_{\lambda} x_j \exp(-[\vartheta_2, \vartheta'_2]_{\lambda} |x|^2) z_1 + [\vartheta_1, \vartheta'_1]_{\lambda} x_j (1 - [\vartheta_2, \vartheta'_2]_{\lambda} |x|^2) \exp(-[\vartheta_2, \vartheta'_2]_{\lambda} |x|^2) z_2) d\lambda \\ &= \int_0^1 ([\vartheta_4, \vartheta'_4]_{\lambda} x_j \exp(-[\vartheta_4, \vartheta'_4]_{\lambda} |x|^2) z_3 + [\vartheta_3, \vartheta'_3]_{\lambda} x_j (1 - [\vartheta_4, \vartheta'_4]_{\lambda} |x|^2) \exp(-[\vartheta_4, \vartheta'_4]_{\lambda} |x|^2) z_4) d\lambda, \end{aligned}$$

for almost every $x \in \mathbb{R}^d$, which further reduces to

$$([\vartheta_2, \vartheta'_2]_{\lambda_x} x_j z_1 + [\vartheta_1, \vartheta'_1]_{\lambda_x} x_j (1 - [\vartheta_2, \vartheta'_2]_{\lambda_x} |x|^2) z_2) \exp(-[\vartheta_2, \vartheta'_2]_{\lambda_x} |x|^2)$$

$$=([\vartheta_4, \vartheta'_4]_{\lambda'_x} x_j z_3 + [\vartheta_3, \vartheta'_3]_{\lambda'_x} x_j (1 - [\vartheta_4, \vartheta'_4]_{\lambda'_x} |x|^2) z_4) \exp(-[\vartheta_4, \vartheta'_4]_{\lambda'_x} |x|^2) \quad (1.40)$$

for almost every $x \in \mathbb{R}^d$, by the mean-value theorem for some $\lambda_x, \lambda'_x \in [0, 1]$, that also respectively depend on $(\vartheta_1, \vartheta_2, \vartheta'_1, \vartheta'_2, z_1, z_2)$ and $(\vartheta_3, \vartheta_4, \vartheta'_3, \vartheta'_4, z_3, z_4)$. A simple sufficient condition is $\vartheta_2 \neq \vartheta_4$: indeed, if ϑ_2 and ϑ_4 take values in disjoint intervals for instance, then for every $\vartheta, \vartheta' \in \Theta$, with $[\vartheta, \vartheta'] \subset \Theta$, we have $[\vartheta_2, \vartheta'_2] \cap [\vartheta_4, \vartheta'_4] = \emptyset$. Then one easily checks that (1.40) cannot hold for sufficiently large $|x|$.

If we only need to verify that $\mathbb{I}(\vartheta)$ is non-degenerate, it is sufficient to take $\vartheta = \vartheta'$ in (1.40) that simply becomes

$$(\vartheta_2 x_j z_1 + \vartheta_1 x_j (1 - \vartheta_2 |x|^2) z_2) \exp(-\vartheta_2 |x|^2) = (\vartheta_4 x_j z_3 + \vartheta_3 x_j (1 - \vartheta_4 |x|^2) z_4) \exp(-\vartheta_4 |x|^2)$$

for almost every $x \in \mathbb{R}^d$, in which case having $\vartheta_2 \neq \vartheta_4$ is sufficient (and somewhat easier to obtain than the asymptotic non-degeneracy of $\mathcal{E}^N$ in the sense of Definition 1.14).

In conclusion, as soon as $\mathcal{F}(\mu_0)$ is non-vanishing almost everywhere and $\Theta \subset (0, \infty)^4$ is a compact such that ϑ_2 and ϑ_4 take values in disjoint intervals, we obtain the asymptotic non-degeneracy of $(\mathcal{E}^N)_{N \geq 1}$.

4.4 A genuinely non-linear example

We end-up this section by inspecting an example where the parametrisation $\nu \mapsto b(\vartheta; t, x, \nu)$ is genuinely non-linear in the measure argument. Consider the model

$$dX_t^i = F\left(\vartheta N^{-1} \sum_{j=1}^N g(X_t^i - X_t^j)\right) dt + dB_t^i, \quad 1 \leq i \leq N,$$

with $X_t^i \in \mathbb{R}$ and $\vartheta > 0$ for simplicity. The functions $F, g : \mathbb{R} \rightarrow \mathbb{R}$ are known and smooth, g is nonnegative, integrable, with positive mass and F is one-to-one on the positive axis. We have $b(\vartheta; t, x, \nu) = F(\vartheta g \star \nu(x))$.

The smoothness of F and g yields Assumptions 1.2, 1.3 and 1.4. As for the identifiability of the model, assume that $\mathbb{P}_\vartheta = \mathbb{P}_{\vartheta'}$, and so $\mu_t^\vartheta = \mu_t^{\vartheta'}$ as well. If, for almost every $x \in \mathbb{R}$, we have

$$F(\vartheta g \star \mu_t^\vartheta(x)) = F(\vartheta' g \star \mu_t^{\vartheta'}(x)),$$

then, since $g(x) \geq 0$ and $\mu_t^\vartheta(x) > 0$ for every $x \in \mathbb{R}$ and $t > 0$, the function $g \star \mu_t^\vartheta$ is positive on the whole real line $\mathbb{R}$. Since F is one-to-one on the positive axis, it follows that

$$\vartheta g \star \mu_t^\vartheta(x) = \vartheta' g \star \mu_t^{\vartheta'}(x)$$

and that can only be true if $\vartheta = \vartheta'$ since $g \star \mu_t^\vartheta(x)$ is nowhere vanishing. As for the non-degeneracy, we have

$$\partial_\vartheta b(\vartheta; t, x, \mu_t) = g \star \mu_t^\vartheta(x) F'(\vartheta g \star \mu_t^\vartheta(x)) = \mathbb{E}_{\mathbb{P}_\vartheta}[g(x - X_t)] F'(\vartheta \mathbb{E}_{\mathbb{P}_\vartheta}[g(x - X_t)])$$

which is continuous at $t = 0$ by the continuity in $\mathbb{P}_\vartheta$ -probability of $t \mapsto X_t$ at $t = 0$. Finally, assume that μ_0 has a positive density. Then $x \mapsto g \star \mu_0(x) F'(\vartheta g \star \mu_0(x))$ is non-identically vanishing since $g \star \mu_0$ is positive and F' is non-vanishing on $(0, \infty)$. We conclude by Proposition 1.15.

5 Proof of the main results

5.1 Preliminaries : couplings

For technical reasons, we will need certain couplings on the canonical space $(\mathcal{C}^N, \mathcal{F}^N)$. We now fix $\vartheta \in \Theta$, and introduce, for every $\vartheta' \in \Theta$, the following two processes

$$X^{(N),\vartheta'} = (X_t^{1,\vartheta'}, \dots, X_t^{N,\vartheta'})_{t \in [0,T]}$$

and

$$\bar{X}^{(N),\vartheta'} = (\bar{X}_t^{1,\vartheta'}, \dots, \bar{X}_t^{N,\vartheta'})_{t \in [0,T]}$$

defined on $(\mathcal{C}^N, \mathcal{F}^N)$ by

$$X_t^{i,\vartheta'} = X_0^i + \int_0^t b(\vartheta'; s, X_s^{i,\vartheta'}, \mu_t^{(N),\vartheta'}) ds + \int_0^t \sigma(s, X_s^{i,\vartheta'}) dB_s^{i,N,\vartheta}, \quad (1.41)$$

with $\mu_t^{(N),\vartheta'} = N^{-1} \sum_{i=1}^N \delta_{X_t^{i,\vartheta'}}$ and

$$\bar{X}_t^{i,\vartheta'} = X_0^i + \int_0^t b(\vartheta'; s, \bar{X}_s^{i,\vartheta'}, \mu_t^{\vartheta'}) ds + \int_0^t \sigma(s, \bar{X}_s^{i,\vartheta'}) dB_s^{i,N,\vartheta}, \quad (1.42)$$

and

$$B_t^{i,N,\vartheta} = \int_0^t c^{-1/2}(s, X_s^i) (dX_s^i - b(\vartheta, s, X_s^i, \mu_s^{(N)}) ds).$$

Note that $X^{(N),\vartheta'}$ and $\bar{X}^{(N),\vartheta'}$ actually depend on ϑ pathwise via $(B_t^{i,N,\vartheta})_{t \in [0,1]}$ under $\mathbb{P}_\vartheta^N$ but not their laws! Indeed, the $(B_t^{i,N,\vartheta})_{t \in [0,1]}$ are standard Brownian motions under $\mathbb{P}_\vartheta^N$. For notational simplicity, we omit the dependence upon ϑ here. Otherwise, we write $X_t^{i,\vartheta',\vartheta}$ or $\bar{X}_t^{i,\vartheta',\vartheta}$. Thus $X^{(N),\vartheta'}$ and $\bar{X}^{(N),\vartheta'}$ are defined as functions of $X^{(N)}$ (as strong solutions of (1.41) and (1.42)) and have law $\mathbb{P}_{\vartheta'}^N$ and $\bar{\mathbb{P}}_{\vartheta'}^{\otimes N}$ under $\mathbb{P}_\vartheta^N$. This is a convenient way to couple $\mathbb{P}_{\vartheta'}^N$ and $\bar{\mathbb{P}}_{\vartheta'}^{\otimes N}$ while still working with the canonical process under $\mathbb{P}_\vartheta^N$. We write $\mu_t^{(N)} = N^{-1} \sum_{i=1}^N \delta_{X_t^i}$, for the empirical measure of the canonical process. We also introduce

$$\bar{\mu}_t^{(N),\vartheta'} = N^{-1} \sum_{i=1}^N \delta_{\bar{X}_t^{i,\vartheta'}}$$

and write $\mu_t^{(N),\vartheta',\vartheta}$ and $\bar{\mu}_t^{(N),\vartheta',\vartheta}$ whenever we want to emphasise that the coupling is constructed with the processes $(B_t^{i,N,\vartheta})_{t \in [0,T]}$ for $1 \leq i \leq N$. We have the following approximation results :

Lemma 1.20. *For every $\vartheta, \vartheta' \in \Theta$ and every $r \geq 1$, we have*

$$\sup_{t \in [0,T]} \mathbb{E}_{\mathbb{P}_\vartheta^N} [\mathcal{W}_1(\mu_t^{(N),\vartheta',\vartheta}, \mu_t^{(N)})^r] \leq \sup_{t \in [0,T]} \mathbb{E}_{\mathbb{P}_\vartheta^N} [N^{-1} \sum_{i=1}^N |X_t^{i,\vartheta',\vartheta} - X_t^i|^r] \leq C|\vartheta' - \vartheta|^r, \quad (1.43)$$

$$\sup_{t \in [0,T]} \mathbb{E}_{\mathbb{P}_\vartheta^N} [\mathcal{W}_1(\bar{\mu}_t^{(N),\vartheta',\vartheta}, \bar{\mu}_t^{(N),\vartheta,\vartheta})^r] \leq \sup_{t \in [0,T]} \mathbb{E}_{\mathbb{P}_\vartheta^N} [N^{-1} \sum_{i=1}^N |\bar{X}_t^{i,\vartheta',\vartheta} - \bar{X}_t^{i,\vartheta,\vartheta}|^r] \leq C|\vartheta' - \vartheta|^r. \quad (1.44)$$

There exists $\delta > 0$ such that for every $r \geq 1$:

$$\sup_{t \in [0, T], \vartheta \in \Theta} \mathbb{E}_{\mathbb{P}_{\vartheta}^N} [W_1(\mu_t^{(N)}, \mu_t^{\vartheta})^r] \leq CN^{-\delta r}, \quad (1.45)$$

$$\sup_{t \in [0, T], \vartheta \in \Theta} N^{-1} \sum_{i=1}^N \mathbb{E}_{\mathbb{P}_{\vartheta}^N} [|X_t^i - \bar{X}_t^{i, \vartheta, \vartheta}|^r] \leq CN^{-\delta r} \quad (1.46)$$

as $N \rightarrow \infty$.

The proof is given in Appendix 7.3.

5.2 Proof of Theorem 1.17

We prove a slightly stronger result, namely a uniform type LAN condition, following Chapter III of [98]. Let $(u_N)_{N \geq 1}$ be a sequence of $\mathbb{R}^p$ such that $u_N \rightarrow u$ and $(\vartheta_N)_{N \geq 1}$ a sequence of Θ such that $\vartheta_N + (N\mathbb{I}(\vartheta_N))^{-1/2}u_N \in \Theta$ for large enough N and such that $\vartheta_N \rightarrow \vartheta$, for some ϑ such that $\mathbb{I}(\vartheta) > 0$. We claim that

$$\zeta_N(\vartheta_N; u_N) = \log \frac{d\mathbb{P}_{\vartheta_N + (N\mathbb{I}(\vartheta_N))^{-1/2}u_N}^N}{d\mathbb{P}_{\vartheta_N}^N} = u^\top \Gamma_N - \frac{1}{2}|u|^2 + r_N(\vartheta_N, u_N), \quad (1.47)$$

where $\Gamma_N \rightarrow \mathcal{N}(0, \text{Id}_{\mathbb{R}^p})$ in distribution under $\mathbb{P}_{\vartheta}^N$ and $r_N(\vartheta_N, u_N) \rightarrow 0$ in $\mathbb{P}_{\vartheta}^N$ -probability. Clearly, (1.47) implies (1.26) and thus Theorem 1.17. Note that since $\mathbb{I}(\vartheta) > 0$ we have that $\mathbb{I}(\vartheta_N)$ is invertible for large enough N , thanks to the continuity of the mapping $\vartheta \mapsto \mathbb{I}(\vartheta)$, recall Proposition 1.11. The asymptotic expansion (1.47) is therefore meaningful for large enough N .

Step 1. (Preliminary expansion.) We have

$$\begin{aligned} \zeta_N(\vartheta_N; u_N) &= \sum_{i=1}^N \int_0^T ((c^{-1/2}b)(\vartheta_N + (N\mathbb{I}(\vartheta_N))^{-1/2}u_N; t, X_t^i, \mu_t^N) - (c^{-1/2}b)(\vartheta_N; t, X_t^i, \mu_t^N))^\top dB_t^{i, N, \vartheta_N} \\ &\quad - \frac{1}{2} \sum_{i=1}^N \int_0^T |(c^{-1/2}b)(\vartheta_N + (N\mathbb{I}(\vartheta_N))^{-1/2}u_N; t, X_t^i, \mu_t^N) - (c^{-1/2}b)(\vartheta_N; t, X_t^i, \mu_t^N)|^2 dt, \end{aligned}$$

where the $B_t^{i, N, \vartheta_N} = \int_0^t c^{-1/2}(s, X_s^i)(dX_s^i - b(\vartheta_N; s, X_s^i, \mu_s^{(N)})ds)$, $1 \leq i \leq N$ are independent Brownian motions on $\mathbb{R}^d$ under $\mathbb{P}_{\vartheta_N}^N$. A first-order Taylor's expansion therefore yields the representation,

$$\zeta_N(\vartheta_N; u_N) = u_N^\top (\mathbb{I}(\vartheta_N)^{-1/2})^\top \Delta_{N, \vartheta_N}(u_N) - \frac{1}{2} u_N^\top (\mathbb{I}(\vartheta_N)^{-1/2})^\top \tilde{\mathbb{I}}_{N, \vartheta_N}(u_N) \mathbb{I}(\vartheta_N)^{-1/2} u_N$$

where

$$\Delta_{N, \vartheta_N}(u) = N^{-1/2} \sum_{i=1}^N \sum_{j=1}^d \int_0^T \nabla_{\vartheta}(c^{-1/2}b)^j([\vartheta_N, \vartheta_N + (N\mathbb{I}(\vartheta_N))^{-1/2}u]; t, X_t^i, \mu_t^N) d(B_t^{i, N, \vartheta_N})^j$$

and

$$\begin{aligned}\widetilde{\mathbb{I}}_{N,\vartheta_N}(u) &= N^{-1} \sum_{i=1}^N \sum_{j=1}^d \int_0^T \nabla_{\vartheta}(c^{-1/2}b)^j([\vartheta_N, \vartheta_N + (N\mathbb{I}(\vartheta_N))^{-1/2}u]; t, X_t^i, \mu_t^N) \\ &\quad \times (\nabla_{\vartheta}(c^{-1/2}b)^j([\vartheta_N, \vartheta_N + (N\mathbb{I}(\vartheta_N))^{-1/2}u]; t, X_t^i, \mu_t^N))^{\top} dt,\end{aligned}$$

with the notation $\phi([\vartheta, \vartheta']) = \int_0^1 \phi(\vartheta + \lambda(\vartheta' - \vartheta))d\lambda$ that we introduced before. We rewrite the above expansion as

$$\begin{aligned}\zeta_N(\vartheta_N; u_N) &= u_N^{\top}(\mathbb{I}(\vartheta_N)^{-1/2})^{\top} \Delta_{N,\vartheta_N}(0) - \frac{1}{2}|u|^2 \\ &\quad + u_N^{\top}(\mathbb{I}(\vartheta_N)^{-1/2})^{\top} \Delta_{N,\vartheta_N}(u_N) - u^{\top}(\mathbb{I}(\vartheta_N)^{-1/2})^{\top} \Delta_{N,\vartheta_N}(0) \\ &\quad - \frac{1}{2}(u_N^{\top}(\mathbb{I}(\vartheta_N)^{-1/2})^{\top} \widetilde{\mathbb{I}}_{N,\vartheta_N}(u_N)\mathbb{I}(\vartheta_N)^{-1/2}u_N - |u|^2)\end{aligned}$$

and thus (1.47) follows from

$$(\mathbb{I}(\vartheta_N)^{-1/2})^{\top} \Delta_{N,\vartheta_N}(0) \rightarrow \mathcal{N}(0, \text{Id}_{\mathbb{R}^p}) \quad (1.48)$$

under $\mathbb{P}_{\vartheta_N}^N$ in distribution together with the convergence to 0 of the last two components.

Step 2. (Convergence of the Gaussian part.) We prove (1.48) or equivalently, the convergence

$$\xi^{\top}(\mathbb{I}(\vartheta_N)^{-1/2})^{\top} \Delta_{N,\vartheta_N}(0) = \sum_{q,q'=1}^d \xi_q(\mathbb{I}(\vartheta_N)^{-1/2})_{q'q} \Delta_{N,\vartheta_N}(0)_{q'} \rightarrow \mathcal{N}(0, |\xi|^2)$$

in distribution under $\mathbb{P}_{\vartheta_N}^N$ for every $\xi \in \mathbb{R}^p$. We apply a classical semimartingale convergence result, following for instance Jacod and Shiryaev [102] (Corollary 3.24). For $t \in [0, T]$, the process

$$\Delta_{N,\vartheta_N}(0)_t = N^{-1/2} \sum_{i=1}^N \sum_{j=1}^d \int_0^t \nabla_{\vartheta}(c^{-1/2}b)^j(\vartheta_N; s, X_s^i, \mu_s^N) d(B_s^{i,N,\vartheta_N})^j$$

is a continuous local martingale under $\mathbb{P}_{\vartheta_N}^N$ and so is $(\xi^{\top}(\mathbb{I}(\vartheta_N)^{-1/2})^{\top} \Delta_{N,\vartheta_N}(0)_t)_{t \in [0, T]}$. It coincides with $\xi^{\top}(\mathbb{I}(\vartheta_N)^{-1/2})^{\top} \Delta_{N,\vartheta_N}(0)$ at $t = T$ and has predictable compensator

$$\begin{aligned}\langle \xi^{\top}(\mathbb{I}(\vartheta_N)^{-1/2})^{\top} \Delta_{N,\vartheta_N}(0) \cdot \rangle_t &= \sum_{q_k, q'_k=1}^p \xi_{q_1} \xi_{q_2} (\mathbb{I}(\vartheta_N)^{-1/2})_{q'_2 q_2} (\mathbb{I}(\vartheta_N)^{-1/2})_{q'_1 q_1} \times \\ &\quad N^{-1} \sum_{i=1}^N \sum_{j=1}^d \int_0^t \partial_{\vartheta_{q'_1}}(c^{-1/2}b)^j(\vartheta_N; s, X_s^i, \mu_s^{(N)}) \partial_{\vartheta_{q'_2}}(c^{-1/2}b)^j(\vartheta_N; s, X_s^i, \mu_s^{(N)}) ds,\end{aligned}$$

that converges to

$$\begin{aligned}&\sum_{q_k, q'_k=1}^p \xi_{q_1} \xi_{q_2} (\mathbb{I}(\vartheta)^{-1/2})_{q'_2 q_2} (\mathbb{I}(\vartheta)^{-1/2})_{q'_1 q_1} \times \\ &\sum_{j=1}^d \int_0^t \mathbb{E}_{\mathbb{P}_{\vartheta}} [\partial_{\vartheta_{q'_1}}(c^{-1/2}b)^j(\vartheta; s, X_s^i, \mu_s^{\vartheta}) \partial_{\vartheta_{q'_2}}(c^{-1/2}b)^j(\vartheta; s, X_s^i, \mu_s^{\vartheta})] ds,\end{aligned}$$

in $\mathbb{P}_{\vartheta_N}^N$ -probability, but that last quantity is exactly $|\xi|^2$ at $t = T$, which proves (1.48). As for the last convergence in probability, it is a simple consequence of the continuity of $\vartheta \mapsto \mathbb{I}(\vartheta)$, see Proposition 1.11 and the following lemma

Lemma 1.21. *Let $\beta > 0$ and $\phi : \Theta \times [0, T] \times \mathbb{R}^d \times \mathcal{P}_\beta \rightarrow \mathbb{R}$ be such that for some $C, \alpha > 0$, we have*

$$\sup_{t \in [0, T]} |\phi(\vartheta'; t, x', \nu') - \phi(\vartheta; t, x, \nu)| \leq C(|\vartheta' - \vartheta| + |x' - x| + \mathcal{W}_1(\nu', \nu))(1 + |x|^\alpha + |x'|^\alpha + \mathbf{m}_\beta(\nu) + \mathbf{m}_\beta(\nu')).$$

Then, there exists $0 < \delta \leq 1/2$ such that for every $t \in [0, T]$ and every $m > 0$, we have

$$\mathbb{E}_{\mathbb{P}_{\vartheta_N}^N} \left[\left| N^{-1} \sum_{i=1}^N \int_0^t \phi(\vartheta_N; s, X_s^i, \mu_s^{(N)}) ds - \int_0^t \int_{\mathbb{R}^d} \phi(\vartheta; s, x, \mu_s^\vartheta) \mu_s^\vartheta(dx) ds \right|^m \right] \leq C(|\vartheta_N - \vartheta|^m + N^{-\delta m}).$$

We apply Lemma 1.21 to $\phi(\vartheta; s, x, \nu) = \partial_{\vartheta_{q'_1}}(c^{-1/2}b)^j(\vartheta; s, x, \nu) \partial_{\vartheta_{q'_2}}(c^{-1/2}b)^j(\vartheta; s, x, \nu)$, thanks to Assumption 1.4. The proof is given in Appendix 7.4.

Step 3. (Convergence of the remainder terms.) We first prove

$$u_N^\top (\mathbb{I}(\vartheta_N)^{-1/2})^\top \Delta_{N, \vartheta_N}(u_N) - u^\top (\mathbb{I}(\vartheta_N)^{-1/2})^\top \Delta_{N, \vartheta_N}(0) \rightarrow 0 \quad (1.49)$$

in $\mathbb{P}_{\vartheta_N}^N$ -probability. Since $\mathbb{I}(\vartheta_N)^{-1/2}$ is well defined for large enough N and converges to $\mathbb{I}(\vartheta)^{-1/2}$ and $u_N \rightarrow u$, it is sufficient to prove $\Delta_{N, \vartheta_N}(u_N) - \Delta_{N, \vartheta_N}(0) \rightarrow 0$ in $\mathbb{P}_{\vartheta_N}^N$ -probability. Introduce the process

$$G_N^{i,r}(\vartheta, u)_t = \int_0^t (\partial_{\vartheta_r}(c^{-1/2}b)(\vartheta + \lambda(N\mathbb{I}(\vartheta))^{-1/2}u; t, X_t^i, \mu_t^{(N)}) - \partial_{\vartheta_r}(c^{-1/2}b)(\vartheta; t, X_t^i, \mu_t^{(N)})) d\lambda$$

for $1 \leq r \leq p$ and $t \in [0, T]$. By Itô's isometry

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_{\vartheta_N}^N} [|\Delta_{N, \vartheta_N}(u_N) - \Delta_{N, \vartheta_N}(0)|^2] &= \sum_{r=1}^p \mathbb{E}_{\mathbb{P}_{\vartheta_N}^N} [|\sum_{i=1}^N \int_0^T G_N^{i,r}(\vartheta_N, u_N)_t^\top dB_t^{i,N, \vartheta_N}|^2] \\ &= \sum_{r=1}^p N^{-1} \sum_{i=1}^N \int_0^T \mathbb{E}_{\mathbb{P}_{\vartheta_N}^N} [|G_N^{i,r}(\vartheta_N, u_N)_t|^2] dt. \end{aligned}$$

Moreover

$$\begin{aligned} |G_N^{i,r}(\vartheta_N, u_N)_t|^2 &\leq \sum_{j=1}^d \sup_{\vartheta \in \Theta} |(\nabla_\vartheta \partial_{\vartheta_r} c^{-1/2}b)^j(\vartheta; t, X_t^i, \mu_t^{(N)}) \lambda(N\mathbb{I}(\vartheta_N))^{-1/2} u_N|^2 \\ &\leq CN^{-1}(1 + |X_t^i|^{2r_1} + \mathbf{m}_{r_2}(\mu_t^{(N)})^2), \end{aligned}$$

for large enough N , thanks to Assumption 1.4. We conclude

$$\sup_{t \in [0, T]} \mathbb{E}_{\mathbb{P}_{\vartheta_N}^N} [|G_N^{i,r}(\vartheta_N, u_N)_t|^2] \leq CN^{-1}$$

for large enough N by Lemma 1.7 and (1.49) follows.

The convergence of the second remainder term is a simple consequence of $\widetilde{\mathbb{I}}_{N, \vartheta_N}(u_N) \rightarrow \mathbb{I}(\vartheta)$ thanks to Lemma 1.21 together with the continuity of $\vartheta \mapsto \mathbb{I}(\vartheta)$ and Proposition 1.11. The Proof of Theorem 1.17 is complete for the experiment $\mathcal{E}^N$.

5.3 Proof of Theorem 1.19

We plan to apply the classical theory of Ibragimov-Hasminski, and more specifically Theorem III.1.1 of [98]. We introduce the notation

$$\mathcal{Z}_N(\vartheta; u) = \frac{d\mathbb{P}_{\vartheta + (N\mathbb{I}(\vartheta))^{-1/2}u}^N}{d\mathbb{P}_{\vartheta}^N}.$$

We first establish two key regularity properties of the likelihood process.

Step 1. (A regularity property for the likelihood process.) Here, we prove that for any $r \geq 2$

$$\mathbb{E}_{\mathbb{P}_{\vartheta}^N} [|\mathcal{Z}_N(\vartheta; u)^{1/r} - \mathcal{Z}_N(\vartheta; v)^{1/r}|^r] \leq C(1 + \kappa^\gamma)|u - v|^{\gamma'}, \quad (1.50)$$

for some positive γ, γ' , uniformly in u, v such that $\mathcal{Z}_N(\vartheta; u)$ and $\mathcal{Z}_N(\vartheta; v)$ are well-defined and $|u|, |v|$ are bounded by $\kappa > 0$. Pick any $r \geq 2$. By a first-order expansion

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_{\vartheta}^N} [|\mathcal{Z}_N(\vartheta; u)^{1/r} - \mathcal{Z}_N(\vartheta; v)^{1/r}|^r] &\leq |u - v|^r \mathbb{E}_{\mathbb{P}_{\vartheta}^N} \left[\left| \int_0^1 \nabla_u (\mathcal{Z}_N)^{1/r}(\vartheta; u + \lambda(v - u)) d\lambda \right|^r \right] \\ &\leq C|u - v|^r \int_0^1 \sum_{q=1}^p \mathbb{E}_{\mathbb{P}_{\vartheta}^N} [|\partial_{u_q} ((\mathcal{Z}_N)^{1/r})(\vartheta; u + \lambda(v - u))|^r] d\lambda. \end{aligned} \quad (1.51)$$

Define, for $t \in [0, T]$, the random process

$$\phi_t(\vartheta; X^i, \mu^{(N)}) = \exp \left(\int_0^t (c^{-1}b)(\vartheta; s, X_s^i, \mu_s^{(N)}) dX_s^i - \frac{1}{2} \int_0^t |(c^{-1/2}b)(\vartheta; s, X_s^i, \mu_s^{(N)})|^2 ds \right).$$

Since

$$\mathcal{Z}_N(\vartheta; u) = \prod_{i=1}^N \frac{\phi_T(\vartheta + (N\mathbb{I}(\vartheta))^{-1/2}u, X^i, \mu^{(N)})}{\phi_T^{1/r}(\vartheta, X^i, \mu^{(N)})},$$

we have

$$\begin{aligned} \partial_{u_q} ((\mathcal{Z}_N)^{1/r})(\vartheta; u) &= \sum_{i=1}^N \frac{\partial_{u_q} (\phi_T^{1/r}(\vartheta + (N\mathbb{I}(\vartheta))^{-1/2}u; X^i, \mu^{(N)}))}{\phi_T^{1/r}(\vartheta; X^i, \mu^{(N)})} \\ &\quad \times \prod_{i' \neq i} \frac{\phi_T^{1/r}(\vartheta + (N\mathbb{I}(\vartheta))^{-1/2}u; X^{i'}, \mu^{(N)})}{\phi_T^{1/r}(\vartheta; X^{i'}, \mu^{(N)})} \\ &= \sum_{i=1}^N \frac{\nabla_{\vartheta} (\phi_T^{1/r})(\vartheta + (N\mathbb{I}(\vartheta))^{-1/2}u; X^i, \mu^{(N)})^\top ((N\mathbb{I}(\vartheta))^{-1/2})_{\cdot q}}{\phi_T^{1/r}(\vartheta + (N\mathbb{I}(\vartheta))^{-1/2}u; X^i, \mu^{(N)})} \\ &\quad \times \prod_{i'=1}^N \frac{\phi_T^{1/r}(\vartheta + (N\mathbb{I}(\vartheta))^{-1/2}u; X^{i'}, \mu^{(N)})}{\phi_T^{1/r}(\vartheta; X^{i'}, \mu^{(N)})} \\ &= (\mathcal{Z}_N)^{1/r}(\vartheta; u) ((N\mathbb{I}(\vartheta))^{-1/2})_{\cdot q} \times \\ &\quad \sum_{i=1}^N \nabla_{\vartheta} (\log(\phi_T^{1/r}))(\vartheta + (N\mathbb{I}(\vartheta))^{-1/2}u, X^i, \mu^{(N)}). \end{aligned}$$

Interpreting $\mathcal{Z}_N(\vartheta; u)$ as a Radon-Nikodym derivative entails

$$\begin{aligned} & \mathbb{E}_{\mathbb{P}_\vartheta^N} \left[\left| \partial_{u_q} ((\mathcal{Z}_N)^{1/r})(\vartheta; u) \right|^r \right] \\ &= \left| (\mathbb{I}(\vartheta))^{-1/2} \right|_q^r \mathbb{E}_{\mathbb{P}_{\vartheta + (N\mathbb{I}(\vartheta))^{-1/2}u}^N} \left[\left| N^{-1/2} \sum_{i=1}^N \nabla_\vartheta (\log(\phi_T^{1/r}))(\vartheta + (N\mathbb{I}(\vartheta))^{-1/2}u) \right|^r \right] \end{aligned} \quad (1.52)$$

by a change of probability between $\mathbb{P}_\vartheta^N$ and $\mathbb{P}_{\vartheta + (N\mathbb{I}(\vartheta))^{-1/2}u}^N$. Next, by definition of ϕ_t , we have, for every $\vartheta' \in \Theta$

$$\begin{aligned} \partial_{\vartheta_{q'}} (\log(\phi_T^{1/r}))(\vartheta'; X^i, \mu^{(N)}) &= \frac{1}{r} \int_0^T \partial_{\vartheta_{q'}} (c^{-1}b)(\vartheta'; t; X_t^i, \mu_t^{(N)}) dX_t^i \\ &\quad - \frac{1}{2r} \int_0^T 2\partial_{\vartheta_{q'}} (c^{-1/2}b)(\vartheta'; t; X_t^i, \mu_t^{(N)})^\top (c^{-1/2}b)(\vartheta'; t; X_t^i, \mu_t^{(N)}) dt \\ &= \frac{1}{r} \int_0^T \partial_{\vartheta_{q'}} (c^{-1/2}b)(\vartheta'; t; X_t^i, \mu_t^{(N)}) dB_t^{i,N,\vartheta'}, \end{aligned}$$

where the $(B_t^{i,N,\vartheta'})_{t \in [0,T]} = (\int_0^t c^{-1/2}(s, X_s^i)(dX_s^i - b(\vartheta'; s, X_s^i, \mu_s^{(N)})ds)_{t \in [0,T]}$, $1 \leq i \leq N$ are independent Brownian motions on $\mathbb{R}^d$ under $\mathbb{P}_{\vartheta'}^N$. Plugging-in this representation in (1.52) at $\vartheta' = \vartheta + (N\mathbb{I}(\vartheta))^{-1/2}u$, we infer

$$\begin{aligned} & \mathbb{E}_{\mathbb{P}_\vartheta^N} \left[\left| \partial_{u_q} ((\mathcal{Z}_N)^{1/r})(\vartheta; u) \right|^r \right] \\ &\leq C \mathbb{E}_{\mathbb{P}_{\vartheta'}^N} \left[\left| N^{-1/2} \sum_{i=1}^N \int_0^T \partial_{\vartheta_{q'}} (c^{-1/2}b)(\vartheta'; t; X_t^i, \mu_t^{(N)}) dB_t^{i,N,\vartheta'} \right|^r \right] \\ &\leq C \mathbb{E}_{\mathbb{P}_{\vartheta'}^N} \left[\left| N^{-1} \sum_{i=1}^N \int_0^T |\partial_{\vartheta_{q'}} (c^{-1/2}b)(\vartheta'; t; X_t^i, \mu_t^{(N)})|^2 dt \right|^{r/2} \right] \\ &\leq CN^{-1} \sum_{i=1}^N \int_0^T (1 + \mathbb{E}_{\mathbb{P}_{\vartheta'}^N} [|X_t^i|^{r_1 r}]) + \mathbf{m}_{r_2}(\mu_t^{(N)})^r dt \\ &\leq C(1 + \int_0^T \mathbb{E}_{\mathbb{P}_{\vartheta'}^N} [|X_t^i|^{r \max(r_1, r_2)}] dt). \end{aligned}$$

where we succesively used the Burckholder-Davis-Gundy, Assumption 1.4 and the fact that $r \geq 2$. Now, we claim that with $\vartheta' = \vartheta + (N\mathbb{I}(\vartheta))^{-1/2}(u + \lambda(v - u))$, we have

$$\int_0^T \mathbb{E}_{\mathbb{P}_{\vartheta + (N\mathbb{I}(\vartheta))^{-1/2}(u + \lambda(v - u))}^N} [|X_t^i|^{r \max(r_1, r_2)}] dt \leq C(1 + \kappa^\gamma) \quad (1.53)$$

for some $\gamma > 0$, uniformly in $|u|, |v|$ bounded by κ and where C depends on Θ only. Indeed, keeping up with the abbreviation ϑ' , we have

$$\mathbb{E}_{\mathbb{P}_{\vartheta'}} [|X_t^i|^{r \max(r_1, r_2)}] \leq C(\mathbb{E}_{\mathbb{P}_{\vartheta'}} [|X_t^i - X_t^{i,\vartheta}|^{r \max(r_1, r_2)}]) + \mathbb{E}_{\mathbb{P}_{\vartheta'}} [|X_t^{i,\vartheta}|^{r \max(r_1, r_2)}].$$

By (1.43) of Lemma 1.20

$$\mathbb{E}_{\mathbb{P}_{\vartheta'}} [|X_t^i - X_t^{i,\vartheta}|^{r \max(r_1, r_2)}] \leq C|\vartheta - \vartheta'|^{r \max(r_1, r_2)}$$

$$\begin{aligned}
 &= C|(N\mathbb{I}(\vartheta))^{-1/2}(u + \lambda(v - u))|^r \max(r_1, r_2) \\
 &\leq CN^{-r \max(r_1, r_2)/2} \kappa^r \max(r_1, r_2)
 \end{aligned}$$

and the second term that only depends on ϑ by coupling is uniformly bounded by Lemma 1.7. The estimate (1.53) follows. Going back to (1.51), we conclude

$$\mathbb{E}_{\mathbb{P}_\vartheta^N} [|\mathcal{Z}_N(\vartheta; u)^{1/r} - \mathcal{Z}_N(\vartheta; u)^{1/r}|^r] \leq C|u - v|^r (1 + \kappa^{r \max(r_1, r_2)})$$

and (1.50) is established with $\gamma = r \max(r_1, r_2)$ and $\gamma' = r$.

Step 2. Here we prove a moment bound for the likelihood ratio process, namely, for every $r > 0$

$$\mathbb{E}_{\mathbb{P}_\vartheta^N} [\mathcal{Z}_N(\vartheta; u)^{1/2}] \leq C|u|^{-r}, \quad (1.54)$$

uniformly in $\vartheta \in \Theta$ and $u = (N\mathbb{I}(\vartheta))^{1/2}(\vartheta' - \vartheta)$ with $\vartheta' \in \Theta$. Introducing for $t \in [0, T]$ the $\mathbb{P}_\vartheta^N$ -martingale

$$\mathcal{M}_t^N(\vartheta; u) = \sum_{i=1}^N \int_0^t ((c^{-1/2}b)(\vartheta + (N\mathbb{I}(\vartheta))^{-1/2}u; s, X_s^i, \mu_s^{(N)}) - (c^{-1/2}b)(\vartheta; s, X_s^i, \mu_s^{(N)}))^\top dB_s^{i,N,\vartheta},$$

we have

$$\mathcal{Z}_N(\vartheta; u) = \exp(\mathcal{M}_T^N(\vartheta; u) - \frac{1}{2}\langle \mathcal{M}^N(\vartheta; u) \rangle_T).$$

It follows that

$$\begin{aligned}
 \mathbb{E}_{\mathbb{P}_\vartheta^N} [\mathcal{Z}_N(\vartheta; u)^{1/2}] &= \mathbb{E}_{\mathbb{P}_\vartheta^N} \left[\exp(-\frac{1}{2}\mathcal{M}_T^N(\vartheta; u) - \frac{3}{16}\langle \mathcal{M}^N(\vartheta; u) \rangle_T) \exp(-\frac{1}{16}\langle \mathcal{M}^N(\vartheta; u) \rangle_T) \right] \\
 &\leq \mathbb{E}_{\mathbb{P}_\vartheta^N} \left[\exp(-\frac{3}{4}\langle \mathcal{M}^N(\vartheta; u) \rangle_T - \frac{1}{2}\langle \frac{3}{4}\mathcal{M}^N(\vartheta; u) \rangle_T) \right]^{\frac{2}{3}} \mathbb{E}_{\mathbb{P}_\vartheta^N} \left[\exp(-\frac{3}{16}\langle \mathcal{M}^N(\vartheta; u) \rangle_T) \right]^{\frac{1}{3}} \\
 &\leq \mathbb{E}_{\mathbb{P}_\vartheta^N} \left[\exp(-\frac{3}{16}\langle \mathcal{M}^N(\vartheta; u) \rangle_T) \right]^{1/3},
 \end{aligned}$$

thanks to Hölder's inequality and the martingale property of $(\frac{3}{4}\mathcal{M}_t^N(\vartheta; u))_{t \in [0, T]}$. With the help of the parametrisation $u = (N\mathbb{I}(\vartheta))^{1/2}(\vartheta' - \vartheta)$, we rewrite $\langle \mathcal{M}^N(\vartheta; u) \rangle_T$ as

$$\begin{aligned}
 &\sum_{i=1}^N \int_0^T |(c^{-1/2}b)(\vartheta'; t, X_t^i, \mu_t^{(N)}) - (c^{-1/2}b)(\vartheta; t, X_t^i, \mu_t^{(N)})|^2 dt \\
 &= \sum_{i=1}^N \sum_{j=1}^d \int_0^T (((N\mathbb{I}(\vartheta))^{-1/2}u)^\top \int_0^1 \nabla_\vartheta(c^{-1/2}b)^j(\vartheta + \lambda(\vartheta' - \vartheta); t, X_t^i, \mu_t^{(N)}) d\lambda)^2 dt \\
 &= u^\top \mathbb{I}(\vartheta)^{-1/2} \Sigma^N([\vartheta, \vartheta']; u) \mathbb{I}(\vartheta)^{-1/2} u,
 \end{aligned}$$

with

$$\Sigma^N([\vartheta, \vartheta']; u) = N^{-1} \sum_{i=1}^N \sum_{j=1}^d \int_0^T \nabla_\vartheta(c^{-1/2}b)^j([\vartheta, \vartheta']; t, X_t^i, \mu_t^{(N)}) (\nabla_\vartheta(c^{-1/2}b)^j([\vartheta, \vartheta']; t, X_t^i, \mu_t^{(N)}))^\top dt$$

which converges by Lemma 1.21 to

$$\Sigma([\vartheta, \vartheta']; u) = \sum_{j=1}^d \int_0^T \int_{\mathbb{R}^d} \nabla_\vartheta(c^{-1/2}b)^j([\vartheta, \vartheta']; t, x, \mu_t^\vartheta) (\nabla_\vartheta(c^{-1/2}b)^j([\vartheta, \vartheta']; t, x, \mu_t^\vartheta))^\top \mu_t(dx) dt$$

in $\mathbb{P}_\vartheta^N$ -probability. Abbreviating further $\tilde{\Sigma}^N(\vartheta; u) = \mathbb{I}(\vartheta)^{-1/2} \Sigma^N([\vartheta, \vartheta']; u) \mathbb{I}(\vartheta)^{-1/2}$ and $\tilde{\Sigma}(\vartheta; u) = \mathbb{I}(\vartheta)^{-1/2} \Sigma([\vartheta, \vartheta']; u) \mathbb{I}(\vartheta)^{-1/2}$, we have

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_\vartheta^N} [\mathcal{Z}_N(\vartheta; u)^{1/2}] &\leq \mathbb{E}_{\mathbb{P}_\vartheta^N} [\exp(-\frac{3}{16} u^\top \tilde{\Sigma}^N(\vartheta; u) u)]^{1/3} \\ &\leq \mathbb{P}_\vartheta^N(u^\top \tilde{\Sigma}^N(\vartheta; u) u \leq \frac{1}{2} u^\top \tilde{\Sigma}(\vartheta; u) u)^{1/3} + \exp(-\frac{1}{32} u^\top \tilde{\Sigma}(\vartheta; u) u). \end{aligned} \quad (1.55)$$

The non-degeneracy assumption (recall Definition 1.14) ensures $u^\top \tilde{\Sigma}(\vartheta; u) u \geq \mathfrak{c} |u|^2$ for some $\mathfrak{c} > 0$ that does not depend on ϑ nor u (but that depends on Θ), hence the remainder term decays faster than any power of $|u|$. The first term in the right-hand side of (1.55) is bounded above by

$$\begin{aligned} \mathbb{P}_\vartheta^N(|u^\top (\tilde{\Sigma}^N(\vartheta; u) - \tilde{\Sigma}(\vartheta; u)) u| \geq \frac{1}{2} u^\top \tilde{\Sigma}(\vartheta; u) u)^{\frac{1}{3}} \\ \leq \mathbb{P}_\vartheta^N(|u^\top (\tilde{\Sigma}^N(\vartheta; u) - \tilde{\Sigma}(\vartheta; u)) u| \geq \frac{1}{2} \mathfrak{c} |u|^2)^{\frac{1}{3}} \\ \leq C |u|^{-\frac{2m}{3}} \mathbb{E}_{\mathbb{P}_\vartheta^N}[|u^\top (\tilde{\Sigma}^N(\vartheta; u) - \tilde{\Sigma}(\vartheta; u)) u|^m]^{\frac{1}{3}} \end{aligned}$$

for every $m > 0$ by Markov's inequality and where we used the non-degeneracy assumption again. For $1 \leq \ell, \ell' \leq p$, introduce

$$\phi_{\ell, \ell'}([\vartheta, \vartheta']; t, x, \nu) = (\partial_{\vartheta_\ell}(c^{-1/2}b)([\vartheta, \vartheta']; t, x, \nu))^\top \partial_{\vartheta_{\ell'}}(c^{-1/2}b)([\vartheta, \vartheta']; t, x, \nu),$$

so that $(\Sigma^N(\vartheta; u) - \Sigma(\vartheta; u))_{\ell, \ell'}$ is simply

$$N^{-1} \sum_{i=1}^N \int_0^T \phi_{\ell, \ell'}([\vartheta, \vartheta']; t, X_t^i, \mu_t^{(N)}) dt - \int_0^T \int_{\mathbb{R}^d} \phi_{\ell, \ell'}([\vartheta, \vartheta']; t, x, \mu_t^\vartheta) \mu_t(dx) dt.$$

By Lemma 1.21, we derive

$$\mathbb{E}_{\mathbb{P}_\vartheta^N}[|(\Sigma^N(\vartheta; u) - \Sigma(\vartheta; u))_{\ell, \ell'}|^m] \leq C N^{-\delta m},$$

therefore

$$\mathbb{E}_{\mathbb{P}_\vartheta^N}[|u^\top (\tilde{\Sigma}^N(\vartheta; u) - \tilde{\Sigma}(\vartheta; u)) u|^m]^{1/3} \leq C |u|^{2m/3} N^{-\delta m/3},$$

and finally

$$\mathbb{P}_\vartheta^N(|u^\top (\tilde{\Sigma}^N(\vartheta; u) - \tilde{\Sigma}(\vartheta; u)) u| \geq \frac{1}{2} u^\top \tilde{\Sigma}(\vartheta; u) u)^{\frac{1}{3}} \leq C N^{-\delta m/3} \quad (1.56)$$

Pick $r \geq 1$. Combining (1.55) and (1.56), we infer

$$|u|^r \mathbb{E}_{\mathbb{P}_\vartheta^N} [\mathcal{Z}_N(\vartheta; u)^{1/2}] \leq C |u|^r N^{-\delta m/3} + |u|^r \exp(-\frac{\mathfrak{c}}{32} |u|^2).$$

For $u = (N \mathbb{I}(\vartheta))^{1/2}(\vartheta' - \vartheta)$ with $\vartheta' \in \Theta$, we have $|u| \leq C N^{1/2}$. The first term in the right-hand side of the previous estimate is thus bounded as soon as $m \geq 3r/(2\delta)$. The second term is bounded uniformly in $|u|$. We thus have established (1.54).

Step 3. We are now ready to apply Theorem III.1.1 of [98] and gather several properties of the maximum likelihood estimator. Note that the continuity of the likelihood function $\vartheta \mapsto \mathcal{L}^N(\vartheta; X^{(N)})$

and the compactness of Θ ensures that a solution $\widehat{\vartheta}_N^{\mathbf{m}1\mathbf{e}}$ to (1.27) exists.

The uniform LAN condition given in the proof of Theorem 1.17 is the Condition N1 of Chapter III [98]. The non-degeneracy assumption (according to Definition (1.14)) is related to the uniform use of Condition N2 of [98]. Step 1 and Step 2 are respectively Condition N3 and N4 of [98]. We may thus apply Theorem III.1.1 of [98] and we readily obtain Statement (i) of Theorem 1.19. Statement (ii) is a consequence of Corollary III.1.1 of [98] while Statement (iii) is a consequence of Theorem III.1.3 of [98].

The proof of Theorem 1.19 is complete.

6 Remaining proofs

6.1 Proof of Proposition 1.9

Anticipating the proof of Lemma 1.20, we prove a slightly stronger result, namely

$$(\mathbb{E}_{\mathbb{P}_\vartheta^N} [|\overline{X}_t^{i,\vartheta} - \overline{X}_t^{i,\vartheta'}|^r])^{1/r} \leq C|\vartheta - \vartheta'| \quad (1.57)$$

for $r \geq 1$. Indeed, Proposition 1.9 is then a consequence of

$$\mathcal{W}_1(\mu_t^\vartheta, \mu_t^{\vartheta'}) \leq \mathbb{E}_{\mathbb{P}_\vartheta^N} [|\overline{X}_t^{i,\vartheta} - \overline{X}_t^{i,\vartheta'}|] \leq (\mathbb{E}_{\mathbb{P}_\vartheta^N} [|\overline{X}_t^{i,\vartheta} - \overline{X}_t^{i,\vartheta'}|^{2r}])^{1/2r},$$

for any $r \geq 1$. From $\overline{X}_0^{i,\vartheta} = \overline{X}_0^{i,\vartheta'}$, we have

$$\overline{X}_t^{i,\vartheta} - \overline{X}_t^{i,\vartheta'} = \int_0^t (b(\vartheta; s, \overline{X}_s^{i,\vartheta}, \mu_s^\vartheta) - b(\vartheta'; s, \overline{X}_s^{i,\vartheta'}, \mu_s^{\vartheta'})) ds + \int_0^t (\sigma(s, \overline{X}_s^{i,\vartheta}) - \sigma(s, \overline{X}_s^{i,\vartheta'})) dB_s^{i,N,\vartheta}.$$

Using Assumption 1.2 and the Burkholder-Davis-Gundy inequality, we infer

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_\vartheta^N} \left[\left| \int_0^t (\sigma(s, \overline{X}_s^{i,\vartheta}) - \sigma(s, \overline{X}_s^{i,\vartheta'})) dB_s^{i,N,\vartheta} \right|^{2r} \right] &\leq C \mathbb{E}_{\mathbb{P}_\vartheta^N} \left[\left(\int_0^t |\overline{X}_s^{i,\vartheta} - \overline{X}_s^{i,\vartheta'}|^2 ds \right)^r \right] \\ &\leq C \mathbb{E}_{\mathbb{P}_\vartheta^N} \left[\int_0^t |\overline{X}_s^{i,\vartheta} - \overline{X}_s^{i,\vartheta'}|^{2r} ds \right] \end{aligned}$$

since $r \geq 1$. Thanks to the smoothness properties of b granted by Assumptions 1.3 and 1.4 and incorporating the previous estimate, we obtain

$$\begin{aligned} &\mathbb{E}_{\mathbb{P}_\vartheta^N} [|\overline{X}_t^{i,\vartheta} - \overline{X}_t^{i,\vartheta'}|^{2r}] \\ &\leq C \int_0^t \mathbb{E}_{\mathbb{P}_\vartheta^N} [|\vartheta - \vartheta'|^{2r} (1 + |\overline{X}_s^{i,\vartheta}|^{2r_1 r} + \mathbf{m}_{r_2}(\mu_s^\vartheta)^{2r}) + |\overline{X}_s^{i,\vartheta} - \overline{X}_s^{i,\vartheta'}|^{2r} + \mathcal{W}_1(\mu_s^\vartheta, \mu_s^{\vartheta'})^{2r}] ds \\ &\leq C(|\vartheta - \vartheta'|^{2r} + \int_0^t (\mathbb{E}_{\mathbb{P}_\vartheta^N} [|\overline{X}_s^{i,\vartheta} - \overline{X}_s^{i,\vartheta'}|^{2r} + \mathcal{W}_1(\mu_s^\vartheta, \mu_s^{\vartheta'})^{2r}] ds) \end{aligned}$$

where we used that $\mathbb{E}_{\mathbb{P}_\vartheta^N} [|\overline{X}_t^{i,\vartheta}|^{r'}] = \mathbf{m}_{r'}(\mu_t^\vartheta)$ is bounded uniformly in $t \in [0, T]$ and $\vartheta \in \Theta$ for all values of $r' \geq 1$ by Lemma 1.8. We obtain (1.57) for $2r$ by Grönwall's lemma, hence for every $r \geq 1$ by Cauchy-Schwarz's inequality. The proposition follows.

6.2 Proof of Proposition 1.10

By Girsanov's theorem,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_\vartheta^{\otimes N}} \left[\log \frac{d\bar{\mathbb{P}}_\vartheta^{\otimes N}}{d\mathbb{P}_\vartheta^{\otimes N}} \right] &= \frac{1}{2} \mathbb{E}_{\mathbb{P}_\vartheta^{\otimes N}} \left[\sum_{i=1}^N \int_0^T |c^{-1/2} b(\vartheta; t, X_t^i, \mu_t^{(N)}) - c^{-1/2} b(\vartheta; t, X_t^i, \mu_t^\vartheta)|^2 dt \right. \\ &\quad \left. \leq \frac{1}{2} \sigma_-^{-2} \sum_{i=1}^N \int_0^T \mathbb{E}_{\mathbb{P}_\vartheta^{\otimes N}} \left[\left| N^{-1} \sum_{j=1}^N (\tilde{b}(\vartheta; t, X_t^i, X_t^j) - \mathbb{E}_{\bar{\mathbb{P}}_\vartheta} [\tilde{b}(\vartheta; t, \zeta, X_t^j)]_{\zeta=X_t^i} \right|^2 \right] dt \right]. \end{aligned}$$

We plan to use the following decomposition

$$\begin{aligned} &N^{-1} \sum_{j=1}^N \tilde{b}(\vartheta; t, X_t^i, X_t^j) - \mathbb{E}_{\bar{\mathbb{P}}_\vartheta} [\tilde{b}(\vartheta; t, \zeta, X_t^j)]_{\zeta=X_t^i} \\ &= N^{-1} (\tilde{b}(\vartheta; t, X_t^i, X_t^i) - \mathbb{E}_{\bar{\mathbb{P}}_\vartheta} [\tilde{b}(\vartheta; t, \zeta, X_t^i)]_{\zeta=X_t^i}) \\ &\quad + \frac{N-1}{N} (N-1)^{-1} \sum_{j=1, j \neq i}^N (\tilde{b}(\vartheta; t, X_t^i, X_t^j) - \mathbb{E}_{\bar{\mathbb{P}}_\vartheta^{\otimes 2}} [\tilde{b}(\vartheta; t, X_t^i, X_t^j) | X_t^i]). \end{aligned}$$

Using the elementary inequality $(a+b)^2 \leq (1+\rho)a^2 + (1+\rho^{-1})b^2$ valid for every $\rho > 0$, we obtain

$$\begin{aligned} &\mathbb{E}_{\mathbb{P}_\vartheta^{\otimes N}} \left[\left| N^{-1} \sum_{j=1}^N \tilde{b}(\vartheta; t, X_t^i, X_t^j) - \mathbb{E}_{\bar{\mathbb{P}}_\vartheta} [\tilde{b}(\vartheta; t, \zeta, X_t^j)]_{\zeta=X_t^i} \right|^2 \right] \\ &\leq (1+\rho^{-1}) N^{-2} \mathbb{E}_{\bar{\mathbb{P}}_\vartheta} [|\tilde{b}(\vartheta; t, X_t^1, X_t^1)|^2] + (1+\rho) \frac{N-1}{N^2} \mathbb{E}_{\bar{\mathbb{P}}_\vartheta^{\otimes 2}} [|\tilde{b}(\vartheta; t, X_t^1, X_t^2)|^2], \end{aligned}$$

and therefore

$$\limsup_{N \rightarrow \infty} \sup_{\vartheta \in \Theta} \mathbb{E}_{\mathbb{P}_\vartheta^{\otimes N}} \left[\log \frac{d\bar{\mathbb{P}}_\vartheta^{\otimes N}}{d\mathbb{P}_\vartheta^{\otimes N}} \right] \leq \frac{1+\rho}{2} \sigma_-^{-2} \sup_{\vartheta \in \Theta} \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} |\tilde{b}(\vartheta; t, x, y)|^2 (\mu_t^\vartheta \otimes \mu_t^\vartheta)(dx, dy) dt.$$

By assumption,

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |\tilde{b}(\vartheta; t, x, y)|^2 (\mu_t^\vartheta \otimes \mu_t^\vartheta)(dx, dy) \leq C \left(1 + \sup_{t \in [0, T], \vartheta \in \Theta} (\mathfrak{m}_{r_1}(\mu_t^\vartheta) + \mathfrak{m}_{r_2}(\mu_t^\vartheta)) \right)$$

which is finite by Lemma 1.8. We thus obtain (1.15). In order to obtain (1.16), we simply apply Pinsker's inequality :

$$\begin{aligned} \limsup_{N \rightarrow \infty} \sup_{\vartheta \in \Theta} \|\mathbb{P}_\vartheta^N - \bar{\mathbb{P}}_\vartheta^{\otimes N}\|_{TV}^2 &\leq \frac{1}{2} \limsup_{N \rightarrow \infty} \sup_{\vartheta \in \Theta} \mathbb{E}_{\mathbb{P}_\vartheta^{\otimes N}} \left[\log \frac{d\bar{\mathbb{P}}_\vartheta^{\otimes N}}{d\mathbb{P}_\vartheta^{\otimes N}} \right] \\ &\leq \sigma_-^{-2} \frac{1+\rho}{4} \sup_{\vartheta \in \Theta} \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} |\tilde{b}(\vartheta; t, x, y)|^2 (\mu_t^\vartheta \otimes \mu_t^\vartheta)(dx, dy). \end{aligned}$$

The conclusion follows by taking ρ sufficiently small.

6.3 Proof of Proposition 1.11

Let

$$(\phi_{\ell,\ell'}(\vartheta; t, x, \nu))_{1 \leq \ell, \ell' \leq p} = (\partial_{\vartheta_\ell}(c^{-1/2}b)(\vartheta; t, x, \mu_t^\vartheta)^\top \partial_{\vartheta_{\ell'}}(c^{-1/2}b)(\vartheta; t, x, \mu_t^\vartheta))_{1 \leq \ell, \ell' \leq p}.$$

We have

$$\begin{aligned} \mathbb{I}(\vartheta)_{\ell,\ell'} - \mathbb{I}(\vartheta')_{\ell,\ell'} &= \int_0^T (\mathbb{E}_{\mathbb{P}_\vartheta}[\phi_{\ell,\ell'}(\vartheta; t, X_t, \mu_t^\vartheta)] - \mathbb{E}_{\mathbb{P}_{\vartheta'}}[\phi_{\ell,\ell'}(\vartheta'; t, X_t, \mu_t^{\vartheta'})]) dt \\ &= \int_0^T (\mathbb{E}_{\mathbb{P}_\vartheta^N}[\phi_{\ell,\ell'}(\vartheta; t, \bar{X}_t^{1,\vartheta}, \mu_t^\vartheta) - \phi_{\ell,\ell'}(\vartheta'; t, \bar{X}_t^{1,\vartheta'}, \mu_t^{\vartheta'})]) dt. \end{aligned}$$

Thanks to the smoothness properties of b and σ granted by Assumptions 1.2, 1.3 and 1.4, we have

$$\begin{aligned} &|\phi_{\ell,\ell'}(\vartheta; t, \bar{X}_t^{1,\vartheta}, \mu_t^\vartheta) - \phi_{\ell,\ell'}(\vartheta'; t, \bar{X}_t^{1,\vartheta'}, \mu_t^{\vartheta'})| \\ &\leq C(|\vartheta - \vartheta'| + |\bar{X}_t^{1,\vartheta} - \bar{X}_t^{1,\vartheta'}| + \mathcal{W}_1(\mu_t^\vartheta, \mu_t^{\vartheta'}))(1 + |\bar{X}_t^{1,\vartheta}|^\alpha + |\bar{X}_t^{1,\vartheta'}|^\alpha + \mathbf{m}_\beta(\mu_t^\vartheta) + \mathbf{m}_\beta(\mu_t^{\vartheta'})). \end{aligned}$$

We have $\mathbb{E}_{\mathbb{P}_\vartheta^N}[|\bar{X}_t^{1,\vartheta}|^r] = \mathbf{m}_r(\mu_t^\vartheta)$, which is uniformly bounded in $t \in [0, T]$, $\vartheta \in \Theta$ for every $r \geq 1$ by Lemma 1.8. Likewise $\mathbb{E}_{\mathbb{P}_\vartheta^N}[|\bar{X}_t^{1,\vartheta'}|^r] \leq C(\mathbf{m}_r(\mu_t^{\vartheta'}) + \mathbb{E}_{\mathbb{P}_\vartheta^N}[|\bar{X}_t^{1,\vartheta} - \bar{X}_t^{1,\vartheta'}|^r])$, therefore, by Cauchy-Schwarz's inequality

$$\begin{aligned} &|\mathbb{E}_{\mathbb{P}_\vartheta^N}[\phi_{\ell,\ell'}(\vartheta, t, \bar{X}_t^{1,\vartheta}, \mu_t^\vartheta) - \phi_{\ell,\ell'}(\vartheta', t, \bar{X}_t^{1,\vartheta'}, \mu_t^{\vartheta'})]| \\ &\leq C(|\vartheta - \vartheta'| + (\mathbb{E}_{\mathbb{P}_\vartheta^N}[|\bar{X}_t^{1,\vartheta} - \bar{X}_t^{1,\vartheta'}|^2])^{1/2} + \mathcal{W}_1(\mu_t^\vartheta, \mu_t^{\vartheta'}))(1 + (\mathbb{E}_{\mathbb{P}_\vartheta^N}[|\bar{X}_t^{1,\vartheta} - \bar{X}_t^{1,\vartheta'}|^2])^{1/2}) \end{aligned}$$

and the Lipschitz smoothness follows by applying (1.44) of Lemma 1.20 and Proposition 1.9. The convergence $N^{-1}\mathbb{I}_{\mathcal{E}^N}(\vartheta) \rightarrow \mathbb{I}(\vartheta)$ is a simple consequence of Lemma 1.21.

6.4 Proof of Proposition 1.16

Note that for every $1 \leq \ell \leq p$:

$$\partial_{\vartheta_\ell} \ell^N(\vartheta; X^{(N)}) = G^N(X^{(N)})_\ell + 2 \sum_{\ell'=1}^p \vartheta_{\ell'} H^N(X^{(N)})_{\ell,\ell'}.$$

Let ϑ, ϑ' in Θ_0 be such that $\mathbb{P}_\vartheta^N = \mathbb{P}_{\vartheta'}^N$ for every $N \geq 1$. (This also implies $\bar{\mathbb{P}}_\vartheta = \bar{\mathbb{P}}_{\vartheta'}$ by Lemma 1.21 for instance.) Using the symmetry of $H^N(X^{(N)})$, we note that

$$\vartheta^\top H^N(X^{(N)}) \vartheta - (\vartheta')^\top H^N(X^{(N)}) \vartheta' = \sum_{\ell=1}^p (\vartheta_\ell - \vartheta'_\ell) \sum_{\ell'=1}^p H^N(X^{(N)})_{\ell,\ell'} (\vartheta_{\ell'} + \vartheta'_{\ell'}).$$

It follows that

$$0 = \ell^N(\vartheta; X^{(N)}) - \ell^N(\vartheta'; X^{(N)}) = \sum_{\ell=1}^p (\vartheta_\ell - \vartheta'_\ell) \left(G^N(X^{(N)})_\ell + 2 \sum_{\ell'=1}^p H^N(X^{(N)})_{\ell,\ell'} \frac{\vartheta_{\ell'} + \vartheta'_{\ell'}}{2} \right)$$

$$\begin{aligned}
 &= \sum_{\ell=1}^p (G^N(X^{(N)})_{\ell} + 2 \sum_{\ell'=1}^p \frac{\vartheta_{\ell'} + \vartheta'_{\ell'}}{2} H^N(X^{(N)})_{\ell, \ell'}) (\vartheta_{\ell} - \vartheta'_{\ell}) \\
 &= \sum_{\ell=1}^p \partial_{\vartheta_{\ell}} \ell^N(\vartheta^*; X^{(N)}) \xi_{\ell} = \nabla_{\vartheta} \ell^N(\vartheta^*; X^{(N)})^{\top} \xi,
 \end{aligned}$$

with $\xi = \vartheta - \vartheta'$ and $\vartheta^* = \frac{1}{2}(\vartheta + \vartheta') \in \Theta_0$ by the convexity of Θ_0 and that does not depend on $X^{(N)}$. Assume now that $\vartheta \neq \vartheta'$. This implies that for some $\xi \neq 0$, we have

$$0 = (\nabla_{\vartheta} \ell^N(\vartheta^*; X^{(N)})^{\top} \xi)^{\top} \nabla_{\vartheta} \ell^N(\vartheta^*; X^{(N)})^{\top} \xi = \xi^{\top} \mathbb{I}_{\mathcal{E}^N}(\vartheta^*) \xi.$$

Thus $\mathbb{I}_{\mathcal{E}^N}(\vartheta^*)$ is degenerate for every $N \geq 1$. Letting $N \rightarrow \infty$ and applying Proposition 1.11, we infer that $\mathbb{I}(\vartheta^*)$ is degenerate as well, a contradiction.

7 Appendix

7.1 Proof of Lemma 1.7

By Assumption 1.3 we have

$$|b(\vartheta; s, X_s^i, \mu_s^{(N)})| \leq C(1 + |X_s^i| + N^{-1} \sum_{i=1}^N |X_s^i|)$$

uniformly in $s \in [0, T]$ and $\vartheta \in \Theta$. For $M > 0$, define $\tau_M = \inf\{s \geq 0, \max_{1 \leq i \leq N} |X_s^i| \geq M\} \wedge T$ and note that τ_M is a $(\mathcal{F}_t^N)$ -stopping time. We have

$$\begin{aligned}
 |X_{t \wedge \tau_M}^i| &\leq |X_0^i| + \int_0^{t \wedge \tau_M} |b(\vartheta, s, X_s^i, \mu_s^{(N)})| ds + \left| \int_0^{t \wedge \tau_M} \sigma(s, X_s^i) dB_s^{i, N, \vartheta} \right| \\
 &\leq |X_0^i| + C \int_0^{t \wedge \tau_M} (1 + |X_s^i| + N^{-1} \sum_{i=1}^N |X_s^i|) ds + \left| \int_0^{t \wedge \tau_M} \sigma(s, X_s^i) dB_s^{i, N, \vartheta} \right| \\
 &\leq |X_0^i| + C \int_0^t (1 + |X_{s \wedge \tau_M}^i| + N^{-1} \sum_{i=1}^N |X_{s \wedge \tau_M}^i|) ds + \sup_{0 \leq t \leq T} \left| \int_0^t \sigma(s, X_s^i) dB_s^{i, N, \vartheta} \right|
 \end{aligned}$$

Taking $\mathbb{P}_{\vartheta}^N$ -expectation of order $r \geq 1$, we obtain

$$\mathbb{E}_{\mathbb{P}_{\vartheta}^N} [|X_{t \wedge \tau_M}^i|^r] \leq C(1 + \mathbb{E}_{\mathbb{P}_{\vartheta}^N} [|X_0^i|^r] + \int_0^t (1 + \mathbb{E}_{\mathbb{P}_{\vartheta}^N} [|X_{s \wedge \tau_M}^i|^r] ds),$$

using Jensen's inequality, the exchangeability of $\mathbb{P}_{\vartheta}^N$ and the Burckholder-Davis-Gundy inequality together with Assumption 1.2 to obtain

$$\mathbb{E} \left[\sup_{0 \leq t \leq T} \left| \int_0^t \sigma(s, X_s^i) dB_s^{i, N, \vartheta} \right|^r \right] \leq C \mathbb{E} \left[\left(\int_0^t |\sigma(s, X_s^i)|^2 ds \right)^{r/2} \right] \leq C'.$$

By Grönwall's lemma, we infer

$$\mathbb{E}_{\mathbb{P}_{\vartheta}^N} [|X_{t \wedge \tau_M}^i|^r] \leq (\mathbb{E}_{\mathbb{P}_{\vartheta}^N} [|X_0^i|^r] + C') \exp(Ct).$$

Letting $M \rightarrow \infty$, we conclude by Fatou's lemma.

7.2 Proof of Lemma 1.8

It is a slight variation of the proof of Lemma 1.7. By Theorem 4.21 in [37], we have $\mathbb{E}_{\mathbb{P}_\vartheta} [\sup_{0 \leq t \leq T} |X_t^i|^r] < \infty$ for every $r \geq 1$. (Their argument is developed for $r = 2$ but the extension to any $r \geq 1$ is straightforward.) Therefore, only the uniformity in ϑ requires a proof. From

$$\mathbb{E}_{\mathbb{P}_\vartheta} [|\bar{X}_t^{i,\vartheta}|^r] \leq C(\mathbb{E}_{\mathbb{P}_\vartheta} [|X_t^i - \bar{X}_t^{i,\vartheta}|^r] + \mathbb{E}_{\mathbb{P}_\vartheta} [|X_t^i|^r]),$$

and the fact that $\mathbb{E}_{\mathbb{P}_\vartheta} [|\bar{X}_t^{i,\vartheta}|^r] = \mathbb{E}_{\mathbb{P}_\vartheta} [|X_t^{i,\vartheta}|^r]$, Lemma 1.8 is now a simple consequence of (1.46) in Lemma 1.20 together with Lemma 1.7.

7.3 Proof of Lemma 1.20

Proof of (1.43)

The first inequality is obvious. Then, since $X_0^i = X_0^{i,\vartheta'}$, we have

$$X_t^i - X_t^{i,\vartheta'} = \int_0^t (b(\vartheta; s, X_s^i, \mu_s^{(N)}) - b(\vartheta'; s, X_s^{i,\vartheta'}, \mu_s^{(N),\vartheta'})) ds + \int_0^t (\sigma(s, X_s^i) - \sigma(s, X_s^{i,\vartheta'})) dB_s^{i,N,\vartheta}.$$

Thanks to the smoothness properties of b and σ granted by Assumptions 1.2, 1.3 and 1.4, taking expectation to the power r on both side and applying the Burckholder-Davis-Gundy inequality, we infer

$$\begin{aligned} & \mathbb{E}_{\mathbb{P}_\vartheta} [|X_t^i - X_t^{i,\vartheta'}|^r] \\ & \leq C \int_0^t \mathbb{E}_{\mathbb{P}_\vartheta} [|\vartheta - \vartheta'|^r (1 + |X_s^i|^{r_1} + \mathbf{m}_{r_2}(\mu_s^{(N)})^r) + |X_s^i - X_s^{i,\vartheta'}|^r + \mathcal{W}_1(\mu_s^{(N),\vartheta'}, \mu_s^{(N)})^r] ds \\ & \leq C(|\vartheta - \vartheta'|^r + \int_0^t (\mathbb{E}_{\mathbb{P}_\vartheta} [|X_s^i - X_s^{i,\vartheta'}|^r + \mathcal{W}_1(\mu_s^{(N),\vartheta'}, \mu_s^{(N)})^r]) ds) \end{aligned}$$

where we used that $\mathbb{E}_{\mathbb{P}_\vartheta} [\mathbf{m}_{r_2}(\mu_t^{(N)})^r] \leq \mathbb{E}_{\mathbb{P}_\vartheta} [|X_t^i|^{r_2 r}]$ which is bounded uniformly in $t \in [0, T]$ and $\vartheta \in \Theta$ by Lemma 1.7. Also, using the first part of (1.43), namely

$$\mathbb{E}_{\mathbb{P}_\vartheta} [\mathcal{W}_1(\mu_s^{(N),\vartheta'}, \mu_s^{(N)})^r] \leq CN^{-1} \sum_{i=1}^N \mathbb{E}_{\mathbb{P}_\vartheta} [|X_s^i - X_s^{i,\vartheta'}|^r]$$

and taking averages over $i = 1, \dots, N$ on both sides, we infer

$$N^{-1} \sum_{i=1}^N \mathbb{E}_{\mathbb{P}_\vartheta} [|X_t^i - X_t^{i,\vartheta'}|^r] \leq C(|\vartheta - \vartheta'|^r + \int_0^t N^{-1} \sum_{i=1}^N \mathbb{E}_{\mathbb{P}_\vartheta} [|X_s^i - X_s^{i,\vartheta'}|^r] ds).$$

We obtain the second part of (1.43) by Grönwall's lemma.

Proof of (1.44)

The first inequality is obvious. The second part is simply (1.57) from the proof of Proposition 1.9.

Proof of (1.45) and (1.46)

By triangle inequality,

$$\mathcal{W}_1(\mu_t^{(N)}, \mu_t^\vartheta) \leq \mathcal{W}_1(\mu_t^{(N)}, \bar{\mu}_t^{(N), \vartheta}) + \mathcal{W}_1(\bar{\mu}_t^{(N), \vartheta}, \mu_t^\vartheta) \leq N^{-1} \sum_{i=1}^N |X_t^i - \bar{X}_t^{i, \vartheta}| + \mathcal{W}_1(\bar{\mu}_t^{(N), \vartheta}, \mu_t^\vartheta). \quad (1.58)$$

By Theorem 2 of [74], we have $\sup_{t \in [0, T], \vartheta \in \Theta} \mathbb{E}_{\mathbb{P}_\vartheta^N} [\mathcal{W}_1(\bar{\mu}_t^{(N), \vartheta}, \mu_t^\vartheta)^r] \leq CN^{-\delta r}$ for every $r \geq 1$ and some $\delta > 0$. The value of δ depends on the dimension d of the state space. The uniformity in (t, ϑ) follows in particular from the uniform moment bounds of Lemma 1.8 (see the conditions of Theorem 2 of [74]). Therefore (1.45) is a consequence of (1.46).

In order to establish (1.46), since $X_0^i = \bar{X}_0^{i, \vartheta}$, we write

$$X_t^i - \bar{X}_t^{i, \vartheta} = \int_0^t (b(\vartheta; X_s^i, \mu_s^{(N)}) - b(\vartheta; s, \bar{X}_s^{i, \vartheta}, \mu_s^\vartheta)) ds + \int_0^t (\sigma(s, X_s^i) - \sigma(s, \bar{X}_s^{i, \vartheta})) dB_s^{i, N, \vartheta}.$$

Taking expectation to the power r on both side and applying the Burckholder-Davis-Gundy inequality, we infer

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_\vartheta^N} [|X_t^i - \bar{X}_t^{i, \vartheta}|^r] &\leq C \int_0^t \mathbb{E}_{\mathbb{P}_\vartheta^N} [|X_s^i - \bar{X}_s^{i, \vartheta}|^r + \mathcal{W}_1(\mu_s^{(N)}, \mu_s^\vartheta)^r] ds \\ &\leq C \int_0^t \mathbb{E}_{\mathbb{P}_\vartheta^N} [|X_s^i - \bar{X}_s^{i, \vartheta}|^r + \mathcal{W}_1(\bar{\mu}_s^{(N), \vartheta}, \mu_s^\vartheta)^r] ds \\ &\leq \varepsilon_N + C \int_0^t \mathbb{E}_{\mathbb{P}_\vartheta^N} [|X_s^i - \bar{X}_s^{i, \vartheta}|^r] ds, \end{aligned}$$

arguing as in (1.58), and where $\varepsilon_N = CT \sup_{t \in [0, T], \vartheta \in \Theta} \mathbb{E}_{\mathbb{P}_\vartheta^N} [\mathcal{W}_1(\bar{\mu}_s^{(N), \vartheta}, \mu_s^\vartheta)^r] \leq CN^{-\delta r}$. Note that the constants in each line are uniformly bounded in $\vartheta \in \Theta$. We obtain (1.46) by Grönwall's lemma.

7.4 Proof of Lemma 1.21

We plan to use the following decomposition

$$N^{-1} \sum_{i=1}^N \int_0^t \phi(\vartheta_N; s, X_s^i, \mu_s^{(N)}) ds - \int_0^t \int_{\mathbb{R}^d} \phi(\vartheta; s, x, \mu_s^\vartheta) \mu_s^\vartheta(dx) ds = I + II + III,$$

with

$$\begin{aligned} I &= N^{-1} \sum_{i=1}^N \int_0^t (\phi(\vartheta_N; s, X_s^i, \mu_s^{(N)}) - \phi(\vartheta_N; s, \bar{X}_s^{i, \vartheta_N}, \mu_s^{\vartheta_N})) ds \\ II &= N^{-1} \sum_{i=1}^N \int_0^t (\phi(\vartheta_N; s, \bar{X}_s^{i, \vartheta_N}, \mu_s^{\vartheta_N}) - \phi(\vartheta; s, \bar{X}_s^{i, \vartheta}, \mu_s^\vartheta)) ds \\ III &= N^{-1} \sum_{i=1}^N \int_0^t \phi(\vartheta; s, \bar{X}_s^{i, \vartheta}, \mu_s^\vartheta) ds - \mathbb{E}_{\mathbb{P}_\vartheta} \left[\int_0^t \phi(\vartheta; s, \bar{X}_s^{i, \vartheta}, \mu_s^\vartheta) ds \right]. \end{aligned}$$

Thanks to the properties of ϕ the term I is bounded by a constant times

$$\begin{aligned}
 & N^{-1} \sum_{i=1}^N \int_0^t (|X_s^i - \bar{X}_s^{i, \vartheta_N}| + \mathcal{W}_1(\mu_s^{(N)}, \mu_s^{\vartheta_N})) (1 + |\bar{X}_s^{i, \vartheta_N}|^\alpha + |X_s^i|^\alpha + \mathfrak{m}_\beta(\mu_s^{\vartheta_N}) + \mathfrak{m}_\beta(\mu_s^{(N)})) ds \\
 & \leq \left(N^{-1} \sum_{i=1}^N \int_0^t (|X_s^i - \bar{X}_s^{i, \vartheta_N}| + \mathcal{W}_1(\mu_s^{(N)}, \mu_s^{\vartheta_N}))^2 ds \right)^{1/2} \\
 & \quad \times C \left(N^{-1} \sum_{i=1}^N \int_0^t (1 + |\bar{X}_s^{i, \vartheta_N}|^{2\alpha} + |X_s^i|^{2\alpha} + \mathfrak{m}_{2\beta}(\mu_s^{\vartheta_N}) + \mathfrak{m}_{2\beta}(\mu_s^{(N)})) ds \right)^{1/2}.
 \end{aligned}$$

by Cauchy-Schwarz's inequality. Applying Cauchy-Schwarz's inequality again together with Jensen's inequality, the $\mathbb{P}_{\vartheta_N}^N$ -expectation to the power m of the first term is then bounded by a constant times

$$\begin{aligned}
 & \mathbb{E}_{\mathbb{P}_{\vartheta_N}^N} \left[\left| N^{-1} \sum_{i=1}^N \int_0^t (|X_s^i - \bar{X}_s^{i, \vartheta_N}|^2 + \mathcal{W}_1(\mu_s^{(N)}, \mu_s^{\vartheta_N})^2) ds \right|^m \right]^{1/2} \\
 & \quad \times \left(\int_0^t (1 + \mathbb{E}_{\mathbb{P}_{\vartheta_N}^N} [|\bar{X}_s^{i, \vartheta_N}|^{2m\alpha} + |X_s^i|^{2m\alpha}] + \mathfrak{m}_{2m\beta}(\mu_s^{\vartheta_N}) + \mathfrak{m}_{2m\beta}(\mu_s^{(N)})) ds \right)^{1/2}.
 \end{aligned}$$

The first term is bounded by a constant times $N^{-m\delta}$ by (1.45) and (1.46) of Lemma 1.20. Also the $\mathbb{P}_{\vartheta_N}^N$ -expectation of $|\bar{X}_s^{i, \vartheta_N}|^{2m\alpha}$, $|X_s^i|^{2m\alpha}$ and $\mathfrak{m}_{2m\beta}(\mu_s^{(N)})$ is uniformly bounded in $s \in [0, T]$ by Lemma 1.7 and so is $\mathfrak{m}_{2m\beta}(\mu_s^{\vartheta_N})$ by Lemma 1.8. We conclude

$$\mathbb{E}_{\mathbb{P}_{\vartheta_N}^N} [|I|^m] \leq CN^{-\delta m}.$$

The second term II is bounded by a constant times

$$N^{-1} \sum_{i=1}^N \int_0^t (|\vartheta_N - \vartheta| + |\bar{X}_s^{i, \vartheta_N} - \bar{X}_s^{i, \vartheta}| + \mathcal{W}_1(\mu_s^{\vartheta_N}, \mu_s^{\vartheta})) ds.$$

Taking $\mathbb{P}_{\vartheta_N}^N$ -expectation to the power m and applying successively the first and second part of (1.44) in Lemma 1.20, we obtain

$$\mathbb{E}_{\mathbb{P}_{\vartheta_N}^N} [|II|^m] \leq C|\vartheta - \vartheta_N|^m.$$

Finally, the third and last term converges to 0 by the law of large numbers, applying for instance Rosenthal's inequality for a precise bound in N . We obtain

$$\mathbb{E}_{\mathbb{P}_{\vartheta_N}^N} [|III|^m] \leq CN^{-m/2}.$$

The proof of Lemma 1.21 is complete.

Acknowledgements We are grateful to the careful comments of two anonymous referees that help to improve a former version of the manuscript.

CHAPITRE 2

NONPARAMETRIC ESTIMATION FOR INTERACTING PARTICLE SYSTEMS : MCKEAN-VLASOV MODELS

1 Introduction

1.1 Setting

We continuously observe a stochastic system of N interacting particles

$$X_t = (X_t^1, \dots, X_t^N), \quad t \in [0, T] \quad (2.1)$$

evolving in an Euclidean ambient space $\mathbb{R}^d$, that solves

$$\begin{cases} dX_t^i = b(t, X_t^i, \mu_t^N)dt + \sigma(t, X_t^i)dB_t^i, & 1 \leq i \leq N, \quad t \in [0, T], \\ \mathcal{L}(X_0^1, \dots, X_0^N) = \mu_0^{\otimes N}, \end{cases} \quad (2.2)$$

where $\mu_t^N = N^{-1} \sum_{i=1}^N \delta_{X_t^i}$ is the empirical measure of the system. The B^i are independent $\mathbb{R}^d$ -valued Brownian motions, and the transport and diffusion coefficients b and σ are sufficiently regular so that $\mu_t^N \rightarrow \mu_t$ weakly as $N \rightarrow \infty$, where μ_t is a weak solution of the parabolic nonlinear equation

$$\begin{cases} \partial_t \mu_t + \operatorname{div}(b(t, \cdot, \mu_t)\mu_t) = \frac{1}{2} \sum_{k,k'=1}^d \partial_{kk'}^2 ((\sigma \sigma^\top)(t, \cdot)_{kk'} \mu_t), \\ \mu_{t=0} = \mu_0, \end{cases} \quad (2.3)$$

see Section 2 below. In this context, we are interested in estimating nonparametrically from data (2.1) the solution $(t, x) \mapsto \mu_t(x)$ of (2.3) and the drift function $(t, x, \mu) \mapsto b(t, x, \mu) \in \mathbb{R}^d$ at the value $(t, x, \mu) = (t, x, \mu_t)$. The time horizon T is fixed and asymptotics are taken as $N \rightarrow \infty$.

A particular case of interest is a homogeneous drift with a linear dependence in the measure argument. The drift term in (2.2) then takes the form

$$b(t, X_t^i, \mu_t^N) = \int_{\mathbb{R}^d} \tilde{b}(X_t^i, y) \mu_t^N(dy) = N^{-1} \sum_{j=1}^N \tilde{b}(X_t^i, X_t^j), \quad (2.4)$$

for some function $\tilde{b} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$. In the paper, when we specialise on this case, we consider $\tilde{b}$ of the form

$$\tilde{b}(x, y) = F(x - y) + G(x)$$

for some regular $F, G : \mathbb{R}^d \rightarrow \mathbb{R}^d$. In this case, we have

$$b(t, X_t^i, \mu_t^N) = F \star \mu_t^N(X_t^i) + G(X_t^i),$$

where $\star$ denotes convolution. The function F plays the role of an interaction force applied to the particle system while G accounts for an external force in the motion of each particle. If the forces F and G derive from smooth potentials $V, W : \mathbb{R}^d \rightarrow \mathbb{R}$, the interaction then takes the form $F(x) = -\nabla W(x)$ and we have a confinement $G(x) = -\nabla V(x)$, see *e.g.* [12, 92, 38]. However, we work on a fixed time horizon $[0, T]$ in the paper, and will not need this point of view¹.

In the semi-linear representation (2.4), we are interested in estimating nonparametrically from data (2.1) the interaction force $x \mapsto F(x)$; the parameter $x \mapsto G(x)$ is considered as a nuisance.

1.2 Motivation

Stochastic systems of interacting particles and associated nonlinear Markov processes in the sense of McKean [146] date back to the 1960's and originated from statistical physics in plasma physics. Their importance in probability theory progressively grew in the following decades, and a versatility of fundamental tools were developed in this context like *e.g.* coupling methods, geometric inequalities, propagation of chaos, concentration and fluctuations in abstract functional spaces, see Sznitman [172, 173], Tanaka and Hitsuda [179], Fernandez *et al.* [71], Méléard [148], Malrieu [142], Cattiaux *et al.* [38], Bolley *et al.* [21], among a myriad of references. However, until the early 2000's, a modern formulation of a statistical inference program in this context was out of reach (with some notable exceptions like *e.g.* Kasonga [111]), at least for two reasons : first, the fine probabilistic tools required for nonparametric adaptive estimation were still in full development ; second and perhaps more importantly, microscopic particles systems issued from statistical physics are not naturally observable and the motivation for statistical inference is not obvious in this context. The situation progressively evolved around the 2010's, with the start of a kind of renaissance of McKean-Vlasov type models in several application fields that model collective and observable dynamics, ranging from mathematical biology (neurosciences, Baladron *et al.* [6], structured models in population dynamics, Mogilner *et al.* [154], Burger *et al.* [32]) to social sciences (opinion dynamics, Chazelle *et al.* [43], cooperative behaviours, Canuto *et al.* [33]) and finance (systemic risk, Fouque and Sun [73]). More recently, mean-field games (Cardaliaguet *et al.* [35], Cardaliaguet and Lehalle [36]) appear as a new frontier for statistical developments, see in particular the recent contribution of Giesecke [84]. The field has reached enough maturity for the necessity and interest of a systematic statistical inference program, starting with nonparametric estimation. This is the topic of the paper.

Parallel to understanding collective dynamics from a statistical point of view, some interest in the study of statistical models related to PDE's has progressively emerged over the last decade. Typical examples include nonparametric Bayes and uncertainty quantification for inverse problems, see Abraham and Nickl [1], Monard *et al.* [155], Nickl [159, 160], and the references therein, or inference in structured models from microscopic data (Doumic *et al.* [67, 68], Hoffmann and Olivier [95], Boumezoued *et al.* [27], Ngoc *et al.* [94], Maïda *et al.* [141]). The analysis of elliptic, parabolic

1. usually required to control the model for convergence to equilibrium when T is large.

or transport-fragmentation equations sheds new light on the underlying nonparametric structure of companion statistical experiments and enrich the classical theory. To that extent, we provide in this paper a first step in that direction for a certain kind of nonlinear parabolic equations, in the sense of McKean [146]. Finally, our work can also be embedded in the framework of functional data analysis, where we observe N diffusion processes with common dynamics, see *e.g.* the recent work of Comte and Genon-Catalot [48] formally contained in our framework, for $d = 1$ and in absence of interactions in the drift b .

1.3 Results and organisation of the paper

In Section 2, we detail the notations and assumptions of the model and build kernel estimators for $\mu_t(x)$ and $b(t, x, \mu_t)$. Whereas the estimation for μ_t is standard, the estimation of the drift requires a smoothing in both time and space of the empirical measure $\pi^N(dt, dx) = N^{-1} \sum_{i=1}^N \delta_{X_t^i}(dx) X_t^i(dt)$, that estimates the intermediate function $\pi(t, x) = b(t, x, \mu_t)\mu_t(x)$. We then use a quotient estimator to recover b .

In Section 3, we adopt the Goldenshluger-Lepski method [87, 88, 89] to tune the bandwidths of both estimators in a data driven way and obtain oracle inequalities in Theorems 2.7 and 2.9 for both μ and b . We further develop a minimax theory in Section 4 when μ and b belong to anisotropic Hölder spaces in time and space variable, that are built upon the solution of the parabolic nonlinear limiting equation (2.3) and prove the optimality and smoothness adaptivity of our estimators in Theorems 2.14 and 2.15. We finally explore in Section 5 the identification of the interaction force F in the Vlasov model where the drift takes the form $b(x, \mu) = F \star (x)\mu + G(x)$, for some sufficiently well localised functions F and G . We prove in Theorem 2.17 that one can consistently estimate F (hence G) by means of a Fourier type estimator, inspired by blind deconvolution methods, see *e.g.* Johannes [106].

We develop the probabilistic tools we need to undertake our statistical estimates in Section 6. We study the fluctuations of μ_t^N around its mean μ_t in Theorem 2.18, with time dependent extension to the fluctuations of $\mu_t^N(dx)\rho(dt)$ around $\nu(dt, dx) = \mu_t(dx)\rho(dt)$ for arbitrary weight measures $\rho(dt)$. We prove a Bernstein concentration inequality that reads

$$\begin{aligned} \text{Prob}\left(N^{-1} \sum_{i=1}^N \int_0^T \phi(t, X_t^i) \rho(dt) - \int_{[0,T] \times \mathbb{R}^d} \phi(t, y) \mu_t(dy) \rho(dt) \geq x\right) \\ \leq \kappa_1 \exp\left(-\frac{\kappa_2 N x^2}{|\phi|_{L^2(\nu)}^2 + |\phi|_{\infty} x}\right), \quad \forall x \geq 0, \end{aligned}$$

over test functions ϕ and for some $\kappa_i = \kappa_i(T, \sigma, b, \mu_0) > 0$. It improves on variance estimates based on coupling or geometric inequalities that usually need ϕ to be 1-Lipschitz, whereas nonparametric statistical estimation requires $\phi = \phi_N$ to mimic a Dirac mass as $N \rightarrow \infty$ that can be controlled in L^2 -norm but behaves badly in Lipschitz norm. Bernstein's inequality for a range of deviation valid for all $x \geq 0$ is also the gateway to nonparametric adaptive estimation; it is not provided by concentration in Wasserstein distance like in Bolley *et al.* [21] that moreover has the drawback of adding an additional unavoidable dimensional penalty in the rates of convergence.

Our method of proof avoids coupling by relying on a Girsanov argument, following classical ideas, recently revisited for instance by Lacker [116], a key reference for our work. We classically

require strong ellipticity for the diffusion coefficient and Lipschitz continuity in the space variable. As for the drift, we assume at least Lipschitz continuity

$$|b(t, x, \mu) - b(t, x', \mu')| \leq C(|x - x'| + \mathcal{W}_1(\mu, \mu'))$$

in Wasserstein-1 metric $\mathcal{W}_1$. In our approximation argument, the logarithm of the Girsanov density between the law of the data and a companion coupled system of independent particles is of order $N \int_0^T \int_{\mathbb{R}^d} |b(s, x, \mu_s^N) - b(s, x, \mu_s)|^2 \mu_s^N(dx) ds \lesssim N \sup_{0 \leq t \leq T} \mathcal{W}_1(\mu_t^N, \mu_t)^2$ for which we need sharp integrability properties uniformly in N . In order to circumvent the unavoidable dimensional effect of the approximation $\mathcal{W}_1(\mu_t^N, \mu_t) \approx N^{-1/\max(2, d)}$, see Fournier and Guillin [74], we assume moreover k -linear differentiability for b in the measure variable, see Assumption 2.4 in Section 2. This enables us to control by a sub-Gaussianity argument for U -statistics each term of a Taylor-like expansion of $b(t, x, \mu_t^N) - b(t, x, \mu_t)$ in the measure variable; we obtain the desired result, provided $k \geq d/2$ (with a special modification in dimension $d = 2$). In particular, in the Vlasov model, we formally have $k = \infty$ and the result is valid in all dimension $d \geq 1$.

Section 6 and 7 are devoted to the proofs and an Appendix (Section 8) contains auxiliary technical results.

2 Model assumptions and construction of estimators

2.1 The system of interacting particles and its limit

We fix an integer $d \geq 1$ and a time horizon $T > 0$. The random processes take their values in $\mathbb{R}^d$. We write $|\cdot|$ for the Euclidean distance (on $\mathbb{R}$ or $\mathbb{R}^d$) or sometimes for the modulus of a complex number, and $x^\top x = |x|^2$.

Functions

We consider functions that are mappings defined on products of metric spaces (typically $[0, T] \times \mathbb{R}^d \times \mathcal{P}_1$ or subsets of these) with values in $\mathbb{R}$ or $\mathbb{R}^d$. Here, $\mathcal{P}_1$ denotes the set of probability measures on $\mathbb{R}^d$ with a first moment, endowed with the Wasserstein 1-metric

$$\mathcal{W}_1(\mu, \nu) = \inf_{m \in \Gamma(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y| m(dx, dy) = \sup_{|\phi|_{\text{Lip}} \leq 1} \int_{\mathbb{R}^d} \phi d(\mu - \nu),$$

where $\Gamma(\mu, \nu)$ denotes the set of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν . All the functions in the paper are implicitly measurable with respect to the Borel-sigma field induced by the product topology. A $\mathbb{R}^d$ -valued function f is written componentwise as $f = (f^k)_{1 \leq k \leq d}$ where the f^k are real-valued. The product $f \otimes g$ of two $\mathbb{R}^d$ -valued functions is the $\mathbb{R}^{2d}$ -valued function with $\mathbb{R}^{2d}$ variables with components $(f \otimes g)^k(x, y) = f^k(x)g^k(y)$. If $f : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$, we set $|f|_\infty = \sup_{t, x} |f(t, x)|$ and $|f|_p = (\int_{[0, T] \times \mathbb{R}^d} |f(t, x)|^p dx dt)^{1/p}$ for $1 \leq p < \infty$. Depending on the context, if $f : [0, T] \rightarrow \mathbb{R}$ or $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is a function of time or space only, we sometimes write $|f|_p$ for $(\int_0^T |f(t)|^p dt)^{1/p}$ or $(\int_{\mathbb{R}^d} |f(x)|^p dx)^{1/p}$ when no confusion is possible.

Constants

We repeatedly use positive quantities $\kappa_i, \varpi_i, C_i, i = 1, 2, \dots$ that do not depend on N , that we call constants, but that actually may (continuously) depend on model parameters. In most

cases, they are explicitly computable. We also use special letters like $\kappa, \delta, \tau, c_{\pm}, \dots$, but they will only appear once. The generic notation C is sometimes used before it is set depending on a model parameter. The notation ϖ_i stands for quantities that need to be tuned in an algorithm (like an estimator).

Assumptions

We work under strong ellipticity and Lipschitz smoothness assumptions on the diffusion matrix $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$ and the drift $b : [0, T] \times \mathbb{R}^d \times \mathcal{P}_1 \rightarrow \mathbb{R}^d$, as well as strong integrability properties for the initial condition μ_0 .

Assumption 2.1. *For some $\gamma_0 > 0, \gamma_1 \geq 1$, the initial condition μ_0 satisfies*

$$\int_{\mathbb{R}^d} \exp(\gamma_0 |x|^2) \mu_0(dx) \leq \gamma_1. \quad (2.5)$$

Assumption 2.2. *The diffusion matrix σ is measurable and for some $C \geq 0$, we have*

$$|\sigma(t, x') - \sigma(t, x)| \leq C|x' - x|.$$

Moreover, $c = \sigma \sigma^\top$ is such that $\sigma_-^2 |y|^2 \leq (c(t, x)y)^\top y \leq \sigma_+^2 |y|^2$ for some $\sigma_{\pm} > 0$.

As for the regularity of the drift

$$b : [0, T] \times \mathbb{R}^d \times \mathcal{P}_1 \rightarrow \mathbb{R}^d,$$

the notion of linear differentiability, commonly used in the literature of mean-field games and McKean-Vlasov equations in order to quantify the smoothness of $\mu \mapsto b(t, x, \mu)$ as a mapping $\mathcal{P}_1 \rightarrow \mathbb{R}^d$ will be the most useful in our setting. We refer in particular to the illuminating section 2.2. in Jourdain and Tse [108] and the references therein.

Definition 2.3. *A mapping $f : \mathcal{P}_1 \rightarrow \mathbb{R}^d$ is said to have a linear functional derivative, if there exists $\delta_\mu f : \mathbb{R}^d \times \mathcal{P}_1 \rightarrow \mathbb{R}^d$ (sometimes denoted by $\frac{\delta f}{\delta \mu}$) such that*

$$f(\mu') - f(\mu) = \int_0^1 \int_{\mathbb{R}^d} \delta_\mu f(y, (1-\vartheta)\mu + \vartheta\mu') (\mu' - \mu)(dy) d\vartheta \quad (2.6)$$

with the following smoothness properties

$$\begin{aligned} |\delta_\mu f(y', \mu') - \delta_\mu f(y, \mu)| &\leq C(\mathcal{W}_1(\mu', \mu) + |y' - y|), \\ |\partial_y(\delta_\mu f(y, \mu') - \delta_\mu f(y, \mu))| &\leq C\mathcal{W}_1(\mu', \mu) \end{aligned}$$

for some $C \geq 0$.

We can iterate the process described in (2.6) and obtain a notion of k -linear functional derivative via the existence of mappings

$$\delta_\mu^\ell f : (\mathbb{R}^d)^\ell \times \mathcal{P}_1 \rightarrow \mathbb{R}^d \text{ for } \ell = 1, \dots, k$$

defined recursively by $\delta_\mu^\ell f = \delta_\mu \circ \delta_\mu^{\ell-1} f$ and enjoying associated smoothness properties.

Assumption 2.4. *The drift $b : [0, T] \times \mathbb{R}^d \times \mathcal{P}_1 \rightarrow \mathbb{R}^d$ is measurable and*

$$b_0 = \sup_{t \in [0, T]} |b(t, 0, \delta_0)| < \infty.$$

Moreover, one of the following three conditions is satisfied for some $C > 0$:

(i) *(Lipschitz continuity.) We have $d = 1$ and*

$$|b(t, x', \mu') - b(t, x, \mu)| \leq C(|x' - x| + \mathcal{W}_1(\mu', \mu)).$$

(ii) *(Existence of a functional derivative of order k .) Let $k \geq 1$. For $(d = 1 \text{ and } k \geq 1)$ or $(d = 2 \text{ and } k \geq 2)$ or $(d \geq 3 \text{ and } k \geq d/2)$, we have (i) and the map*

$$\mu \mapsto b(t, x, \mu)$$

admits a functional derivative of order k in the sense of Definition 2.3. Moreover, the following representation holds

$$\delta_\mu^k b(t, x, (y_1, \dots, y_k), \mu) = \sum_{\mathcal{J} \subset \{1, \dots, k\}, m \geq 1} \bigotimes_{j \in \mathcal{J}} (\delta_\mu^k b)_{\mathcal{J}, j, m}(t, x, y_j, \mu), \quad (2.7)$$

where the sum in m is finite with at most m_b terms and the mappings $(\delta_\mu^k b)_{\mathcal{J}, j, m} : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_1 \rightarrow \mathbb{R}^d$ are such that

$$|(\delta_\mu^k b)_{\mathcal{J}, j, m}(t, x', y', \mu) - (\delta_\mu^k b)_{\mathcal{J}, j, m}(t, x, y, \mu)| \leq C(|x' - x| + |y' - y|).$$

(iii) *(Vlasov case.) We have $d \geq 1$ and*

$$b(t, x, \mu) = \int_{\mathbb{R}^d} \tilde{b}(t, x, y) \mu(dy)$$

for some measurable $\tilde{b} : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that :

$$|\tilde{b}(t, x', y') - \tilde{b}(t, x, y)| \leq C(|x' - x| + |y' - y|).$$

We let $|b|_{\text{Lip}}$ denote the smallest $C \geq 0$ for which Assumption 2.4 (i) holds and $|\delta_\mu^k b|_{\text{Lip}}$ the smallest constant $C \geq 0$ for which Assumption 2.4 (ii) holds for the highest order of differentiability.

Remark 2.5. 1) *The representation (2.7) in Assumption 2.4 (ii) is merely technical and enables one to obtain a control in $\mathcal{W}_1$ -distance of the remainder term in Taylor-like expansions of $b(t, x, \mu)$ in an easy way, see in particular the proof of Proposition 2.19, Step 2 below. It can presumably be relaxed, but will be sufficient for the level of generality intended in the paper. It accomodates in particular drifts of the form*

$$b(t, x, \mu) = \sum_j F_j \left(t, x, \int_{\mathbb{R}^{q_2}} G_j(t, x, \cdot) H_j(t, x, z) \mu(dz), z' \right) \lambda_j(dz')$$

for smooth mappings $F_j(t, x, \cdot) : \mathbb{R}^{q_3} \rightarrow \mathbb{R}^d$, $G_j(t, x, \cdot) : \mathbb{R}^{q_1} \times \mathbb{R}^{q_2} \rightarrow \mathbb{R}^{q_3}$, $H_j(t, x, \cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^{q_1}$ and positive measures λ_j on $\mathbb{R}^{q_2}$ in some cases and combinations of these, see Jourdain and Tse [108]. Explicit examples of mean-field models where the structure of the drift is of the form 2.4 (ii) rather

than 2.4 (i) or (iii) are given for instance in [46, 162, 148, 107]. **2)** Condition (iii) is stronger than (ii) : under Assumption 2.4 (iii),

$$\int_{\mathbb{R}^d} \tilde{b}(t, x, \cdot) d(\mu' - \mu) \leq C \sup_{|\phi|_{\text{Lip}} \leq 1} \int_{\mathbb{R}^d} \phi d(\mu' - \mu) = C\mathcal{W}_1(\mu', \mu),$$

thus $|b(t, x', \mu') - b(t, x, \mu)| = |\int_{\mathbb{R}^d} (\tilde{b}(t, x', \cdot) - \tilde{b}(t, x, \cdot)) d\mu' - \int_{\mathbb{R}^d} \tilde{b}(t, x, \cdot) d(\mu - \mu')| \leq C(|x' - x| + \mathcal{W}_1(\mu', \mu))$ and Assumption 2.4 (i) holds true. Moreover $\delta_\mu b(t, x, y, \mu) = \tilde{b}(t, x, y)$ and Assumption 2.4 (ii) holds true as well.

We let $\mathcal{C} = \mathcal{C}([0, T], (\mathbb{R}^d)^N)$ denote the space of continuous functions on $(\mathbb{R}^d)^N$, equipped with the filtration $(\mathcal{F}_t)_{0 \leq t \leq T}$ induced by our observation, namely the canonical mappings

$$X_t(\omega) = (X_t^1(\omega), \dots, X_t^N(\omega)) = \omega_t$$

and modified to be right-continuous for safety. For $\mu_0 \in \mathcal{P}_1$, the probability $\mathbb{P}^N$ on $(\mathcal{C}, \mathcal{F}_T)$ under which the canonical process $X = (X_t^1, \dots, X_t^N)_{0 \leq t \leq 1}$ is a weak solution of (2.2) for the initial condition $\mu_0^{\otimes N}$ is uniquely defined under Assumptions 2.1, 2.2 and 2.4. Recommended reference (that covers our set of assumptions) is the textbook by Carmona and Delarue [37] or the lectures notes of Lacker [115]).

2.2 Kernel estimators

We pick two bounded and compactly supported kernel functions $H : (0, T) \rightarrow \mathbb{R}$ and $K : \mathbb{R}^d \rightarrow \mathbb{R}$ such that

$$\int_0^T H(s) ds = \int_{\mathbb{R}^d} K(y) dy = 1.$$

Let $\ell \geq 1$ be an integer. We say that the kernels H or K have order ℓ if, for $k = 0, \dots, \ell - 1$, we have

$$\int_0^T s^k H(s) ds = \int_{\mathbb{R}^d} (y^1)^k K(y) dy^1 = \dots = \int_{\mathbb{R}^d} (y^d)^k K(y) dy^d = \mathbf{1}_{\{k=0\}}, \quad (2.8)$$

with $y = (y^1, \dots, y^d)$.

Construction of an estimator of $\mu_t(x) \in \mathbb{R}$

Let $(t_0, x_0) \in (0, T] \times \mathbb{R}^d$. For $h > 0$ we obtain a family of estimators of $\mu_{t_0}(x_0)$ by setting

$$\hat{\mu}_h^N(t_0, x_0) = \int_{\mathbb{R}^d} K_h(x_0 - x) \mu_{t_0}^N(dx), \quad (2.9)$$

with $K_h(x) = h^{-d} K(h^{-1}x)$.

Construction of an estimator of $b(t, x, \mu_t) \in \mathbb{R}^d$

Let $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$. Abusing notation slightly, define the $\mathbb{R}^d$ -valued random measure

$$\pi^N(dt, dx) = N^{-1} \sum_{i=1}^N X^i(dt) \delta_{X_t^i}(dx),$$

defined by

$$\int_{[0,T] \times \mathbb{R}^d} \phi(t, x) \pi^N(dt, dx) = N^{-1} \sum_{i=1}^N \int_0^T \phi(t, X_t^i) dX_t^i$$

for a test function $\phi : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$. Letting

$$(H \otimes K)_{\mathbf{h}}(t, x) = H_{h_1}(t) K_{h_2}(x) \text{ for } \mathbf{h} = (h_1, h_2), h_i > 0,$$

with $H_h(t) = h^{-1} H(h^{-1}t)$, we obtain a family of estimators of $\pi(t_0, x_0) = b(t_0, x_0, \mu_{t_0}) \mu_{t_0}(x_0)$ by setting

$$\begin{aligned} \hat{\pi}_{\mathbf{h}}^N(t_0, x_0) &= \int_{[0,T] \times \mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) \pi^N(dt, dx) \\ &= \left(\int_{[0,T] \times \mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) (\pi^N)^k(dt, dx) \right)_{1 \leq k \leq d}, \end{aligned} \quad (2.10)$$

i.e. by smoothing componentwise $\pi^N(dt, dx)$. We finally estimate $b(t_0, x_0, \mu_{t_0})$ by

$$\hat{b}_{\mathbf{h}, \varpi}^N(t_0, x_0) = \frac{\hat{\pi}_{\mathbf{h}}^N(t_0, x_0)}{\hat{\mu}_{\mathbf{h}}^N(t_0, x_0) \vee \varpi} \in \mathbb{R}^d, \quad \mathbf{h} = (h_1, h_2),$$

with $h, h_i > 0$ and some threshold $\varpi > 0$ that prevents the estimator to blow-up for small values of $\hat{\mu}_{\mathbf{h}}^N(t_0, x_0)$.

3 Nonparametric oracle estimation

Our results involve positive quantities that continuously depend on real-valued parameters of the problem, namely

$$\mathbf{b} = (\gamma_0, \gamma_1, b_0, |b|_{\text{Lip}}, |\delta_{\mu}^k b|_{\text{Lip}}, m_b, \sigma_{\pm}, T, d),$$

defined in Assumptions 2.1, 2.2 and 2.4, together with the dimension $d \geq 1$ of the ambient space and the value of the terminal time $T > 0$. In the following, the notation $A_N \lesssim B_N$ means the existence of $C > 0$ (possibly depending on $\mathbf{b}$ but not N) such that $A_N \leq C B_N$ for every $N \geq 1$.

3.1 Oracle estimation of $\mu_t(x)$

We fix $(t_0, x_0) \in (0, T] \times \mathbb{R}^d$ and implement a variant of the Goldenshluger-Lepski's algorithm [87, 88, 89] for pointwise estimation. Pick a discrete set

$$\mathcal{H}_1^N \subset [N^{-1/d}(\log N)^{2/d}, 1],$$

of admissible bandwidths such that $\text{Card}(\mathcal{H}_1^N) \lesssim N$. The algorithm, based on Lepski's principle, requires the family of estimators

$$\left(\hat{\mu}_{\mathbf{h}}^N(t_0, x_0), h \in \mathcal{H}_1^N \right)$$

defined in (2.9) and selects an appropriate bandwidth $\hat{h}^N$ from data $\mu_{t_0}^N(dx)$. Writing $\{x\}_+ = \max(x, 0)$, define

$$\mathbf{A}_h^N = \max_{h' \leq h, h' \in \mathcal{H}_1^N} \left\{ \left(\hat{\mu}_{\mathbf{h}}^N(t_0, x_0) - \hat{\mu}_{\mathbf{h}'}^N(t_0, x_0) \right)^2 - (\mathbf{V}_{\mathbf{h}}^N + \mathbf{V}_{\mathbf{h}'}^N) \right\}_+, \quad (2.11)$$

where

$$\mathbf{V}_h^N = \varpi_1 |K|_2^2 (\log N) N^{-1} h^{-d}, \quad \varpi_1 > 0. \quad (2.12)$$

Let

$$\hat{h}^N \in \operatorname{argmin}_{h \in \mathcal{H}_1^N} (\mathbf{A}_h^N + \mathbf{V}_h^N). \quad (2.13)$$

The data driven Goldenshluger-Lepski estimator of $\mu_{t_0}(x_0)$ defined by

$$\hat{\mu}_{\text{GL}}^N(t_0, x_0) = \hat{\mu}_{\hat{h}^N}^N(t_0, x_0)$$

is specified by K and ϖ_1 .

Remark 2.6. *The choice of the penalty $\mathbf{A}_h^N$ and the threshold $\mathbf{V}_h^N$ in (2.11) and (2.12) are standard in the GL methodology : $\mathbf{A}_h^N$ is a kind of proxy for the estimation of the mean of $\hat{\mu}_h^N(t_0, x_0)$ while $\mathbf{V}_h^N$ is the exact penalty needed in order to balance the size of the variance of the estimator in h , of order $|K|_2^2 N^{-1} h^{-d}$, inflated by a logarithmic term $\log N$ and tuned with $\varpi_1 > 0$. This enables one to control all the stochastic deviation terms. See in particular the proof of Theorem 2.7. We also refer to the original sources in Goldenshluger and Lepski [87, 88, 89].*

Oracle estimate

We need some notation. Given a kernel K , the *bias at scale* $h > 0$ of μ at point (t_0, x_0) is defined as

$$\mathcal{B}_h^N(\mu)(t_0, x_0) = \sup_{h' \leq h, h' \in \mathcal{H}_1^N} \left| \int_{\mathbb{R}^d} K_{h'}(x_0 - x) \mu_{t_0}(x) dx - \mu_{t_0}(x_0) \right|. \quad (2.14)$$

We are ready to give the performance of our estimator of $\mu_t(x)$, by means of an oracle inequality.

Theorem 2.7. *Work under Assumptions 2.1, 2.2 and 2.4. Let $(t_0, x_0) \in (0, T] \times \mathbb{R}^d$. The following oracle inequality holds true :*

$$\mathbb{E}_{\mathbb{P}^N} [(\hat{\mu}_{\text{GL}}^N(t_0, x_0) - \mu_{t_0}(x_0))^2] \lesssim \min_{h \in \mathcal{H}_1^N} (\mathcal{B}_h^N(\mu)(t_0, x_0)^2 + \mathbf{V}_h^N),$$

for large enough N , up to a constant depending on (t_0, x_0) , $|K|_\infty$ and $\mathfrak{b}$, provided $\hat{\mu}_{\text{GL}}^N(t_0, x_0)$ is calibrated with $\varpi_1 \geq 16\kappa_2^{-1}\kappa_3$, where κ_2 is specified in Theorem 2.18 and κ_3 is a (local) upper bound of μ_{t_0} , see Lemma 2.23 below.

Some remarks are in order : **1)** Up to an inessential logarithmic factor, our estimator achieves the optimal bias-variance tradeoff among every possible bandwidth $h \in \mathcal{H}_1^N$. **2)** The requirement $Nh^d \geq (\log N)^2$ for $h \in \mathcal{H}_1^N$ could be tightened to $Nh^d \geq (\log N)^{1+\epsilon}$ for an arbitrary $\epsilon > 0$; this is slightly more stringent than the usual bound $Nh^d \geq \log N$ in the literature [87, 88, 89], but this has no consequence for the subsequent minimax results. **3)** The choice of a pointwise loss function at (t_0, x_0) is inessential here : other integrated norms like $|\cdot|_p$ would work as well, following the general strategies of Lepski's principle. **4)** The construction of the estimator of $\mu_{t_0}(x_0)$ requires a lower bound on ϖ_1 that has to be set prior to the data analysis. The bound we obtain are presumably too large. In practice, ϖ_1 has to be tuned by other methods, possibly using data. Such approaches in the context of Lepski's methods have been recently introduced by Lacour *et al.* [117]. This weakness is common to all nonparametric methods that depend on a data-driven bandwidth.

3.2 Oracle estimation of $b(t, x, \mu_t)$

Similarly to the estimation of $\mu_t(x)$, we pick a discrete set

$$\mathcal{H}_2^N \subset [N^{-1/(d+1)}(\log N)^{2/(d+1)}, (\log N)^{-2}] \times [N^{-1/(d+1)}(\log N)^{2/(d+1)}, 1], \quad (2.15)$$

with cardinality $\text{Card } \mathcal{H}_2^N \lesssim N$. We assume that $\mathcal{H}_2^N$ is equipped with some ordering $\preceq$ such that for every $\mathbf{h}, \mathbf{h}' \in \mathcal{H}_2^N$, we have either $\mathbf{h} \preceq \mathbf{h}'$ or $\mathbf{h}' \preceq \mathbf{h}$. The construction uses $\hat{\mu}_{\text{GL}}^N(t_0, x_0)$, given in addition the family of estimators

$$(\hat{\pi}_{\mathbf{h}}^N(t_0, x_0), \mathbf{h} \in \mathcal{H}_2^N)$$

defined in (2.10) and constructed with the kernel $H \otimes K$. Define

$$\mathbf{A}_{\mathbf{h}}^N = \max_{\mathbf{h}' \preceq \mathbf{h}, \mathbf{h}' \in \mathcal{H}_2^N} \left\{ |\hat{\pi}_{\mathbf{h}}^N(t_0, x_0) - \hat{\pi}_{\mathbf{h}'}^N(t_0, x_0)|^2 - (\mathbf{V}_{\mathbf{h}}^N + \mathbf{V}_{\mathbf{h}'}^N) \right\}_+, \quad (2.16)$$

where

$$\mathbf{V}_{\mathbf{h}}^N = \varpi_2 |H \otimes K|_2^2 (\log N) N^{-1} h_1^{-1} h_2^{-d}, \quad \varpi_2 > 0. \quad (2.17)$$

Let

$$\hat{\mathbf{h}}^N \in \text{argmin}_{\mathbf{h} \in \mathcal{H}_2^N} (\mathbf{A}_{\mathbf{h}}^N + \mathbf{V}_{\mathbf{h}}^N).$$

The data-driven Goldenshluger-Lepski estimator of $b(t_0, x_0, \mu_{t_0})$ is defined as

$$\hat{b}_{\text{GL}}^N(t_0, x_0) = \hat{b}_{\hat{\mathbf{h}}^N, \hat{\mathbf{h}}^N}^N(t_0, x_0) \varpi_3 = \frac{\hat{\pi}_{\hat{\mathbf{h}}^N}^N(t_0, x_0)}{\hat{\mu}_{\hat{\mathbf{h}}^N}^N(t_0, x_0) \vee \varpi_3} \quad (2.18)$$

and is specified by H, K, ϖ_1, ϖ_2 and the threshold $\varpi_3 > 0$ that prevents the estimator to blow-up for small values of $\hat{\mu}_{\hat{\mathbf{h}}^N}^N(t_0, x_0)$.

Remark 2.8. *The same comments as in Remark 2.6 apply here for the specification of $\mathbf{A}_{\mathbf{h}}^N$ in (2.16) and $\mathbf{V}_{\mathbf{h}}^N$ in (2.17), noting that in the anisotropic case, the variance of $\hat{\pi}_{\mathbf{h}}^N(t_0, x_0)$ is now of order $|H \otimes K|_2^2 N^{-1} h_1^{-1} h_2^{-d}$.*

Oracle estimates

Given a kernel $H \otimes K$, the bias at scale $\mathbf{h}$ of $\pi = b\mu$ at point (t_0, x_0) is defined as

$$\mathcal{B}_{\mathbf{h}}^N(\pi)(t_0, x_0) = \sup_{\mathbf{h}' \preceq \mathbf{h}, \mathbf{h}' \in \mathcal{H}_2^N} \left| \int_{[0, T] \times \mathbb{R}^d} (H \otimes K)_{\mathbf{h}'}(t_0 - t, x_0 - x) \pi(t, x) dx dt - \pi(t_0, x_0) \right|. \quad (2.19)$$

We are ready to give an oracle bound for the estimation of $b(t, x, \mu_t)$.

Theorem 2.9. *Work under Assumptions 2.1, 2.2 and 2.4. Let $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$. The following oracle inequality holds*

$$\mathbb{E}_{\mathbb{P}^N} [|\hat{b}_{\text{GL}}^N(t_0, x_0) - b(t_0, x_0, \mu_{t_0})|^2] \lesssim \min_{\mathbf{h} \in \mathcal{H}_1^N} (\mathcal{B}_{\mathbf{h}}^N(\mu)(t_0, x_0)^2 + \mathbf{V}_{\mathbf{h}}^N) + \min_{\mathbf{h} \in \mathcal{H}_2^N} (\mathcal{B}_{\mathbf{h}}^N(\pi)(t_0, x_0)^2 + \mathbf{V}_{\mathbf{h}}^N),$$

for large enough N , up to a constant depending on (t_0, x_0) , $|H \otimes K|_{\infty}$, and $\mathbf{b}$, provided

$$\varpi_2 \geq 12d\kappa_3 \max(12T\kappa_2^{-1}\kappa_5^2, 25|\text{Tr}(c)|_{\infty}) \quad \text{and} \quad \varpi_3 \leq \kappa_4,$$

where κ_2 is defined in Theorem 2.18, and $\kappa_3, \kappa_4, \kappa_5$ are (local) upper or lower bounds on μ and b defined in Lemma 2.23 below.

Some remarks : **1)** The same remarks as **1), 2), 3** and **4)** after Theorem 2.7 are in order. This includes the calibration of ϖ_2, ϖ_3 and the requirement $Nh_1h_2^d \geq (\log N)^2$ for $h \in \mathcal{H}_2^N$ that could be tightened to $Nh_1h_2^d \geq (\log N)^{1+\epsilon}$ for an arbitrary $\epsilon > 0$. However, the requirement $h_1 \leq (\log N)^{-2}$ is a bit unusual and necessary for technical reason : it enables us to manage the delicate term II in the proof. Fortunately, this has no consequence for the subsequent minimax results, since we always look for oracle bandwidths of the form $N^{-\epsilon}$ that are much smaller than $(\log N)^{-2}$. **2)** The estimator $\widehat{b}_{\text{GL}}^N$ is a quotient estimator that estimates the ratio of $\pi(t, x) = b(t, x, \mu_t)\mu_t(x)$ and $\mu_t(x)$, very much in the sense of a Nadaraya-Watson (NW) type estimator in regression [157]. Its performance is similar to the worst performance of the estimation of the product π and μ . However, the smoothness of μ is usually no worse than the smoothness of b and we do not lose in terms of approximation results, see Section 4 and Proposition 2.13 below for the formulation of a minimax theory in this setting.

4 Adaptive minimax estimation

4.1 Anisotropic Hölder smoothness classes for McKean-Vlasov models

Definition 2.10. Let $x_0 \in \mathbb{R}^d$ and $\mathcal{U}$ a neighbourhood of x_0 . We say that $f : \mathbb{R}^d \rightarrow \mathbb{R}$ belongs to $\mathcal{H}^\alpha(x_0)$ with $\alpha > 0$ if for every $x, y \in \mathcal{U}$

$$|D^s f(y) - D^s f(x)| \leq C|y - x|^{\alpha - \lfloor \alpha \rfloor} \quad (2.20)$$

for any s such that $|s| \leq \lfloor \alpha \rfloor$, the largest integer strictly smaller than α ; $s \in \mathbb{N}^d$ is a multi-index with $|s| = s_1 + \dots + s_d$ and $D^s = \frac{\partial^{|s|}}{\partial_1^{s_1} \dots \partial_d^{s_d}}$.

The definition depends on x_0 via $\mathcal{U}$ but this is further omitted in the notation for simplicity. We obtain a semi-norm by setting

$$|f|_{\mathcal{H}^\alpha(x_0)} = \sup_{x \in \mathcal{U}} |f(x)| + C(f),$$

where $C(f)$ is the smallest constant C for which (2.20) holds. The extension of Definition 2.10 for $\mathbb{R}^d$ -valued functions is straightforward by considering coordinate functions. For time-varying functions defined on $(0, T)$ we have the

Definition 2.11. Let $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$ and $\alpha, \beta > 0$. The function $f : (0, T) \times \mathbb{R}^d \rightarrow \mathbb{R}$ belongs to the anisotropic Hölder class $\mathcal{H}^{\alpha, \beta}(t_0, x_0)$ if there exist relatively compact neighbourhoods $\mathcal{U}_{t_0}$ of t_0 in $(0, T)$ and $\mathcal{U}_{x_0}$ of x_0 in $\mathbb{R}^d$ such that $f(\cdot, x) \in \mathcal{H}^\alpha(t_0)$ uniformly in $x \in \mathcal{U}_{x_0}$ and $f(t, \cdot) \in \mathcal{H}^\beta(x_0)$ uniformly in $t \in \mathcal{U}_{t_0}$, so that

$$|f|_{\mathcal{H}^{\alpha, \beta}(t_0, x_0)} = \sup_{x \in \mathcal{U}_{x_0}} |\mu^b(\cdot, x)|_{\mathcal{H}^\alpha(t_0)} + \sup_{t \in \mathcal{U}_{t_0}} |\mu^b(t, \cdot)|_{\mathcal{H}^\beta(x_0)} < \infty. \quad (2.21)$$

Again, the extension to $\mathbb{R}^d$ -valued functions is straightforward : a mapping $f = (f^k)_{1 \leq k \leq d} : (0, T) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ belongs to $\mathcal{H}^{\alpha, \beta}(t_0, x_0)$ if $f^k \in \mathcal{H}^{\alpha, \beta}(t_0, x_0)$ for every $k = 1, \dots, d$.

We model Hölder smoothness classes for the density function $\mu_t(x)$ and the drift $b(t, x, \mu_t)$. The McKean-Vlasov model (2.2) is parametrised by (b, σ, μ_0) , or rather (b, c, μ_0) , with $c = \sigma\sigma^\top$.

We denote by $\mathcal{P} = \mathcal{P}(\mathbf{b})$ the class of (b, c, μ_0) satisfying Assumptions 2.1, 2.2 and 2.4 with model parameter $\mathbf{b}$. We let

$$(b, c, \mu_0) \mapsto \mu = \mathcal{S}(b, c, \mu_0)$$

denote the solution (or forward) map of (2.3).

Definition 2.12. Let $\alpha, \beta > 0$. The anisotropic Hölder class $\mathcal{S}^{\alpha, \beta}(t_0, x_0)$ to the solution of (2.3) is defined by

$$\mathcal{S}^{\alpha, \beta}(t_0, x_0) = \left\{ (b, c, \mu_0) \in \mathcal{P}, \mu = \mathcal{S}(b, c, \mu_0) \in \mathcal{H}^{\alpha, \beta}(t_0, x_0) \right\}.$$

Before establishing minimax rates of convergence, we briefly investigate how rich is the class $\mathcal{S}^{\alpha, \beta}(t_0, x_0)$. **1)** If $b(t, x, \mu) = b(t, x)$ does not involve an interaction, then we have explicit formulas for μ in some cases when c is regular, see *e.g.* Genon-Catalot and Jacod [80] and the formulas become relatively tractable in dimension $d = 1$, especially when $b(t, x) = b(x)$, $c(t, x) = c(x)$ and μ_0 is the invariant distribution of the diffusion process X_t^i provided it exists. In that case, one can construct $\mu \in \mathcal{H}^{\alpha, \beta}(t_0, x_0)$ with arbitrary $\alpha, \beta > 0$ for specific choices of c and b thanks to Feller's classification of scalar diffusions (see *e.g.* Revuz and Yor [165]). **2)** For a non-trivial representation of $b(t, x, \mu)$ as in the Vlasov model, we have the following result, that shows how versatile the classes $\mathcal{H}^{\alpha, \beta}(t_0, x_0)$ can be.

Proposition 2.13. Let $c(t, x) = \frac{1}{2}\sigma^2 \text{Id}$ with $\sigma > 0$ and $b(t, x, \mu) = b(x, \mu) = F \star \mu(x) + G(x)$, with F having compact support, $\mu_0 \in \mathcal{P}_1$ with a continuous bounded density satisfying Assumption 2.1 and

$$|G|_{\mathcal{H}^\beta} + |F|_{\mathcal{H}^{\beta'}} + |\mu_0|_{\mathcal{H}^{\beta''}} < \infty,$$

for some $\beta, \beta' > 1$ and $\beta'' > 0$ (and β non-integer for technical reason). Here, $\mathcal{H}^\beta$ denotes the global Hölder space (obtained when taking $\mathcal{U} = \mathbb{R}^d$ in Definition 2.10). Then, for every $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$, we have $\mu \in \mathcal{H}^{\alpha, \beta+1}(t_0, x_0)$ and $b(\cdot, \cdot, \mu) \in \mathcal{H}^{\alpha, \beta}(t_0, x_0)$ with $\alpha = (\beta + 1)/2$.

The proof relies on classical estimates for parabolic equations, see *e.g.* the textbook by Bogachev *et al.* [17]. See also Méléard and Jourdain [107] for analogous results. It is sketched in Appendix. 8.3.

4.2 Minimax adaptive estimation of $\mu_t(x)$

For $\alpha, \beta, L > 0$, we set

$$\mathcal{S}_L^{\alpha, \beta}(t_0, x_0) = \left\{ (b, c, \mu_0) \in \mathcal{P}, |\mathcal{S}(b, c, \mu_0)|_{\mathcal{H}^{\alpha, \beta}(t_0, x_0)} \leq L \right\}, \quad (2.22)$$

that shall serve as a smoothness model for the unknown μ , where the semi-norm $|\cdot|_{\mathcal{H}^{\alpha, \beta}(t_0, x_0)}$ is defined in (2.21).

Since the estimator $\hat{\mu}_{\text{GL}}^N(t_0, x_0)$ is built on $\mu_{t_0}^N$ solely and not the whole process $(\mu_t^N)_{0 \leq t \leq T}$, we study minimax rates of convergence in restriction to the experiment generated by $\mu_{t_0}^N$. We have the following adaptive upper bound and accompanying lower bound for estimating μ :

Theorem 2.14. Work under Assumptions 2.1, 2.2 and 2.4. Let $\hat{\mu}_{\text{GL}}^N(t_0, x_0)$ be specified with a kernel K of order $\ell \geq 1$ as defined in (2.8). For every $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$, we have,

$$\sup_{(b, c, \mu_0)} \left(\mathbb{E}_{\mathbb{P}^N} \left[\left(\hat{\mu}_{\text{GL}}^N(t_0, x_0) - \mu_{t_0}(x_0) \right)^2 \right] \right)^{1/2} \lesssim \left(\frac{\log N}{N} \right)^{\beta \wedge \ell / (2\beta \wedge \ell + d)} \quad (2.23)$$

for large enough N , up to a constant that depends on $\mathfrak{b}$, K , β and L only. Moreover

$$\inf_{\widehat{\mu}} \sup_{(b,c,\mu_0)} \mathbb{E}_{\mathbb{P}^N} [|\widehat{\mu} - \mu_{t_0}(x_0)|] \gtrsim N^{-\beta/(2\beta+d)} \quad (2.24)$$

for large enough N . The infimum in (2.24) is taken over all estimators constructed with $\mu_{t_0}^N$. The supremum in (2.23) and (2.24) is taken over $\mathcal{S}_L^{\alpha,\beta}(t_0, x_0)$ for arbitrary $\alpha, \beta, L > 0$.

Some remarks : **1)** We obtain the (nearly) optimal rate of convergence $N^{-\beta/(2\beta+d)}$ for estimating $\mu_{t_0}(x_0)$ as a function of the d -dimensional state variable x_0 for fixed t_0 . The result obviously does not depend on the smoothness of $t \mapsto \mu_t(x_0)$ and is constructed from data $(X_{t_0}^1, \dots, X_{t_0}^N)$ for fixed t_0 . **2)** The extra logarithmic payment is unavoidable for pointwise estimation, as a result of the classical Lepski-Low phenomenon [127, 135]. **3)** Although the result is stated for arbitrary $\alpha, \beta > 0$, the mapping $(t, x) \mapsto \mu_t(x)$ is locally smooth; there is no contradiction and result must be understood as bounds that are valid over smooth functions μ having prescribed $\mathcal{S}^{\alpha,\beta}(t_0, x_0)$ semi-norms. **4)** Our concentration result Theorem 2.18 for the fluctuation of μ_t^N around μ_t enables us to improve on the estimation result in Proposition 2.1 of Bolley *et al.* [21] that achieves the (non-adaptive) suboptimal rate $N^{-\beta/(2\beta+2d+2)}$. This is due to the fact that Bolley *et al.* [21] rely on controlling the fluctuations of μ_t^N around μ_t in Wasserstein distance, and therefore have to accomodate a dimension effect that we do not have here.

4.3 Minimax adaptive estimation of $b(t, x, \mu_t)$

For $\alpha, \beta > 0$ and $L > 0$, in analogy to the class $\mathcal{S}_L^{\alpha,\beta}(t_0, x_0)$ defined in (2.22) above, we model the smoothness of the function $(t, x) \mapsto b(t, x, \mu_t)$ via the class

$$\mathcal{D}_L^{\alpha,\beta}(t_0, x_0) = \{(b, c, \mu_0) \in \mathcal{P}, |b(\cdot, \cdot, \mathcal{S}(b, c, \mu_0))|_{\mathcal{H}^{\alpha,\beta}(t_0, x_0)} \leq L\}.$$

Define the effective anisotropic smoothness $s_d(\alpha, \beta)$ by

$$\frac{1}{s_d(\alpha, \beta)} = \frac{1}{\alpha} + \frac{d}{\beta}.$$

We have the following adaptive upper bound and accompanying lower bound for estimating the drift b :

Theorem 2.15. *Work under Assumptions 2.1, 2.2 and 2.4. Let $\widehat{b}_{\text{GL}}^N(t_0, x_0)$ be constructed with kernels H and K of order $\ell \geq 1$. For every $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$, we have*

$$\sup_{(b,c,\mu_0)} (\mathbb{E}_{\mathbb{P}^N} [|\widehat{b}_{\text{GL}}^N(t_0, x_0) - b(t_0, x_0, \mu_{t_0})|^2])^{1/2} \lesssim \left(\frac{\log N}{N} \right)^{s_d(\alpha,\beta) \wedge \ell_d / (2s_d(\alpha,\beta) \wedge \ell_d + 1)}, \quad (2.25)$$

for large enough N , with $\ell_d = \ell/(d+1)$, up to constants that depend on $\mathfrak{b}$, $H \otimes K$ and α, β, L only. Moreover,

$$\inf_{\widehat{b}} \sup_{(b,c,\mu_0)} \mathbb{E}_{\mathbb{P}^N} [|\widehat{b} - b(t_0, x_0, \mu_{t_0})|] \gtrsim N^{-s_d(\alpha,\beta)/(2s_d(\alpha,\beta)+1)}, \quad (2.26)$$

for large enough N . The infimum in (2.26) is taken over all estimators constructed with $(\mu_t^N)_{0 \leq t \leq T}$. The supremum in (2.25) and (2.26) is taken over $\mathcal{S}_L^{\alpha,\beta}(t_0, x_0) \cap \mathcal{D}_L^{\alpha,\beta}(t_0, x_0)$ for some known non-decreasing parametrisation $\alpha = \alpha(\beta) > 0$ such that $\alpha(\beta)/\beta$ is non-increasing, with $\beta > 0$ and $L > 0$.

Some remarks : **1)** Theorem 2.15 establishes that estimating $b(t, x, \mu_t)$ has the same complexity as estimating d functions of $1 + d$ variables in the time and space domain. Each component $t \mapsto b^k(t, x, \mu_t)$ has smoothness α , while $x^\ell \mapsto b^k(t, (x^1, \dots, x^\ell, \dots, x^d), \mu_t)$ has smoothness β for $\ell = 1, \dots, d$, resulting in an anisotropic function $(t, x) \mapsto b^k(t, x, \mu_t)$ of $1 + d$ variables with smoothness index $(\alpha, \beta, \dots, \beta)$. We therefore recover the usual anisotropic minimax rate of convergence, with effective smoothness $s_d(\alpha, \beta)$ obtained as the arithmetico-geometric mean of the smoothness index $(\alpha, \beta, \dots, \beta)$. **2)** For technical simplicity, we consider α and β to be linked, as for instance in Proposition 2.13 where we have $\alpha = \alpha(\beta) = (\beta + 1)/2$. This somehow weakens our anisotropic adaptation result, but enables us to easily construct a well-behaved ordering $\preceq$ for $\mathcal{H}_2^N$ that behaves well with respect to the bias at scale $\mathbf{h}$ as defined in (2.19). Dropping this restriction is possible in principle, as in the original paper of Goldenshluger and Lepski [88], yet at a significant additional technical cost. **3)** The remarks **2)** and **3)** of Theorem 2.14 are valid here as well.

5 Estimation of the interaction in the Vlasov model

In this section, we work under Assumptions 2.1 and 2.2 with a constant $c(t, x) = \sigma^2 \text{Id}$ for some $\sigma > 0$ and Assumption 2.4 (iii), *i.e.* in the Vlasov case

$$b(t, x, \mu) = b(x, \mu) = \int_{\mathbb{R}^d} \tilde{b}(x, y) \mu(dy),$$

for $d \geq 1$ and with a time homogeneous drift kernel

$$\tilde{b}(x, y) = F(x - y) + G(x), \quad x, y \in \mathbb{R}^d.$$

Model (2.2) then reads

$$\begin{cases} dX_t^i = G(X_t^i)dt + N^{-1} \sum_{j=1}^N F(X_t^i - X_t^j)dt + \sigma dB_t^i, & 1 \leq i \leq N, \quad t \in [0, T], \\ \mathcal{L}(X_0^1, \dots, X_0^N) = \mu_0^{\otimes N}. \end{cases}$$

We assume that each component $F^k \in L^1(\mathbb{R}^d)$ for every $k = 1, \dots, d$. We are interested in identifying the interaction function $x \mapsto F(x)$ from data (2.1) and possibly $x \mapsto G(x)$, rather considered here as a nuisance parameter². We have

$$\begin{aligned} b(x, \mu_t) &= G(x) + \int_{\mathbb{R}^d} F(x - y) \mu_t(y) dy \\ &= G(x) + F \star \mu_t(x), \end{aligned} \tag{2.27}$$

where $f \star \mu_t(x) = \left(\int_{\mathbb{R}^d} f^k(x - y) \mu_t(y) dy \right)_{1 \leq k \leq d}$ denotes the convolution between $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and μ_t .

5.1 Identification of the interaction F

Introduce the linear form $\mathcal{L}$ acting on test functions $\varphi : [0, T] \rightarrow \mathbb{C}$ defined by

$$\mathcal{L}\varphi = \int_{[0, T]} \varphi(t) w(t) \rho(dt), \tag{2.28}$$

2. In particular, it is a first step toward the interesting problem of testing the hypothesis $F = 0$ against a set of local alternatives that quantify how far F is from being constant.

where ρ is a probability distribution on $[0, T]$ and $w : [0, T] \rightarrow \mathbb{R}$ a bounded weight function such that $\int_{[0, T]} w(t) \rho(dt) = 0$. Note that $\mathcal{L}1 = 0$ where 1 denotes the constant function. Applying $\mathcal{L}$ on both sides of (2.27), we obtain

$$\mathcal{L}b(x, \mu) = F \star \mathcal{L}\mu(x) \quad (2.29)$$

by Fubini's theorem. For $f = (f^k)_{1 \leq k \leq d} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ with $f^k \in L^1(\mathbb{R}^d)$, define a Fourier transform

$$\mathcal{F}(f)(\xi) = \left(\int_{\mathbb{R}^d} e^{-2i\pi\xi^\top x} f^k(x) dx \right)_{1 \leq k \leq d}, \quad \xi \in \mathbb{R}^d,$$

so that whenever $g \in L^1(\mathbb{R}^d)$, we have $\mathcal{F}(f \star g) = \mathcal{F}(f) \cdot \mathcal{F}(g)$. We infer from (2.29)

$$\mathcal{F}(\mathcal{L}b(\cdot, \mu)) = \mathcal{F}(F) \cdot \mathcal{F}(\mathcal{L}\mu).$$

This yields the formal decomposition

$$\mathcal{F}(F) = \frac{\mathcal{F}(\mathcal{L}b(\cdot, \mu))}{\mathcal{F}(\mathcal{L}\mu)}, \quad (2.30)$$

provided the quotient is well defined.

5.2 Consistent estimation of F

A first estimation strategy consists in plugging-in our estimators of $b(t, x, \mu_t)$ and μ_t in (2.30) above. A somewhat simpler estimator of $\mathcal{F}(\mathcal{L}\mu)(\xi) = \mathcal{L}\mathcal{F}(\mu)(\xi)$ is given by the periodogram

$$\mathcal{L} \left(\int_{\mathbb{R}^d} e^{-2i\pi\xi^\top x} \mu^N(dx) \right) = \int_{\mathbb{R}^d} e^{-2i\pi\xi^\top x} \mathcal{L}\mu^N(dx) = \mathcal{F}(\mathcal{L}\mu^N)(\xi)$$

for which we need not tune a bandwidth. Following Johannes [106], we obtain an estimator of F by the formula

$$\mathcal{F}(\widehat{F}_{\varpi, \varpi'}^N) = \frac{\mathcal{F}(\mathcal{L}((\widehat{b}_{h, \mathbf{h}}^N)_{\varpi'}^r)) \cdot \overline{\mathcal{F}(\mathcal{L}\mu^N)}}{|\mathcal{F}(\mathcal{L}\mu^N)|^2} \mathbf{1}_{\{|\mathcal{F}(\mathcal{L}\mu^N)|^2 \geq \varpi\}} \in \mathbb{R}^d \quad (2.31)$$

for some threshold $\varpi > 0$ vanishing as $N \rightarrow \infty$, with the estimator

$$\widehat{b}_{h, \mathbf{h}}^N(t_0, x_0)_{\varpi'}^r = \widehat{b}_{h, \mathbf{h}}^N(t_0, x_0)_{\varpi'} \mathbf{1}_{\{|x| \leq r\}}$$

of $b(t_0, x_0, \mu_{t_0})$, constructed in Section 2.2 for some threshold $\varpi' > 0$ and bandwidths $h > 0$ and $\mathbf{h} > 0$. We also set the estimator to be equal to 0 outside $|x| \leq r$ for some $r > 0$. We obtain a consistency result under the following additional assumption :

Assumption 2.16. *We have $|\mathcal{F}(\mathcal{L}\mu)(\xi)| > 0$ $d\xi$ -almost everywhere.*

Theorem 2.17. *Work under the assumptions of Proposition 2.13 and Assumption 2.16. Assume moreover that G is in $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ componentwise. If w has compact support in $(0, T)$, there exists a choice of $(\varpi, \varpi', h, \mathbf{h}) = (\varpi_N, \varpi'_N, h_N, \mathbf{h}_N) \rightarrow 0$ and $r = r_N \rightarrow \infty$ such that*

$$\mathbb{E}_{\mathbb{P}^N} [|\widehat{F}_{\varpi, \varpi'}^N - F|_2^2] \rightarrow 0 \text{ as } N \rightarrow \infty.$$

Some remarks : **1)** Theorem 2.17 proves that we can reconstruct the interaction force F from data (2.1), while the function G remains a nuisance parameter. This is a first step for the construction of a statistical test based on data (2.1) for the presence against the absence of interaction between particles in the Vlasov model. **2)** Although we obtain consistency, we do not have a rate of convergence and our result is not uniform in the model parameter. A glance at the proof of Theorem 2.17 shows that it is possible to cook-up a result with a rate of convergence and some uniformity in the parameter, provided we have a sharp control from below on the decay $|\mathcal{F}(\mathcal{L}\mu)(\xi)|$ or rather $|\mathcal{F}(\mu)(\xi)|$ as $|\xi| \rightarrow \infty$ as well as the decay of $\inf_{t \in [r_1, r_2], |x| \leq r} \mu_t(x)$ for given $[r_1, r_2] \subset (0, T)$ as $r \rightarrow \infty$. This requires the exact knowledge of the smoothness of the solution map $\mu = \mathcal{S}(b, c, \mu_0)$, and it is a delicate issue, see Proposition 2.13; we can anticipate ill-posedness. **3)** From our estimator of F , we may construct a plug-in estimator for the function G by setting

$$\widehat{G}^N(x)_{\varpi, \varpi'} = -\widehat{b}_{h, h}^N(t, x)_{\varpi'}^r - \widehat{F}_{\varpi, \varpi'}^N \star_x \widehat{\mu}_h^N(t, x).$$

A consistency result can be obtained in the same way as for Theorem 2.17. **4)**, the deconvolution method we employ here requires quite a stringent localisation assumption on the external force G and the interaction force F . As pointed out by a referee, it does not apply for gradient forces of the form $G = -\nabla V$ and $F = -\nabla W$ where V and W diverge polynomially at infinity like *e.g.* in [12, 92, 38] for which alternative methods yet need to be constructed.

6 Probabilistic tools : a concentration inequality

6.1 A Bernstein inequality

Let $\rho(dt)$ be a probability measure on $[0, T]$. We establish a deviation inequality for the sequence of signed measures $\nu^N(dt, dx) - \nu(dt, dx)$, where

$$\nu^N(dt, dx) = \mu_t^N(dx) \otimes \rho(dt) \quad \text{and} \quad \nu(dt, dx) = \mu_t(x)dx \otimes \rho(dt).$$

We have a Bernstein concentration inequality :

Theorem 2.18. *Work under Assumptions 2.1, 2.2 and 2.4. Let $(\mu_t)_{0 \leq t \leq T}$ denote the unique solution of (2.3) with $\mu_{t=0} = \mu_0$ satisfying Assumption 2.1. Then there exist $\kappa_1, \kappa_2 > 0$ depending on $\mathfrak{b}$ such that*

$$\mathbb{P}^N \left(\int_{[0, T] \times \mathbb{R}^d} \phi(t, y) (\nu^N(dt, dy) - \nu(dt, dy)) \geq x \right) \leq \kappa_1 \exp \left(-\kappa_2 \frac{Nx^2}{|\phi|_{L^2(\nu)}^2 + |\phi|_\infty x} \right)$$

for every $x \geq 0$, for every bounded $\phi : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$, and for any probability measure $\rho(dt)$ on $[0, T]$.

As a corollary, for a bounded $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ and any $0 \leq t_0 \leq T$, picking $\rho(dt) = \delta_{t_0}(dt)$, we obtain

$$\mathbb{P}^N \left(\int_{\mathbb{R}^d} \phi(y) (\mu_{t_0}^N(dy) - \mu_{t_0}(y)dy) \geq x \right) \leq \kappa_1 \exp \left(-\kappa_2 \frac{Nx^2}{|\phi|_{L^2(\mu_{t_0})}^2 + |\phi|_\infty x} \right).$$

Several remarks : **1)** Up to the constants κ_i , the result is quite satisfactory and comparable to the Bernstein deviation inequality for independent data, see *e.g.* Massart [145]. Theorem 2.18 is the

gateway to derive sharp nonparametric estimators, although it has an independent interest as a deviation inequality. **2)** The constants κ_i are explicitly computable, but certainly far from being optimal with the method of proof employed here. **3)** Our method of proof uses a change of measure argument based on Girsanov's theorem, in the spirit of the recent work of Lacker [116]. It can presumably be extended to path dependent coefficients in (2.2), but it is essential that the diffusion coefficient does not depend on μ_t^N . **4)** We have an interplay between the smoothness k of the drift b in its measure argument and the dimension d of the ambient state space. This is explained by the fact that we need to control an exponential moment of $\sum_{i=1}^N \int_t^{t+\delta} |b(s, X_s^i, \mu_s^N) - b(s, X_s^i, \mu_s)|^2 ds$ over small intervals $[t, t+\delta]$ in order to approximate the law of the data by the law of independent particles. This approximation is roughly controlled by $NW_1(\mu_t^N, \mu_t)^2$ for which a dimensional effect drastically deteriorates the rate of convergence, see *e.g.* Fournier and Guillin [74]. The k -linear differentiability of $\mu \mapsto b(t, x, \mu)$ enables us to mitigate this effect. In particular, in the Vlasov case covered by Assumption 2.4 (iii), we formally have $k = \infty$, and the result is valid in any dimension $d \geq 1$.

The remainder of Section 6 is devoted to the proof of Theorem 2.18.

6.2 Preparation for the proof of Theorem 2.18

We let $\bar{\mathbb{P}}^N$ denote the unique probability measure on $(\mathcal{C}, \mathcal{F}_T)$ under which the canonical process $X = (X^1, \dots, X^N)$ solves

$$\begin{cases} dX_t^i = b(t, X_t^i, \mu_t)dt + \sigma(t, X_t^i)d\bar{B}_t^i, & 1 \leq i \leq N, t \in [0, T], \\ \mathcal{L}(X_0^1, \dots, X_0^N) = \mu_0^{\otimes N}, \end{cases} \quad (2.32)$$

where

$$\bar{B}_t^i = \int_0^t c(s, X_s^i)^{-1/2} (dX_s^i - b(s, X_s^i, \mu_s)ds), \quad 1 \leq i \leq N,$$

are independent d -dimensional $\bar{\mathbb{P}}^N$ -Brownian motions. The existence of $\bar{\mathbb{P}}^N$ follows from Carmona and Delarue [37] or the lectures notes of Lacker [115]. In turn, the real-valued process

$$\bar{M}_t^N = \sum_{i=1}^N \int_0^t ((c^{-1/2}b)(s, X_s^i, \mu_s^N) - (c^{-1/2}b)(s, X_s^i, \mu_s))^\top d\bar{B}_s^i, \quad (2.33)$$

is a $\bar{\mathbb{P}}^N$ -local martingale. Here, $c^{-1/2}$ denotes any square-root of $c^{-1} = (\sigma\sigma^\top)^{-1}$.

The following estimate is the central result of the section. It is the key ingredient that enables us to implement a change of probability argument in order to obtain our concentration estimates. Its proof is delayed until Section 6.4.

Proposition 2.19. *Work under Assumptions 2.1, 2.2 and 2.4. For every $\tau > 0$, there exists $\delta_0 > 0$ depending on τ and $\mathbf{b}$ such that*

$$\sup_{N \geq 1} \sup_{t \in [0, T-\delta]} \mathbb{E}_{\bar{\mathbb{P}}^N} \left[\exp \left(\tau (\langle \bar{M}^N \rangle_{t+\delta} - \langle \bar{M}^N \rangle_t) \right) \right] \leq C_1,$$

for every $0 \leq \delta \leq \delta_0$ and some C_1 that depends on $\mathbf{b}$ and τ .

Let $\mathcal{E}_t(\overline{M}^N) = \exp(\overline{M}_t^N - \frac{1}{2}\langle \overline{M}^N \rangle_t)$ denote the (martingale) exponential of $\overline{M}_t^N$ and $\langle \overline{M}^N \rangle_t$ its predictable compensator. By Novikov's criterion – in its version developed in the classical textbook [110], Lemma 5.14, p.198 – Proposition 2.19 shows that the local martingale $\mathcal{E}_t(\overline{M}^N)$ is indeed a true martingale. This enables us to define a new probability measure on $(\mathcal{C}, \mathcal{F}_T)$ by setting

$$\tilde{\mathbb{P}}^N = \mathcal{E}_T(\overline{M}^N) \cdot \mathbb{P}^N.$$

By Girsanov's theorem, under $\tilde{\mathbb{P}}^N$, the canonical process solves (2.2). By uniqueness of the weak solution of (2.2), this proves $\tilde{\mathbb{P}}^N = \mathbb{P}^N$ and shows in particular that $\mathbb{P}^N \ll \tilde{\mathbb{P}}^N$ and

$$\frac{d\mathbb{P}^N}{d\tilde{\mathbb{P}}^N} = \mathcal{E}_T(\overline{M}^N).$$

6.3 Proof of Theorem 2.18

Step 1 : Let $\mathcal{A}^N \in \mathcal{F}_T$. Since $\mathbb{P}^N$ and $\tilde{\mathbb{P}}^N$ coincide on $\mathcal{F}_0$, we have

$$\mathbb{P}^N(\mathcal{A}^N) = \mathbb{E}_{\mathbb{P}^N}[\mathbb{P}^N(\mathcal{A}^N | \mathcal{F}_0)] = \mathbb{E}_{\tilde{\mathbb{P}}^N}[\mathbb{P}^N(\mathcal{A}^N | \mathcal{F}_0)].$$

Next, for any subdivision $0 = t_0 < t_1 < \dots < t_K \leq T$ and any $\mathcal{F}_T$ -measurable event $\mathcal{A}^N$, we claim

$$\mathbb{E}_{\tilde{\mathbb{P}}^N}[\mathbb{P}^N(\mathcal{A}^N | \mathcal{F}_0)] \leq \mathbb{E}_{\tilde{\mathbb{P}}^N}[\mathbb{P}^N(\mathcal{A}^N | \mathcal{F}_{t_K})]^{1/4^K} \prod_{j=1}^K \mathbb{E}_{\tilde{\mathbb{P}}^N} \left[\exp(2(\langle \overline{M}^N \rangle_{t_j} - \langle \overline{M}^N \rangle_{t_{j-1}})) \right]^{j/4}. \quad (2.34)$$

It follows that

$$\begin{aligned} \mathbb{P}^N(\mathcal{A}^N) &\leq \mathbb{E}_{\tilde{\mathbb{P}}^N}[\mathbb{P}^N(\mathcal{A}^N | \mathcal{F}_T)]^{1/4^K} \prod_{j=1}^K \mathbb{E}_{\tilde{\mathbb{P}}^N} \left[\exp(2(\langle \overline{M}^N \rangle_{t_j} - \langle \overline{M}^N \rangle_{t_{j-1}})) \right]^{j/4} \\ &\leq \tilde{\mathbb{P}}^N(\mathcal{A}^N)^{1/4^K} \sup_{N \geq 1} \sup_{t \in [0, T - \delta_0]} \left(\mathbb{E}_{\tilde{\mathbb{P}}^N} \left[\exp(2(\langle \overline{M}^N \rangle_{t+\delta_0} - \langle \overline{M}^N \rangle_t)) \right] \right)^{K(K+1)/8} \\ &\leq C_1^{K(K+1)/8} \tilde{\mathbb{P}}^N(\mathcal{A}^N)^{1/4^K} \end{aligned} \quad (2.35)$$

by (2.34) and Proposition 2.19 with $\tau = 2$, $t_j = jT/K$ and K large enough so that $t_j - t_{j-1} \leq \delta_0$.

Step 2 : Let

$$\begin{aligned} \mathcal{A}^N &= \left\{ \int_{[0, T] \times \mathbb{R}^d} \phi(t, y) (\nu^N(dt, dy) - \nu(dt, dy)) \geq x \right\} \\ &= \left\{ \sum_{i=1}^N \left(\int_0^T \phi(t, X_t^i) \rho(dt) - \int_{[0, T] \times \mathbb{R}^d} \phi(t, y) \mu_t(dy) \rho(dt) \right) \geq Nx \right\}, \end{aligned}$$

so that $\mathcal{A}^N \in \mathcal{F}_T$. Recall Bernstein's inequality : if $Z_1, \dots, Z_N$ are real-valued independent random variables bounded by some constant Q and such that $\mathbb{E}[Z_i] = 0$, we have

$$\mathbb{P}\left(\sum_{i=1}^N Z_i \geq y\right) \leq \exp\left(-\frac{y^2}{2(\sum_{i=1}^N \mathbb{E}[Z_i^2] + \frac{Qy}{3})}\right) \text{ for every } y \geq 0.$$

Under $\bar{\mathbb{P}}^N$, the random processes $(X_t^i)_{0 \leq t \leq T}$ are independent and identically distributed processes. Noticing that $\int_{[0,T] \times \mathbb{R}^d} \phi(t, x) \mu_t(dx) \rho(dt) = \mathbb{E}_{\bar{\mathbb{P}}^N}[\int_0^T \phi(t, X_t^i) \rho(dt)]$, we apply Bernstein inequality with $Z_i = \int_0^T \phi(t, X_t^i) \rho(dt) - \mathbb{E}_{\bar{\mathbb{P}}^N}[\int_0^T \phi(t, X_t^i) \rho(dt)]$, $y = Nx$ and $Q = 2|\phi|_\infty$ to infer

$$\bar{\mathbb{P}}^N(\mathcal{A}^N) \leq \exp\left(-\frac{Nx^2}{2(|\phi|_{L^2(\nu)}^2 + \frac{2}{3}|\phi|_\infty x)}\right),$$

using $\mathbb{E}[Z_i^2] \leq \mathbb{E}_{\bar{\mathbb{P}}^N}[(\int_0^T \phi(t, X_t^i) \rho(dt))^2] \leq |\phi|_{L^2(\nu)}^2$ by Jensen's inequality. By (2.35), we also have

$$\mathbb{P}^N(\mathcal{A}^N) \leq C_1^{K(K+1)/8} \bar{\mathbb{P}}^N(\mathcal{A}^N)^{1/4^K} \leq C_1^{K(K+1)/8} \exp\left(-\frac{Nx^2}{2 \cdot 4^K(|\phi|_{L^2(\nu)}^2 + \frac{2}{3}|\phi|_\infty x)}\right)$$

and we obtain Theorem 2.18 with $\kappa_1 = C_1^{K(K+1)/8}$ and $\kappa_2 = 2^{-1}4^{-K}$.

Step 3 : It remains to prove the key estimate (2.34), adapted from large deviation techniques, see *e.g.* Gärtner [78] and the estimate (4.2) in Theorem 2.6 in Lacker [116]. We proceed by induction. First,

$$\begin{aligned} \mathbb{E}_{\bar{\mathbb{P}}^N}[\mathbb{P}^N(\mathcal{A}^N | \mathcal{F}_{t_{j-1}})] &= \mathbb{E}_{\bar{\mathbb{P}}^N}[\mathbb{E}_{\mathbb{P}^N}[\mathbb{P}^N(\mathcal{A}^N | \mathcal{F}_{t_j}) | \mathcal{F}_{t_{j-1}}]] \\ &= \mathbb{E}_{\bar{\mathbb{P}}^N}\left[\mathbb{E}_{\bar{\mathbb{P}}^N}\left[\frac{\mathcal{E}_{t_j}(\bar{M}^N)}{\mathcal{E}_{t_{j-1}}(\bar{M}^N)} \mathbb{P}^N(\mathcal{A}^N | \mathcal{F}_{t_j}) | \mathcal{F}_{t_{j-1}}\right]\right] \\ &= \mathbb{E}_{\bar{\mathbb{P}}^N}\left[\frac{\mathcal{E}_{t_j}(\bar{M}^N)}{\mathcal{E}_{t_{j-1}}(\bar{M}^N)} \mathbb{P}^N(\mathcal{A}^N | \mathcal{F}_{t_j})\right], \end{aligned}$$

see *e.g.* Lemma 3.5.3 p. 193 in [?]. Next,

$$\frac{\mathcal{E}_{t_j}(\bar{M}^N)}{\mathcal{E}_{t_{j-1}}(\bar{M}^N)} = \mathcal{E}_{t_j}(2(\bar{M}^N - \bar{M}_{t_{j-1}}^N))^{1/2} (\exp(\langle \bar{M}^N \rangle_{t_j} - \langle \bar{M}^N \rangle_{t_{j-1}}))^{1/2}. \quad (2.36)$$

As shown before, under $\bar{\mathbb{P}}^N$, the process $\mathcal{E}_t(2(\bar{M}^N - \bar{M}_{t_{j-1}}^N))_{t \geq t_{j-1}}$ is a martingale and

$$\mathbb{E}_{\bar{\mathbb{P}}^N}[\mathcal{E}_{t_j}(2(\bar{M}^N - \bar{M}_{t_{j-1}}^N))] = 1. \quad (2.37)$$

Using (2.36) and Cauchy-Schwarz's inequality twice together with (2.37), we obtain

$$\begin{aligned} \mathbb{E}_{\bar{\mathbb{P}}^N}\left[\frac{\mathcal{E}_{t_j}(\bar{M}^N)}{\mathcal{E}_{t_{j-1}}(\bar{M}^N)} \mathbb{P}^N(\mathcal{A}^N | \mathcal{F}_{t_j})\right] &\leq \mathbb{E}_{\bar{\mathbb{P}}^N}[\mathbb{P}^N(\mathcal{A}^N | \mathcal{F}_{t_j})^2 \exp(\langle \bar{M}^N \rangle_{t_j} - \langle \bar{M}^N \rangle_{t_{j-1}})]^{1/2} \\ &\leq \mathbb{E}_{\bar{\mathbb{P}}^N}[\mathbb{P}^N(\mathcal{A}^N | \mathcal{F}_{t_j})^4]^{1/4} \mathbb{E}_{\bar{\mathbb{P}}^N}[\exp(2(\langle \bar{M}^N \rangle_{t_j} - \langle \bar{M}^N \rangle_{t_{j-1}}))]^{1/4}. \end{aligned}$$

By Jensen's inequality, we infer

$$\mathbb{E}_{\bar{\mathbb{P}}^N}[\mathbb{P}^N(\mathcal{A}^N | \mathcal{F}_{t_{j-1}})] \leq \mathbb{E}_{\bar{\mathbb{P}}^N}[\mathbb{P}^N(\mathcal{A}^N | \mathcal{F}_{t_j})]^{1/4} \mathbb{E}_{\bar{\mathbb{P}}^N}[\exp(2(\langle \bar{M}^N \rangle_{t_j} - \langle \bar{M}^N \rangle_{t_{j-1}}))]^{1/4}.$$

Repeating the argument over the subdivision $0 = t_0 < t_1 < \dots < t_K$ proves (2.34). The proof of Theorem 2.18 is complete.

6.4 Proof of Proposition 2.19

Preparation

Recall that the notation $A_N \lesssim B_N$: it means the existence of $C > 0$ possibly depending on $\mathfrak{b}$ and also τ in this part of the paper, but not N , such that $A_N \leq CB_N$ for every $N \geq 1$. The following classical moment estimate will be needed.

Lemma 2.20. *In the setting of Theorem 2.18, for every $p \geq 1$, we have*

$$\sup_{t \in [0, T]} \mathbb{E}_{\mathbb{P}^N} [|X_t^i|^{2p}] \leq p! C_2^p.$$

for some $C_2 > 0$ that depends on $\mathfrak{b}$ only.

In particular, the X_t^i are sub-Gaussian under $\mathbb{P}^N$ and satisfy

$$\mathbb{E}_{\mathbb{P}^N} [e^{\frac{1}{2C_2} |X_t^i|^2}] = 1 + \sum_{p \geq 1} \frac{2^{-p}}{p! C_2^p} \mathbb{E}_{\mathbb{P}^N} [|X_t^i|^{2p}] \leq 2. \quad (2.38)$$

The proof is classical (see *e.g.* estimates of this type in Méléard [148] or Sznitman [173]) and postponed to Appendix 8.1.

Completion of Proof of Proposition 2.19

Let $\tau > 0$. All we need to show is that for small enough δ , we have

$$\sup_{t \in [0, T - \delta]} \mathbb{E}_{\mathbb{P}^N} \left[\exp \left(\tau (\langle \overline{M}^N \rangle_{t+\delta} - \langle \overline{M}^N \rangle_t) \right) \right] \lesssim 1. \quad (2.39)$$

We start with a useful estimate.

Lemma 2.21. *Let $\xi_s^N(x) = b(s, x, \mu_s^N) - b(s, x, \mu_s)$. We have*

$$|\xi_s^N(X_s^i)|^2 \leq C \begin{cases} \mathcal{W}_1(\mu_s^N, \mu_s)^2 & \text{under Assumption 2.4(i),} \\ \sum_{\ell=1}^{k-1} \left| \int_{(\mathbb{R}^d)^\ell} \delta_\mu^\ell b(s, X_s^i, y^\ell, \mu_s) (\mu_s^N - \mu_s)^{\otimes \ell} (dy^\ell) \right|^2 \\ \quad + \mathcal{W}_1(\mu_s^N, \mu_s)^2 \wedge \mathcal{W}_1(\mu_s^N, \mu_s)^{2k} & \text{under Assumption 2.4(ii),} \\ \left| \int_{\mathbb{R}^d} \tilde{b}(s, X_s^N, y) (\mu_s^N - \mu_s) (dy) \right|^2 & \text{under Assumption 2.4(iii),} \end{cases}$$

for some explicit $C \geq 1$ depending on $\mathfrak{b}$.

Démonstration. The estimate follows from the Lipschitz continuity of b under Assumption 2.4 (i) with $C = |b|_{\text{Lip}}^2$ and from the definition of the Vlasov case with $C = 1$ under Assumption 2.4 (iii). We turn to the estimate under (ii). The k -linear differentiability of b enables us to write

$$\xi_s^N(X_s^i) = \sum_{\ell=1}^{k-1} \frac{1}{\ell!} \int_{(\mathbb{R}^d)^\ell} \delta_\mu^\ell b(s, X_s^i, y^\ell, \mu_s) (\mu_s^N - \mu_s)^{\otimes \ell} (dy^\ell) + \mathcal{R}_k, \quad (2.40)$$

where $y^\ell = (y_1, \dots, y_\ell) \in (\mathbb{R}^d)^\ell$ and

$$\mathcal{R}_k = \frac{1}{(k-1)!} \int_0^1 (1-\vartheta)^{k-1} \int_{(\mathbb{R}^d)^k} \delta_\mu^k b(s, X_s^i, y^k, [\mu_s^N, \mu_s]_\vartheta) (\mu_s^N - \mu_s)^{\otimes k} (dy^k) d\vartheta,$$

with $[\mu_s^N, \mu_s]_\vartheta = (1-\vartheta)\mu_s + \vartheta\mu_s^N$, as stems from the definition of linear differentiability and the iteration $\delta_\mu^\ell b = \delta_\mu \circ \delta_\mu^{\ell-1} b$, see also Lemma 2.2. of Chassagneux *et al.* [42] where (2.40) is established by induction.

Thanks to representation (2.7), the remainder term $\mathcal{R}_k$ equals

$$\frac{1}{(k-1)!} \int_0^1 (1-\vartheta)^{k-1} \sum_{\mathcal{J} \subset \{1, \dots, k\}, m \geq 1} \int_{(\mathbb{R}^d)^k} \bigotimes_{j \in \mathcal{J}} (\delta_\mu^k b)_{\mathcal{J}, j, m}(s, X_s^i, y_j, [\mu_s^N, \mu_s]_\vartheta) (\mu_s^N - \mu_s)^{\otimes k} (dy^k) d\vartheta.$$

Note that the product integral vanishes for all terms in the sum in $\mathcal{J}$ except $\mathcal{J} = \{1, \dots, k\}$ since $(\mu_s^N - \mu_s)(\mathbb{R}^d) = 0$. By definition,

$$|(\delta_\mu^k b)_{\{1, \dots, k\}, j, m}(s, X_s^i, \cdot, [\mu_s^N, \mu_s]_\vartheta)|_{\text{Lip}} \leq |\delta_\mu^k b|_{\text{Lip}} \text{ for every } (j, m)$$

and the sum in m has at most m_b terms by assumption. It follows that

$$\begin{aligned} |\mathcal{R}_k| &= \frac{1}{(k-1)!} \left| \int_0^1 (1-\vartheta)^{k-1} \sum_{m=1}^{m_b} \prod_{j=1}^k \int_{\mathbb{R}^d} (\delta_\mu^k b)_{\{1, \dots, k\}, j, m}(s, X_s^i, y_j, [\mu_s^N, \mu_s]_\vartheta) (\mu_s^N - \mu_s)(dy_j) d\vartheta \right| \\ &\leq \frac{m_b}{(k-1)!} \left| \int_0^1 (1-\vartheta)^{k-1} |\delta_\mu^k b|_{\text{Lip}}^k \left(\sup_{|\varphi|_{\text{Lip}} \leq 1} \int_{\mathbb{R}^d} \varphi d(\mu_s^N - \mu_s) \right)^k d\vartheta \right| \\ &\leq \frac{m_b |\delta_\mu^k b|_{\text{Lip}}^k}{k!} \mathcal{W}_1(\mu_s^N, \mu_s)^k. \end{aligned}$$

Writing $y^\ell = (y^{\ell-1}, y) \in (\mathbb{R}^d)^{\ell-1} \times \mathbb{R}^d$, we also have the rough bound

$$\begin{aligned} &\left| \int_{(\mathbb{R}^d)^\ell} \delta_\mu^\ell b(s, X_s^i, y^\ell, \mu_s) (\mu_s^N - \mu_s)^{\otimes \ell} (dy^\ell) \right| \\ &\leq \int_{(\mathbb{R}^d)^{\ell-1}} \left| \int_{\mathbb{R}^d} \delta_\mu^\ell b(s, X_s^i, (y^{\ell-1}, y), \mu_s) (\mu_s^N - \mu_s)(dy) \right| (\mu_s^N + \mu_s)^{\otimes (\ell-1)} (dy^{\ell-1}) \\ &\leq \sup_{y^{\ell-1} \in (\mathbb{R}^d)^{\ell-1}} |\delta_\mu^\ell b(s, X_s^i, (y^{\ell-1}, \cdot), \mu_s)|_{\text{Lip}} \\ &\quad \times \left| \sup_{|\varphi|_{\text{Lip}} \leq 1} \int_{\mathbb{R}^d} \varphi(y) (\mu_s^N - \mu_s)(dy) \right| \int_{(\mathbb{R}^d)^{\ell-1}} (\mu_s^N + \mu_s)^{\otimes (\ell-1)} (dy^{\ell-1}) \\ &= 2^{\ell-1} \sup_{y^{\ell-1} \in (\mathbb{R}^d)^{\ell-1}} |\delta_\mu^\ell b(s, X_s^i, (y^{\ell-1}, \cdot), \mu_s)|_{\text{Lip}} \mathcal{W}_1(\mu_s^N, \mu_s) \\ &\leq 2^{\ell-1} |\delta_\mu^\ell b|_{\text{Lip}} \mathcal{W}_1(\mu_s^N, \mu_s). \end{aligned}$$

Plugging this estimate in (2.40) and using the Lipschitz property for b , we obtain

$$|\mathcal{R}_k| \leq (|b|_{\text{Lip}} + \sum_{\ell=1}^{k-1} \frac{2^{\ell-1}}{\ell!} |\delta_\mu^\ell b|_{\text{Lip}}) \mathcal{W}_1(\mu_s^N, \mu_s)$$

and we conclude

$$|\mathcal{R}_k| = |\mathcal{R}_k(s, X_s^i, \mu_s^N, \mu_s)| \leq C' \mathcal{W}_1(\mu_s^N, \mu_s) \wedge \mathcal{W}_1(\mu_s^N, \mu_s)^k, \quad (2.41)$$

with

$$C' = \max\left(\frac{mb}{k!} |\delta_\mu^k b|_{\text{Lip}}^k, |b|_{\text{Lip}} + \sum_{\ell=1}^{k-1} \frac{2^{\ell-1}}{\ell!} |\delta_\mu^\ell b|_{\text{Lip}}\right).$$

From (2.40) and (2.41) we conclude

$$\begin{aligned} |\xi_s^N(X_s^i)|^2 &\leq k \left(\sum_{\ell=1}^{k-1} \frac{1}{(\ell!)^2} \left| \int_{(\mathbb{R}^d)^\ell} \delta_\mu^\ell b(s, X_s^i, y^\ell, \mu_s) (\mu_s^N - \mu_s)^{\otimes \ell} (dy^\ell) \right|^2 + |\mathcal{R}_k(s, X_s^i, \mu_s^N, \mu_s)|^2 \right) \\ &\leq C \left(\sum_{\ell=1}^{k-1} \left| \int_{(\mathbb{R}^d)^\ell} \delta_\mu^\ell b(s, X_s^i, y^\ell, \mu_s) (\mu_s^N - \mu_s)^{\otimes \ell} (dy^\ell) \right|^2 + \mathcal{W}_1(\mu_s^N, \mu_s)^2 \wedge \mathcal{W}_1(\mu_s^N, \mu_s)^{2k} \right), \end{aligned}$$

where $C = k \max(1, (C')^2)$ incorporates the constant in (2.41). $\square$

We now establish (2.39). By (2.33) and Lemma 2.21, we have

$$\begin{aligned} &\tau(\langle \overline{M}^N \rangle_{t+\delta} - \langle \overline{M}^N \rangle_t) \\ &= \tau \sum_{i=1}^N \int_t^{t+\delta} (b(s, X_s^i, \mu_s^N) - b(s, X_s^i, \mu_s))^\top c^{-1} (b(s, X_s^i, \mu_s^N) - b(s, X_s^i, \mu_s)) ds \\ &\leq \tau |\text{Tr}(c^{-1})|_\infty \sum_{i=1}^N \int_t^{t+\delta} |\xi_s^N(X_s^i)|^2 ds, \\ &\leq \kappa \begin{cases} N \int_t^{t+\delta} \mathcal{W}_1(\mu_s^N, \mu_s)^2 ds & \text{under 2.4(i),} \\ \int_t^{t+\delta} \sum_{i=1}^N \sum_{\ell=1}^{k-1} \left| \int_{(\mathbb{R}^d)^\ell} \delta_\mu^\ell b(s, X_s^i, y^\ell, \mu_s) (\mu_s^N - \mu_s)^{\otimes \ell} (dy^\ell) \right|^2 ds \\ \quad + N \int_t^{t+\delta} \mathcal{W}_1(\mu_s^N, \mu_s)^2 \wedge \mathcal{W}_1(\mu_s^N, \mu_s)^{2k} ds & \text{under 2.4(ii),} \\ \int_t^{t+\delta} \sum_{i=1}^N \left| \int_{\mathbb{R}^d} \tilde{b}(s, X_s^i, y) (\mu_s^N - \mu_s)(dy) \right|^2 ds, & \text{under 2.4(iii).} \end{cases} \quad (2.42) \end{aligned}$$

with $\kappa = \tau |\text{Tr}(c^{-1})|_\infty C$, where C is the constant of Lemma 2.21. We now heavily rely on the sharp deviation estimate

$$\sup_{0 \leq s \leq T} \overline{\mathbb{P}}^N(\mathcal{W}_1(\mu_s^N, \mu_s) \geq x) \lesssim \varepsilon_N(x), \quad (2.43)$$

with

$$\varepsilon_N(x) = \begin{cases} \exp(-\mathfrak{C}Nx^2) & \text{if } d = 1, \\ \exp(-\mathfrak{C}N \frac{x^2}{(\log(2+1/x))^2}) \mathbf{1}_{\{x \leq 1\}} + \exp(-\mathfrak{C}Nx^2) \mathbf{1}_{\{x > 1\}} & \text{if } d = 2, \\ \exp(-\mathfrak{C}Nx^d) \mathbf{1}_{\{x \leq 1\}} + \exp(-\mathfrak{C}Nx^2) \mathbf{1}_{\{x > 1\}} & \text{if } d \geq 3, \end{cases} \quad (2.44)$$

extracted from Theorem 2 of Fournier and Guillin [74]. Here $\mathfrak{C}$ depends on C_2 and d only, thanks to (2.38) that guarantees that Condition (1) of Theorem 2 in [74] is satisfied, hence the uniformity

in $s \in [0, T]$.

We complete the proof of (2.39) under Assumption 2.4 (i) that implies in particular $d = 1$. From (2.42), we infer

$$\begin{aligned}
 & \mathbb{E}_{\mathbb{P}^N} \left[\exp \left(\tau \left(\langle \overline{M}^N \rangle_{t+\delta} - \langle \overline{M}^N \rangle_t \right) \right) \right] \\
 & \leq \delta^{-1} \int_t^{t+\delta} \mathbb{E}_{\mathbb{P}^N} \left[\exp \left(\kappa \delta N \mathcal{W}_1(\mu_s^N, \mu_s)^2 \right) \right] ds \\
 & \leq \sup_{s \in [0, T]} \mathbb{E}_{\mathbb{P}^N} \left[\exp \left(\kappa \delta N \mathcal{W}_1(\mu_s^N, \mu_s)^2 \right) \right] \\
 & \leq 1 + \kappa \delta \sup_{s \in [0, T]} \int_0^\infty \exp(\kappa \delta z) \mathbb{P}^N(\mathcal{W}_1(\mu_s^N, \mu_s) \geq N^{-1/2} z^{1/2}) dz \\
 & \lesssim \int_0^\infty \exp((\kappa \delta - \mathfrak{C})z) dz,
 \end{aligned} \tag{2.45}$$

where $\mathfrak{C}$ is the constant in (2.44). The integral in (2.45) is finite as soon as $\delta \leq \tau^{-1} |\text{Tr}(c^{-1})|_\infty^{-1} |b|_{\text{Lip}}^{-2} \mathfrak{C}$ and (2.39) follows.

We next complete the proof of (2.39) under Assumption 2.4 (ii). When $d = 1$, we can rely on the previous case. Assume now that $(d = 2 \text{ and } k \geq 2)$ or $(d \geq 3 \text{ and } k \geq d/2)$. By Jensen's inequality

$$\mathbb{E}_{\mathbb{P}^N} \left[\exp \left(\tau \left(\langle \overline{M}^N \rangle_{t+\delta} - \langle \overline{M}^N \rangle_t \right) \right) \right] \leq I + II,$$

with

$$\begin{aligned}
 I &= \frac{1}{\delta k N} \int_t^{t+\delta} \sum_{i=1}^N \sum_{\ell=1}^{k-1} \mathbb{E}_{\mathbb{P}^N} \left[\exp \left(\kappa \delta k N \left| \int_{(\mathbb{R}^d)^\ell} \delta_\mu^\ell b(s, X_s^i, y^\ell, \mu_s) (\mu_s^N - \mu_s)^{\otimes \ell} (dy^\ell) \right|^2 \right) \right] ds, \\
 II &= \frac{1}{\delta k} \int_t^{t+\delta} \mathbb{E}_{\mathbb{P}^N} \left[\exp \left(\kappa \delta k N \mathcal{W}_1(\mu_s^N, \mu_s)^2 \wedge \mathcal{W}_1(\mu_s^N, \mu_s)^{2k} \right) \right] ds.
 \end{aligned}$$

We first estimate the remainder term II : by inequality (2.43), we have

$$\begin{aligned}
 II &\leq \frac{1}{k} \left(1 + \sup_{t \in [0, T]} \kappa \delta k \int_0^\infty e^{\kappa \delta k z} \mathbb{P}^N(N \mathcal{W}_1(\mu_t^N, \mu_t)^2 \wedge \mathcal{W}_1(\mu_t^N, \mu_t)^{2k} \geq z) dz \right) \\
 &\lesssim 1 + \sup_{t \in [0, T]} \int_0^N e^{\kappa \delta k z} \mathbb{P}^N(\mathcal{W}_1(\mu_t^N, \mu_t) \geq N^{-1/(2k)} z^{1/(2k)}) dz \\
 &\quad + \sup_{t \in [0, T]} \int_N^\infty e^{\kappa \delta k z} \mathbb{P}^N(\mathcal{W}_1(\mu_t^N, \mu_t) \geq N^{-1/2} z^{1/2}) dz.
 \end{aligned}$$

We first estimate the integral over $[0, N]$:

$$\begin{aligned}
 \int_0^N e^{\kappa \delta k z} \mathbb{P}^N(\mathcal{W}_1(\mu_t^N, \mu_t) \geq N^{-1/(2k)} z^{1/(2k)}) dz &\lesssim \int_0^N \exp(\kappa \delta k z - \mathfrak{C} N^{1-d/(2k)} z^{d/(2k)}) dz \\
 &= N \int_0^1 \exp(N(\kappa \delta k z - \mathfrak{C} z^{d/(2k)})) dz \lesssim 1
 \end{aligned}$$

for $2 < d \leq 2k$ as soon as $\delta \leq k^{-1}\kappa^{-1}\mathfrak{C}$. The case ($d = 2$ and $k \geq 2$) is slightly more technical but elementary and we omit it. For the integral over $[N, \infty)$, we proceed as under Assumption 2.4 (i) to obtain

$$\int_N^\infty e^{\kappa\delta kz} \mathbb{P}^N(\mathcal{W}_1(\mu_t^N, \mu_t) \geq N^{-1/2} z^{1/2}) dz \lesssim \int_0^\infty \exp((\kappa\delta k - \mathfrak{C})z) dz \lesssim 1$$

for $\delta \leq k^{-1}\kappa^{-1}\mathfrak{C}$ and we conclude $II \lesssim 1$ in that case.

We next turn to the term I . Observe first that by exchangeability

$$\begin{aligned} I &\leq \sup_{1 \leq \ell \leq k-1, t \in [0, T]} \mathbb{E}_{\mathbb{P}^N} \left[\exp(\kappa\delta k N \left| \int_{(\mathbb{R}^d)^\ell} \delta_\mu^\ell b(t, X_t^N, y^\ell, \mu_t) (\mu_t^N - \mu_t)^{\otimes \ell} (dy^\ell) \right|^2) \right] \\ &= 1 + \sum_{p \geq 1} \frac{(\kappa\delta k N)^p}{p!} \sup_{1 \leq \ell \leq k-1, t \in [0, T]} \mathbb{E}_{\mathbb{P}^N} \left[\left| \int_{(\mathbb{R}^d)^\ell} \delta_\mu^\ell b(t, X_t^N, y^\ell, \mu_t) (\mu_t^N - \mu_t)^{\otimes \ell} (dy^\ell) \right|^{2p} \right]. \end{aligned}$$

We then use the following estimate, reminiscent of moment bounds for U -statistics, however in a weaker and simpler form in our context. For an integer $\ell \geq 1$, we call $\mathcal{G}_\ell$ the class of functions $f : [0, T] \times \mathbb{R}^d \times (\mathbb{R}^d)^\ell \rightarrow \mathbb{R}^d$ that are Lipschitz continuous in the space variables.

Lemma 2.22. *Let $p \geq 1$. For $1 \leq \ell \leq k$, $f \in \mathcal{G}_\ell$ and $N \geq k + 1$, we have*

$$\mathcal{V}_{2p, \ell}^N(f(t, \cdot)) = \mathbb{E}_{\mathbb{P}^N} \left[\left| \int_{(\mathbb{R}^d)^\ell} f(t, X_t^N, y^\ell) (\mu_t^N - \mu_t)^{\otimes \ell} (dy^\ell) \right|^{2p} \right] \leq \frac{p! K_\ell^p}{(N - k)^p} \|f(t, \cdot)\|_{\text{Lip}}^{2p}, \quad (2.46)$$

for some explicitly computable $K_\ell = K_\ell(\mathbf{b}) > 0$.

The proof of Lemma 2.22 is quite elementary, yet technical, and is delayed until Appendix 8.2. The remainder of the proof of (2.39) is then straightforward : by Lemma 2.22 with $f(t, x, y^\ell) = \delta_\mu^\ell b(t, x, y^\ell, \mu_t)$, it follows that

$$I \leq 1 + \sum_{p \geq 1} (\kappa\delta k N)^p \frac{K_\ell^p}{(N - k)^p} \sup_{t \in [0, T]} |\delta_\mu^\ell b(t, \cdot, \mu_t)|_{\text{Lip}}^{2p} \lesssim 1$$

as soon as $\delta < \kappa^{-1} k^{-1} (k + 1)^{-1} K_\ell^{-1} \sup_{t \in [0, T]} |\delta_\mu^\ell b(t, \cdot, \mu_t)|_{\text{Lip}}^{-2}$. Since $II \lesssim 1$ is established as well, we obtain Proposition 2.19 under Assumption 2.4 (ii) provided Lemma 2.22 is proved.

We finally prove (2.39) under Assumption 2.4 (iii) when $d \geq 1$ is arbitrary. By (2.42) and Jensen's inequality together with the exchangeability of $(X_t^i)_{1 \leq i \leq N}$, we have

$$\begin{aligned} &\mathbb{E}_{\mathbb{P}^N} \left[\exp(\tau(\langle \overline{M}^N \rangle_{t+\delta} - \langle \overline{M}^N \rangle_t)) \right] \\ &\leq \frac{1}{\delta} \int_t^{t+\delta} \mathbb{E}_{\mathbb{P}^N} \left[\exp(\kappa\delta \sum_{i=1}^N \left| \int_{(\mathbb{R}^d)^\ell} \tilde{b}(s, X_s^i, y) (\mu_s^N - \mu_s)(dy) \right|^2) \right] ds \\ &\leq \sup_{s \in [0, T]} \mathbb{E}_{\mathbb{P}^N} \left[\exp(\kappa\delta N \left| \int_{\mathbb{R}^d} \tilde{b}(s, X_s^N, y) (\mu_s^N - \mu_s)(dy) \right|^2) \right] \\ &\leq 1 + \sum_{p \geq 1} \frac{(\kappa\delta N)^p}{p!} \sup_{s \in [0, T]} \mathbb{E}_{\mathbb{P}^N} \left[\left| \int_{\mathbb{R}^d} \tilde{b}(s, X_s^N, y) (\mu_s^N - \mu_s)(dy) \right|^{2p} \right] \lesssim 1 \end{aligned}$$

as soon as $\delta < \frac{1}{2} \kappa^{-1} K_1^{-1} \sup_{s \in [0, T]} |\tilde{b}(s, \cdot)|_{\text{Lip}}^{-2}$ by Lemma 2.22. Therefore (2.39) is established under Assumption 2.4 (iii) and this completes the proof of Proposition 2.19.

7 Proof of the nonparametric estimation results

We will repeatedly use estimates of the form

$$\int_{\nu}^{\infty} \exp(-z^r) dz \leq 2r^{-1} \nu^{1-r} \exp(-\nu^r), \quad \nu, r > 0, \quad \nu \geq (2/r)^{1/r}. \quad (2.47)$$

and

$$\int_0^{\infty} \exp\left(-\frac{az^p}{b + cz^{p/2}}\right) dz \leq C_p \max\left(\left(\frac{a}{b}\right)^{-1/p}, \left(\frac{a}{c}\right)^{-2/p}\right), \quad a, b, c, p > 0, \quad (2.48)$$

with $C_p = 2 \int_0^{\infty} \exp(-\frac{1}{2}(\min(\sqrt{z}, z))^p) dz$, stemming from the rough bound

$$\exp\left(-\frac{az^p}{b + cz^{p/2}}\right) \leq \exp\left(-\frac{az^p}{2b}\right) + \exp\left(-\frac{az^{p/2}}{2c}\right), \quad z > 0.$$

The estimate (2.48) is far from being optimal, but will be sufficient for our purpose.

7.1 Proof of Theorem 2.7

Preliminaries

We first state local upper and lower estimates on $(t, x) \mapsto \mu_t(x)$.

Lemma 2.23. *Work under Assumptions 2.1, 2.2 and 2.4. Let $(t_0, x_0) \in (0, T] \times \mathbb{R}^d$. Let $r > 0$ and $[r_1, r_2] \subset (0, T)$.*

(i) *There exists κ_5 depending on $(t_0, x_0), r$ and $\mathbf{b}$ such that*

$$\sup_{t \in [0, T], |x - x_0| \leq r} |b(t, x, \mu_t)| \leq \kappa_5. \quad (2.49)$$

(ii) *There exist κ_3, κ_4 depending on x_0, r_1, r_2, r and $\mathbf{b}$ such that*

$$0 < \kappa_4 \leq \inf_{t \in [r_1, r_2], |x - x_0| \leq r} \mu_t(x) \leq \sup_{t \in [r_1, r_2], |x - x_0| \leq r} \mu_t(x) \leq \kappa_3. \quad (2.50)$$

In turn, for a compactly supported kernel K , this implies the existence of $r = r(K)$ such that the estimate

$$|K_h(x_0 - \cdot)|_{L^2(\mu_{t_0})}^2 = \int_{\mathbb{R}^d} h^{-2d} K(h^{-1}x)^2 \mu_{t_0}(x_0 - x) dx \leq \kappa_3(r) h^{-d} |K|_2^2 \quad (2.51)$$

holds true.

Démonstration. For x such that $|x - x_0| \leq r$, we have

$$\begin{aligned} |b(t, x, \mu_t)| &\leq |b(t, 0, \delta_0)| + |b|_{\text{Lip}}(|x| + \mathcal{W}_1(\mu_t, \delta_0)) \\ &\leq \sup_{t \in [0, T]} |b(t, 0, \delta_0)| + |b|_{\text{Lip}}(|x_0| + r + \sup_{t \in [0, T]} \int_{\mathbb{R}^d} |y| \mu_t(dy) \end{aligned} \quad (2.52)$$

that defines κ_5 thanks to Assumption 2.4 and Lemma 2.20. This establishes (2.49). The estimate (2.50) follows from classical Gaussian tail estimates for the solution of parabolic equations. We refer

for example in our context to Corollary 8.2.2 of [17] : for every compact interval $[r_1, r_2] \subset (0, T)$, there exist constants $\mathbf{c}_\pm > 0$ depending on r_1, r_2 and $\mathbf{b}$ only such that

$$\exp(-\mathbf{c}_-(1 + |x|^2)) \leq \mu_t(x) \leq \exp(\mathbf{c}_+(1 + |x|^2))$$

for every $(t, x) \in [r_1, r_2] \times \mathbb{R}^d$. This establishes (2.50). Actually, if we moreover have 1/2-Hölder smoothness in time for the diffusion coefficient, investigating further Theorem 7.3.3 and Example 8.3.10 of [17], it is possible to prove $\sup_{x \in \mathbb{R}^d} \mu_t(x) < \infty$ uniformly in $t \in [r_1, r_2]$, hence κ_3 can be taken independently of r . $\square$

We next prove a standard bias-variance estimate for the quadratic risk of $\hat{\mu}_h^N(t_0, x_0)$.

Lemma 2.24. *In the setting of Theorem 2.7, if K is a bounded and compactly supported kernel and $h \in \mathcal{H}_1^N$, we have*

$$\mathbb{E}_{\mathbb{P}^N}[(\hat{\mu}_h^N(t_0, x_0) - \mu_{t_0}(x_0))^2] \lesssim \mathcal{B}_h^N(\mu)(t_0, x_0)^2 + \mathcal{V}_h^N,$$

up to a constant that depends (continuously) on (t_0, x_0) , $|K|_\infty$ and $\mathbf{b}$, and where $\mathcal{B}_h^N(\mu)(t_0, x_0)$ is defined in (2.14) and $\mathcal{V}_h^N$ in (2.12).

Démonstration. Write $\hat{\mu}_h^N(t_0, x_0) - \mu_{t_0}(x_0) = I + II$, with

$$I = \int_{\mathbb{R}^d} K_h(x_0 - x) \mu_{t_0}(x) dx - \mu_{t_0}(x_0)$$

and

$$II = \int_{\mathbb{R}^d} K_h(x_0 - x) (\mu_{t_0}^N(dx) - \mu_{t_0}(x) dx).$$

We have $I^2 \leq \mathcal{B}_h^N(\mu)(t_0, x_0)^2$ for the squared bias term. For the variance term, using successively Theorem 2.18 and the estimate (2.51) we have

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^N}[II^2] &= \int_0^\infty \mathbb{P}^N(|II| \geq z^{1/2}) dz \\ &\leq 2\kappa_1 \int_0^\infty \exp\left(-\frac{\kappa_2 N z}{|K_h(x_0 - \cdot)|_{L^2(\mu_{t_0})}^2 + |K_h(x_0 - \cdot)|_\infty z^{1/2}}\right) dz \\ &\leq 2\kappa_1 \int_0^\infty \exp\left(-\frac{\kappa_2 N h^d z}{\kappa_3 |K|_2^2 + |K|_\infty z^{1/2}}\right) dz \\ &\lesssim (N h^d)^{-1} (1 + (N h^d)^{-1}) \\ &\lesssim \mathcal{V}_h^N \end{aligned}$$

where we used (2.48) and the fact that $\max_{h \in \mathcal{H}_1^N} (N h^d)^{-1} \lesssim 1$. $\square$

Completion of proof of Theorem 2.7

We essentially repeat the main argument of the Goldenshluger-Lepski method (see *e.g.* [87, 88, 89] for the pointwise risk). We nevertheless give a proof for sake of completeness. Recall that $\hat{h}^N$

denotes the data-driven bandwidth defined in (2.13).

Step 1 : For $h \in \mathcal{H}_1^N$, we successively have

$$\begin{aligned}
 & \mathbb{E}_{\mathbb{P}^N} [(\hat{\mu}_{GL}^N(t_0, x_0) - \mu_{t_0}(x_0))^2] \\
 & \lesssim \mathbb{E}_{\mathbb{P}^N} [(\hat{\mu}_{GL}^N(t_0, x_0) - \hat{\mu}_h^N(t_0, x_0))^2] + \mathbb{E}_{\mathbb{P}^N} [(\hat{\mu}_h^N(t_0, x_0) - \mu_{t_0}(x_0))^2] \\
 & \lesssim \mathbb{E}_{\mathbb{P}^N} [\{(\hat{\mu}_{h^N}^N(t_0, x_0) - \hat{\mu}_h^N(t_0, x_0))^2 - \mathbf{V}_h^N - \mathbf{V}_{h^N}^N\}_+ + \mathbf{V}_h^N + \mathbf{V}_{h^N}^N] + \mathbb{E}_{\mathbb{P}^N} [(\hat{\mu}_h^N(t_0, x_0) - \mu_{t_0}(x_0))^2] \\
 & \lesssim \mathbb{E}_{\mathbb{P}^N} [\mathbf{A}_{\max(h^N, h)}^N + \mathbf{V}_h^N + \mathbf{V}_{h^N}^N] + \mathbb{E}_{\mathbb{P}^N} [(\hat{\mu}_h^N(t_0, x_0) - \mu_{t_0}(x_0))^2] \\
 & \lesssim \mathbb{E}_{\mathbb{P}^N} [\mathbf{A}_h^N] + \mathbf{V}_h^N + \mathbb{E}_{\mathbb{P}^N} [\mathbf{A}_{h^N}^N + \mathbf{V}_{h^N}^N] + \mathbb{E}_{\mathbb{P}^N} [(\hat{\mu}_h^N(t_0, x_0) - \mu_{t_0}(x_0))^2] \\
 & \lesssim \mathbb{E}_{\mathbb{P}^N} [\mathbf{A}_h^N] + \mathbf{V}_h^N + \mathcal{B}_h^N(\mu)(t_0, x_0)^2,
 \end{aligned}$$

where we applied Lemma 2.24 in order to obtain the last line.

Step 2 : We first estimate $\mathbf{A}_h^N$. Write $\mu_h(t_0, x_0)$ for $\int_{\mathbb{R}^d} K_h(x_0 - x) \mu_{t_0}(x) dx$. For $h, h' \in \mathcal{H}_1^N$ with $h' \leq h$, since

$$\begin{aligned}
 & (\hat{\mu}_h^N(t_0, x_0) - \hat{\mu}_{h'}^N(t_0, x_0))^2 \\
 & \leq 4(\hat{\mu}_h^N(t_0, x_0) - \mu_h(t_0, x_0))^2 + 4(\mu_h(t_0, x_0) - \mu_{t_0}(x_0))^2 + 4(\mu_{h'}(t_0, x_0) - \mu_{t_0}(x_0))^2 \\
 & \quad + 4(\hat{\mu}_{h'}^N(t_0, x_0) - \mu_{h'}(t_0, x_0))^2,
 \end{aligned}$$

we have

$$\begin{aligned}
 & (\hat{\mu}_h^N(t_0, x_0) - \hat{\mu}_{h'}^N(t_0, x_0))^2 - \mathbf{V}_h^N - \mathbf{V}_{h'}^N \leq 8\mathcal{B}_h^N(\mu)(t_0, x_0)^2 + (4(\hat{\mu}_h^N(t_0, x_0) - \mu_h(t_0, x_0))^2 - \mathbf{V}_h^N) \\
 & \quad + (4(\hat{\mu}_{h'}^N(t_0, x_0) - \mu_{h'}(t_0, x_0))^2 - \mathbf{V}_{h'}^N)
 \end{aligned}$$

using $h' \leq h$ in order to bound $(\hat{\mu}_{h'}^N(t_0, x_0) - \mu_{h'}(t_0, x_0))^2$ by the bias at scale h . Taking maximum over $h' \leq h$, we obtain

$$\begin{aligned}
 & \max_{h' \leq h} \{(\hat{\mu}_h^N(t_0, x_0) - \hat{\mu}_{h'}^N(t_0, x_0))^2 - \mathbf{V}_h^N - \mathbf{V}_{h'}^N\}_+ \\
 & \leq 8\mathcal{B}_h^N(\mu)(t_0, x_0)^2 + \{4(\hat{\mu}_h^N(t_0, x_0) - \mu_h(t_0, x_0))^2 - \mathbf{V}_h^N\}_+ \\
 & \quad + \max_{h' \leq h} \{4(\hat{\mu}_{h'}^N(t_0, x_0) - \mu_{h'}(t_0, x_0))^2 - \mathbf{V}_{h'}^N\}_+.
 \end{aligned} \tag{2.53}$$

Step 3 : We estimate the expectation of the first stochastic term in the right-hand side of (2.53). We refine the computation of the term II in the proof of Lemma 2.24. By Theorem 2.18 and using estimates of the form (2.47) and (2.48), we have

$$\begin{aligned}
 & \mathbb{E}_{\mathbb{P}^N} [\{4(\hat{\mu}_h^N(t_0, x_0) - \mu_h(t_0, x_0))^2 - \mathbf{V}_h^N\}_+] \\
 & = \int_0^\infty \mathbb{P}^N(4(\hat{\mu}_h^N(t_0, x_0) - \mu_h(t_0, x_0))^2 - \mathbf{V}_h^N \geq z) dz \\
 & = \int_0^\infty \mathbb{P}^N(|\hat{\mu}_h^N(t_0, x_0) - \mu_h(t_0, x_0)| \geq \frac{1}{2}(\mathbf{V}_h^N + z)^{1/2}) dz
 \end{aligned}$$

$$\begin{aligned}
 &\leq 2\kappa_1 \int_{V_h^N}^\infty \exp\left(-\frac{\kappa_2 N h^{\frac{d}{4}} z}{\kappa_3 |K|_2^2 + |K|_\infty^{\frac{1}{2}} z^{1/2}}\right) dz \\
 &\lesssim \int_{V_h^N}^\infty \exp\left(-\frac{\kappa_2 N h^d z}{8\kappa_3 |K|_2^2}\right) dz + \int_{V_h^N}^\infty \exp\left(-\frac{\kappa_2 N h^d z^{1/2}}{4|K|_\infty}\right) dz \\
 &\lesssim (N h^d)^{-1} \exp\left(-\frac{\kappa_2 N h^d V_h^N}{8\kappa_3 |K|_2^2}\right) + (N h^d)^{-2} N h^d (V_h^N)^{1/2} \exp\left(-\frac{\kappa_2 N h^d (V_h^N)^{1/2}}{4|K|_\infty}\right) \\
 &\lesssim (N h^d)^{-1} N^{-\varpi_1 \kappa_2 / (8\kappa_3)} + (N h^d)^{-3/2} (\log N)^{1/2} \exp\left(-\frac{\kappa_2 |K|_2 \varpi_1^{1/2}}{4|K|_\infty} (\log N)^{5/2}\right), \\
 &\lesssim N^{-2}
 \end{aligned}$$

as soon as $\varpi_1 \geq 16\kappa_2^{-1}\kappa_3$, thanks to $\max_{h \in \mathcal{H}_1^N} (N h^d)^{-1} \lesssim 1$, and using $\min_{h \in \mathcal{H}_1^N} h \geq (N^{-1}(\log N)^2)^{1/d}$ to show that the second term is negligible in front of N^{-2} .

Step 4 : For the second stochastic term, we have the rough estimate

$$\begin{aligned}
 &\mathbb{E}_{\mathbb{P}^N} \left[\max_{h' \leq h} \left\{ 4(\hat{\mu}_{h'}^N(t_0, x_0) - \mu_{h'}(t_0, x_0))^2 - V_{h'}^N \right\}_+ \right] \\
 &\leq \sum_{h' \leq h} \mathbb{E}_{\mathbb{P}^N} \left[\left\{ 4(\hat{\mu}_{h'}^N(t_0, x_0) - \mu_{h'}(t_0, x_0))^2 - V_{h'}^N \right\}_+ \right] \lesssim \text{Card}(\mathcal{H}_1^N) N^{-2} \lesssim N^{-1}
 \end{aligned}$$

where we used Step 3 to bound each term $\mathbb{E}_{\mathbb{P}^N} \left[\left\{ 4(\hat{\mu}_{h'}^N(t_0, x_0) - \mu_{h'}(t_0, x_0))^2 - V_{h'}^N \right\}_+ \right]$ independently of h together with $\text{Card}(\mathcal{H}_1^N) \lesssim N$. In conclusion, we have through Steps 2-4 that $\mathbb{E}_{\mathbb{P}^N} [A_h^N] \lesssim N^{-1} + \mathcal{B}_h^N(\mu)(t_0, x_0)^2$. Therefore, from Step 1, we conclude

$$\mathbb{E}_{\mathbb{P}^N} \left[(\hat{\mu}_{GL}^N(t_0, x_0) - \mu_{t_0}(x_0))^2 \right] \lesssim \mathcal{B}_h^N(\mu)(t_0, x_0)^2 + V_h^N + N^{-1}$$

for any $h \in \mathcal{H}_1^N$. Since $N^{-1} \lesssim V_h^N$ always, the proof of Theorem 2.7 is complete.

7.2 Proof of Theorem 2.9

Preliminaries

The assumptions of Theorem 2.9 are in force in this section. We first study the fluctuations of the random measure $\pi^N(dt, dx) - \pi(t, x)dt dx$, where $\pi^N(dt, dx) = N^{-1} \sum_{i=1}^N \delta_{X_t^i}(dx) X^i(dt)$.

Lemma 2.25. *Let $\phi : (0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ be bounded and compactly supported. The following decomposition holds*

$$\begin{aligned}
 &\int_{[0, T] \times \mathbb{R}^d} \phi(t, x) (\pi^N(dt, dx) - \pi(t, x)dt dx) \\
 &= \int_0^T \int_{\mathbb{R}^d} \phi(t, x) (b(t, x, \mu_t)(\mu_t^N(dx) - \mu_t(x)dx) + \xi_t^N(x) \mu_t^N(x)dx) dt + \mathcal{M}_T^N(\phi),
 \end{aligned}$$

where $\xi_t^N(x) = b(t, x, \mu_t^N) - b(t, x, \mu_t)$ and $\mathcal{M}_t^N(\phi) = (\mathcal{M}_t^N(\phi)^1, \dots, \mathcal{M}_t^N(\phi)^d)$ is a d -dimensional $\mathbb{P}^N$ -continuous martingale with predictable compensator such that

$$\langle \mathcal{M}_t^N(\phi)^k \rangle_t \leq N^{-1} |\text{Tr}(c)|_\infty \int_0^t \int_{\mathbb{R}^d} \phi(s, x)^2 \mu_s^N(dx) ds. \quad (2.54)$$

Démonstration. we have

$$\begin{aligned}
 & \int_{[0,T] \times \mathbb{R}^d} \phi(t, x) (\pi^N(dt, dx) - \pi(t, x) dt dx) \\
 &= N^{-1} \sum_{i=1}^N \int_0^T \phi(t, X_t^i) (\sigma(t, X_t^i) dB_t^i + b(t, X_t^i, \mu_t^N) dt) - \int_{[0,T] \times \mathbb{R}^d} \phi(t, x) b(t, x, \mu_t) \mu_t(x) dt dx \\
 &= \int_0^T \int_{\mathbb{R}^d} \phi(t, x) (b(t, x, \mu_t^N) \mu_t^N(dx) - b(t, x, \mu_t) \mu_t(x) dx) dt + \mathcal{M}_T^N(\phi),
 \end{aligned}$$

where

$$\mathcal{M}_t^N(\phi) = N^{-1} \sum_{i=1}^N \int_0^t \phi(s, X_s^i) \sigma(s, X_s^i) dB_s^i$$

is a martingale with bracket satisfying (2.54). The result follows. $\square$

We next have a bias-variance estimate for the quadratic risk of $\hat{\pi}_{\mathbf{h}}(t_0, x_0)$, in the same spirit as in Lemma 2.24.

Lemma 2.26. *Assume that $H \otimes K$ is a bounded and compactly supported kernel on $(0, T) \times \mathbb{R}^d$. Let $\mathbf{h} \in \mathcal{H}_2^N$. Then*

$$\mathbb{E}_{\mathbb{P}^N} [|\hat{\pi}_{\mathbf{h}}^N(t_0, x_0) - \pi(t_0, x_0)|^2] \lesssim \mathcal{B}_{\mathbf{h}}^N(\pi)(t_0, x_0)^2 + \mathcal{V}_{\mathbf{h}}^N,$$

up to a constant that (continuously) depends on (t_0, x_0) , $|H \otimes K|_{\infty}$ and $\mathbf{b}$, and where $\mathcal{B}_{\mathbf{h}}^N(\pi)(t_0, x_0)$ is defined in (2.19) and $\mathcal{V}_{\mathbf{h}}^N$ in (2.17).

Démonstration. Write $\hat{\pi}_{\mathbf{h}}^N(t_0, x_0) - \pi(t_0, x_0) = I + II$, with

$$I = \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) \pi(t, x) dx dt - \pi(t_0, x_0)$$

and

$$II = \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) (\pi^N(dt, dx) - \pi(t, x) dt dx).$$

We have $|I|^2 \leq \mathcal{B}_{\mathbf{h}}^N(\pi)(t_0, x_0)^2$ for the squared bias term. For the variance term, applying the decomposition of Lemma 2.25 with test function $\phi(t, x) = (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x)$, we obtain

$$|II|^2 \lesssim III + IV + V,$$

with

$$\begin{aligned}
 III &= \left| \int_0^T \int_{\mathbb{R}^d} \phi(t, x) b(t, x, \mu_t) (\mu_t^N(dx) - \mu_t(x) dx) dt \right|^2, \\
 IV &= \left| \int_0^T \int_{\mathbb{R}^d} \phi(t, x) \xi_t^N(x) \mu_t^N(dx) dt \right|^2, \\
 V &= |\mathcal{M}_T^N(\phi)|^2,
 \end{aligned}$$

where $\xi_t^N(x) = b(t, x, \mu_t^N) - b(t, x, \mu_t)$. Writing $b = (b^1, \dots, b^d)$ in components, note first that for $\nu(dt, dx) = \mu_t(dx)T^{-1}dt$, we have

$$|\phi b^k(\cdot, \mu_\cdot)|_{L^2(\nu)}^2 \leq \kappa_3 \kappa_5^2 T^{-1} |H \otimes K|_2^2 (h_1 h_2^d)^{-1}, \quad |\phi b^k(\cdot, \mu_\cdot)|_\infty \leq \kappa_5 |H \otimes K|_\infty (h_1 h_2^d)^{-1}, \quad (2.55)$$

by Lemma 2.23 and the compactness of the support of ϕ . By Theorem 2.18 applied to $\nu^N(dt, dx) - \nu(dt, dx) = (\mu_t^N(dx) - \mu_t(dx))T^{-1}dt$, it follows that

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^N}[III] &\lesssim \sum_{k=1}^d \int_0^\infty \mathbb{P}^N(|\int_{[0,T] \times \mathbb{R}^d} \phi(t, x) b^k(t, x, \mu_t) (\mu_t^N(dx) - \mu_t(x)dx) T^{-1}dt| \geq z^{1/2}) dz \\ &\leq 2d\kappa_1 \int_0^\infty \exp\left(-\frac{\kappa_2 N h_1 h_2^d z}{\kappa_3 \kappa_5^2 T^{-1} |H \otimes K|_2^2 + \kappa_5 |H \otimes K|_\infty z^{1/2}}\right) dz \\ &\lesssim (N h_1 h_2^d)^{-1} (1 + (N h_1 h_2^d)^{-1}) \\ &\lesssim V_{\mathbf{h}}^N, \end{aligned}$$

using $\max_{\mathbf{h} \in \mathcal{H}_2^N} (N h_1 h_2^d) \lesssim 1$. We conclude

$$\mathbb{E}_{\mathbb{P}^N}[III] \lesssim V_{\mathbf{h}}^N. \quad (2.56)$$

We next turn to the term *IV*. We need a deviation result for the fluctuation $\xi_t^N(x) = b(t, x, \mu_t^N) - b(t, x, \mu_t)$ that will also be helpful later.

Lemma 2.27. *There exist positive numbers κ_6, κ_7 and κ_8 , depending on $\mathbf{b}$, such that for large enough N*

$$\sup_{0 \leq t \leq T} \mathbb{P}^N(|\xi_t^N(X_t^N)| \geq u) \leq \kappa_6 \exp\left(-\frac{\kappa_7 N u^2}{1 + N^{1/2} u}\right) \text{ for } u \geq \kappa_8 N^{-1/2}.$$

Démonstration. Writing $\xi_t^N(X_t^N) = (\xi_t^N(X_t^N)^1, \dots, \xi_t^N(X_t^N)^d)$ in components, we have

$$\mathbb{P}^N(|\xi_t^N(X_t^N)| \geq u) \leq \sum_{\ell=1}^d \mathbb{P}^N(|\xi_t^N(X_t^N)^\ell| \geq u d^{-1})$$

It suffices thus to prove the result for each component $\xi_t^N(X_t^N)^\ell$, substituting κ_6 and κ_8 by $d\kappa_6$ and $d\kappa_8$ to obtain the general case. For notational simplicity, we drop the superscript ℓ and prove the result for $\xi_t^N(X_t^N)$ instead of $|\xi_t^N(X_t^N)|$, up to a inflation of κ_6 by a factor 2.

By (2.40) in the proof of Proposition 2.19, we may write

$$\xi_t^N(X_t^N) = \zeta_t^N(X_t^N) + \mathcal{R}_k,$$

with

$$\zeta_t^N(X_t^N) = \sum_{\ell=1}^{k-1} \frac{1}{\ell!} \int_{(\mathbb{R}^d)^\ell} \delta_\mu^\ell b(t, X_t^N, y^\ell, \mu_t) (\mu_t^N - \mu_t)^{\otimes \ell} (dy^\ell),$$

having $k = 1$ under Assumption 2.4 (i), with $|\mathcal{R}_k| \lesssim \mathcal{W}_1(\mu_t^N, \mu_t) \wedge \mathcal{W}_1(\mu_t^N, \mu_t)^k$ under Assumption 2.4 (ii) by (2.41), and having $k = 1$ with $\delta_\mu^1 b = \bar{b}$ and $\mathcal{R}_k = 0$ under Assumption 2.4 (iii). It is

enough to prove the deviation bound for each term separately.

Let $u \geq 0$. We first bound the remainder term $\mathcal{R}_k$. Applying (2.35) in the proof of Theorem 2.18 for the event $\mathcal{A}^N = \{\mathcal{W}_1(\mu_t^N, \mu_t) \wedge \mathcal{W}_1(\mu_t^N, \mu_t)^k \geq u\}$ we obtain

$$\begin{aligned} \mathbb{P}^N(\mathcal{W}_1(\mu_t^N, \mu_t) \wedge \mathcal{W}_1(\mu_t^N, \mu_t)^k \geq u) &\leq C_1^{K(K+1)/8} \bar{\mathbb{P}}^N(\mathcal{W}_1(\mu_t^N, \mu_t) \wedge \mathcal{W}_1(\mu_t^N, \mu_t)^k \geq u)^{4^{-K}} \\ &\leq C_1^{K(K+1)/8} (\varepsilon_N(u^{1/k}) \wedge \varepsilon_N(u))^{4^{-K}} \end{aligned}$$

where the last estimate stems from the deviation inequality (2.43) of Fournier and Guillin [74]. Under Assumption 2.4 (i), with $(d = 1$ and $k = 1)$ or under Assumption 2.4 (ii) with $(d = 2$ and $k \geq 2)$ or $(d \geq 3$ and $k \geq d/2)$,

$$\varepsilon_N(u^{1/k}) \wedge \varepsilon_N(u) \lesssim \exp(-\mathfrak{C}Nu^2) \text{ for every } u \geq 0$$

as follows from the definition of $\varepsilon_N(x)$ in (2.43). Therefore $\mathcal{R}_k$ has the right order. As for the main term, we first note that for every $p \geq 2$, we have

$$\begin{aligned} \mathbb{E}_{\bar{\mathbb{P}}^N} [|\zeta_t^N(X_t^N)|^p] &\leq e^{p-1} \sum_{\ell=1}^{k-1} \frac{1}{\ell!} \mathbb{E}_{\bar{\mathbb{P}}^N} \left[\left| \int_{(\mathbb{R}^d)^\ell} \delta_\mu^\ell b(t, X_t^N, y^\ell, \mu_t) (\mu_t^N - \mu_t)^{\otimes \ell} (dy^\ell) \right|^p \right] \\ &\leq N^{-p/2} p! C_5^p, \end{aligned}$$

by Lemma 2.22 and Cauchy-Schwarz's inequality, for large enough N and some C_5 depending on $\mathfrak{b}$. With no loss of generality, we take $C_5 \geq 1$. In particular, by Cauchy-Schwarz's inequality,

$$|\mathbb{E}_{\bar{\mathbb{P}}^N} [\zeta_t^N(X_t^N)]| \leq \sqrt{2} C_5 N^{-1/2} = \kappa_8 N^{-1/2}$$

that defines the constant κ_8 . We next use the following version of Bernstein inequality that can be found in Lemma 8 in Birgé and Massart [15] : if Z is a real-valued random variable such that $\mathbb{E}[|Z|^p] \leq \frac{p!}{2} v^2 c^{p-2}$ for $c, v > 0$ and every $p \geq 2$, then

$$\mathbb{P}(Z - \mathbb{E}[Z] \geq u) \leq \exp \left(- \frac{u^2/2}{v^2 + cu} \right) \text{ for every } u \geq 0. \quad (2.57)$$

We then apply (2.57) to $Z = \zeta_t^N(X_t^N)$ with $c = C_5 N^{-1/2}$ and $v = \sqrt{2} N^{-1/2} C_5$ and obtain

$$\bar{\mathbb{P}}^N(\zeta_t^N(X_t^N) - \mathbb{E}_{\bar{\mathbb{P}}^N}[\zeta_t^N(X_t^N)] \geq u) \leq \exp \left(- \frac{C_6 N u^2}{1 + N^{1/2} u} \right) \text{ for every } u \geq 0,$$

with $C_6 = (4C_5^2)^{-1}$ using $C_5 \geq 1$. Finally, for $u \geq \kappa_8 N^{-1/2}$, setting $u' = u - \kappa_8 N^{-1/2}$ and applying (2.35) in the proof of Theorem 2.18 for the event $\{\zeta_t^N(X_t^N) \geq u\}$, we derive

$$\begin{aligned} \mathbb{P}^N(\zeta_t^N(X_t^N) \geq u) &\lesssim \bar{\mathbb{P}}^N(\zeta_t^N(X_t^N) \geq u)^{4^{-K}} \\ &\leq \bar{\mathbb{P}}^N(\zeta_t^N(X_t^N) - \mathbb{E}_{\bar{\mathbb{P}}^N}[\zeta_t^N(X_t^N)] \geq u')^{4^{-K}} \\ &\leq \exp \left(- \frac{C_6 4^{-K} N (u')^2}{1 + N^{1/2} u'} \right) \\ &\lesssim \exp \left(- \frac{C_6 4^{-K} N u^2}{1 + N^{1/2} u} \right) \end{aligned}$$

and the lemma follows with $\kappa_7 = 4^{-K} C_6$. $\square$

We are ready to bound the term IV . Applying Cauchy-Schwarz's inequality twice, we have

$$\mathbb{E}_{\mathbb{P}^N}[IV] \leq \mathbb{E}_{\mathbb{P}^N} \left[\left(\int_0^T \int_{\mathbb{R}^d} |\phi(t, x)|^2 \mu_t^N(dx) dt \right)^2 \right]^{\frac{1}{2}} \mathbb{E}_{\mathbb{P}^N} \left[\left(\int_0^T \int_{\mathbb{R}^d} |\xi_t^N(x)|^2 \mu_t^N(dx) dt \right)^2 \right]^{\frac{1}{2}}.$$

On the one hand, by exchangeability and Lemma 2.27, we have

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^N} \left[\left(\int_0^T \int_{\mathbb{R}^d} |\xi_t^N(x)|^2 \mu_t^N(dx) dt \right)^2 \right]^{\frac{1}{2}} &\leq \left(T \int_0^T \mathbb{E}_{\mathbb{P}^N} \left[|\xi_t^N(X_t^N)|^4 \right] dt \right)^{\frac{1}{2}} \\ &\lesssim \sup_{0 \leq t \leq T} \left(\int_0^\infty \mathbb{P}^N(|\xi_t^N(X_t^N)| \geq z^{1/4}) dz \right)^{1/2} \\ &\lesssim \left(\kappa_8^4 N^{-2} + \kappa_6 \int_0^\infty \exp \left(-\frac{\kappa_7 N z^{1/2}}{1 + N^{1/2} z^{1/4}} \right) dz \right)^{1/2} \\ &\lesssim N^{-1}. \end{aligned} \quad (2.58)$$

On the other hand

$$\begin{aligned} &\mathbb{E}_{\mathbb{P}^N} \left[\left(\int_0^T \int_{\mathbb{R}^d} \phi(t, x)^2 \mu_t^N(dx) dt \right)^2 \right]^{1/2} \\ &\lesssim \int_0^T \int_{\mathbb{R}^d} \phi(t, x)^2 \mu_t(x) dx dt + \mathbb{E}_{\mathbb{P}^N} \left[\left| \int_{[0, T] \times \mathbb{R}^d} \phi(t, x)^2 (\mu_t^N(dx) - \mu_t(dx)) dt \right|^2 \right]^{1/2} \\ &\lesssim |\phi|_2^2 + \left(\int_0^\infty \mathbb{P}^N \left(\left| \int_{[0, T] \times \mathbb{R}^d} \phi(t, x)^2 (\mu_t^N(dx) - \mu_t(dx)) T^{-1} dt \right| \geq z^{1/2} \right) \right)^{1/2} \\ &\lesssim |\phi|_2^2 + \left(2\kappa_1 \int_0^\infty \exp \left(-\frac{\kappa_2 N z}{\kappa_3 T^{-1} |\phi|_4^4 + |\phi|_\infty^2 z^{1/2}} \right) dz \right)^{1/2} \\ &\lesssim |\phi|_2^2 + (N^{-1} |\phi|_4^4 + N^{-2} |\phi|_\infty^4)^{1/2} \lesssim |\phi|_2^2 + N^{-1/2} |\phi|_4^2 + N^{-1} |\phi|_\infty^2, \end{aligned} \quad (2.59)$$

using Lemma 2.23 and the fact that ϕ is compactly supported to obtain the first term and Theorem 2.18 applied to $\nu^N(dt, dx) - \nu(dt, dx) = (\mu_t^N(dx) - \mu_t(dx)) T^{-1} dt$ together with $|\phi^2|_{L^2(\nu)}^2 \leq \kappa_3 T^{-1} |\phi|_4^4$ to obtain the second term. Putting together (2.58) and (2.59), we conclude

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^N}[IV] &\lesssim N^{-1} (|\phi|_2^2 + N^{-1/2} |\phi|_4^2 + N^{-1} |\phi|_\infty^2) \\ &\lesssim (Nh_1 h_2^d)^{-1} (1 + (Nh_1 h_2^d)^{-1/2} + (Nh_1 h_2^d)^{-1}) \lesssim \mathbf{V}_{\mathbf{h}}^N. \end{aligned} \quad (2.60)$$

Finally, by Lemma 2.25 and Theorem 2.18 again we have

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^N}[V] &= \sum_{k=1}^d \mathbb{E}_{\mathbb{P}^N} [\langle \mathcal{M}^N(\phi)^k \rangle_T] \\ &\leq d N^{-1} |\text{Tr}(c)|_\infty \mathbb{E}_{\mathbb{P}^N} \left[\int_{[0, T] \times \mathbb{R}^d} \phi(s, x)^2 \mu_s^N(dx) ds \right] \\ &\lesssim N^{-1} |\phi|_2^2 + N^{-1} \mathbb{E}_{\mathbb{P}^N} \left[\left| \int_{[0, T] \times \mathbb{R}^d} \phi(s, x)^2 (\mu_s^N(dx) - \mu_s(dx)) T^{-1} ds \right|^2 \right] \\ &\lesssim N^{-1} |\phi|_2^2 + N^{-1} \int_0^\infty \mathbb{P}^N \left(\left| \int_{[0, T] \times \mathbb{R}^d} \phi(s, x)^2 (\mu_s^N(dx) - \mu_s(x)) T^{-1} ds \right| \geq z \right) dz \end{aligned}$$

$$\begin{aligned}
 &\leq N^{-1}|\phi|_2^2 + N^{-1}2\kappa_1 \int_0^\infty \exp\left(-\frac{\kappa_2 N z^2}{\kappa_3 T^{-1}|\phi|_4^4 + |\phi|_\infty^2 z}\right) dz \\
 &\lesssim N^{-1}|\phi|_2^2 + N^{-3/2}|\phi|_4^2 + N^{-2}|\phi|_\infty^2 \\
 &\lesssim (Nh_1 h_2^d)^{-1} (1 + (Nh_1 h_2^d)^{-1/2} + (Nh_1 h_2^d)^{-1}) \lesssim \mathbf{V}_{\mathbf{h}}^N.
 \end{aligned} \tag{2.61}$$

Putting together (2.56), (2.60) and (2.61) establishes $\mathbb{E}_{\mathbb{P}^N}[II^2] \lesssim \mathbf{V}_{\mathbf{h}}^N$ and concludes the proof of Lemma 2.26. $\square$

Completion of proof of Theorem 2.9

Let $(h, \mathbf{h}) \in \mathcal{H}_1^N \times \mathcal{H}_2^N$ and $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$. Remember that we set $\pi(t, x) = b(t, x, \mu_t)\mu_t(x)$.

Step 1 : We plan to use the decomposition

$$\widehat{b}_{h, \mathbf{h}}^N(t_0, x_0)_{\varpi_3} - b(t_0, x_0, \mu_{t_0}) = I + II,$$

with

$$I = \frac{\pi(t_0, x_0)(\mu_{t_0}(x_0) - \widehat{\mu}_{\mathbf{h}}^N(t_0, x_0) \vee \varpi_3)}{\mu_{t_0}(x_0)\widehat{\mu}_{\mathbf{h}}^N(t_0, x_0) \vee \varpi_3}$$

and

$$II = \frac{(\widehat{\pi}_{\mathbf{h}}^N(t_0, x_0) - \pi(t_0, x_0))\mu_{t_0}(x_0)}{\mu_{t_0}(x_0)\widehat{\mu}_{\mathbf{h}}^N(t_0, x_0) \vee \varpi_3}.$$

First, we have

$$|I| \leq \frac{\kappa_5}{\varpi_3} |\mu_{t_0}(x_0) - \widehat{\mu}_{\mathbf{h}}^N(t_0, x_0) \vee \varpi_3| \lesssim |\mu_{t_0}(x_0) - \widehat{\mu}_{\mathbf{h}}^N(t_0, x_0)|$$

as soon as $\varpi_3 \leq \kappa_4$ by Lemma 2.23, for some (small) $r > 0$ fixed throughout. In the same way,

$$|II| \leq \varpi_3^{-1} |\widehat{\pi}_{\mathbf{h}}^N(t_0, x_0) - \pi(t_0, x_0)|.$$

Picking $h = \widehat{h}^N$, $\mathbf{h} = \widehat{\mathbf{h}}^N$, taking square and expectation, we have thus established

$$\begin{aligned}
 &\mathbb{E}_{\mathbb{P}^N} [|\widehat{b}_{\widehat{h}, \widehat{\mathbf{h}}}^N(t_0, x_0) - b(t_0, x_0, \mu_{t_0})|^2] \\
 &\lesssim \mathbb{E}_{\mathbb{P}^N} [(\widehat{\mu}_{\widehat{\mathbf{h}}}^N(t_0, x_0) - \mu_{t_0}(x_0))^2] + \mathbb{E}_{\mathbb{P}^N} [|\widehat{\pi}_{\widehat{\mathbf{h}}}^N(t_0, x_0) - \pi(t_0, x_0)|^2]
 \end{aligned} \tag{2.62}$$

as soon as $\varpi_3 \leq \kappa_4$. By Theorem 2.7, we already have the desired bound for the first term.

Step 2 : We study the second term in the right-hand side of (2.62). For any $\mathbf{h} \in \mathcal{H}_2^N$, similarly to the proof of Step 1 in Theorem 2.7, we have

$$\mathbb{E}_{\mathbb{P}^N} [|\widehat{\pi}_{\mathbf{h}}^N(t_0, x_0) - \pi(t_0, x_0)|^2] \lesssim \mathbb{E}_{\mathbb{P}^N} [\mathbf{A}_{\mathbf{h}}^N] + \mathbf{V}_{\mathbf{h}}^N + \mathcal{B}_{\mathbf{h}}^N(\pi)(t_0, x_0)^2$$

thanks to Lemma 2.26. In order to estimate $\mathbb{E}_{\mathbb{P}^N} [\mathbf{A}_{\mathbf{h}}^N]$, we repeat Step 2 of the proof of Theorem 2.7 and obtain

$$\max_{\mathbf{h}' \preceq \mathbf{h}} \{|\widehat{\pi}_{\mathbf{h}}^N(t_0, x_0) - \widehat{\pi}_{\mathbf{h}'}^N(t_0, x_0)|^2 - \mathbf{V}_{\mathbf{h}}^N - \mathbf{V}_{\mathbf{h}'}^N\}_+ \tag{2.63}$$

$$\begin{aligned} &\lesssim \mathcal{B}_{\mathbf{h}}^N(\pi)(t_0, x_0)^2 + \left\{ 4 \left| \widehat{\pi}_{\mathbf{h}}^N(t_0, x_0) - \pi_{\mathbf{h}}(t_0, x_0) \right|^2 - \mathbf{V}_{\mathbf{h}}^N \right\}_+ \\ &\quad + \max_{\mathbf{h}' \preceq \mathbf{h}} \left\{ 4 \left| \widehat{\pi}_{\mathbf{h}'}^N(t_0, x_0) - \pi_{\mathbf{h}'}(t_0, x_0) \right|^2 - \mathbf{V}_{\mathbf{h}'}^N \right\}_+ \end{aligned}$$

with the notation $\pi_{\mathbf{h}}(t_0, x_0) = \int_0^T \int_{\mathbb{R}^d} (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x) \pi(t, x) dx dt$.

Step 3 : We estimate the expectation of the first stochastic term in the right-hand side of (2.63). By Lemma 2.25, setting $\phi(t, x) = (H \otimes K)_{\mathbf{h}}(t_0 - t, x_0 - x)$, we have

$$\left\{ 4 \left| \widehat{\pi}_{\mathbf{h}}^N(t_0, x_0) - \pi_{\mathbf{h}}(t_0, x_0) \right|^2 - \mathbf{V}_{\mathbf{h}}^N \right\}_+ \leq I + II + III,$$

with

$$\begin{aligned} I &= \left\{ 12 \left| \int_{[0, T] \times \mathbb{R}^d} \phi(t, x) b(t, x, \mu_t) (\mu_t^N(dx) - \mu_t(x) dx) dt \right|^2 - \frac{1}{3} \mathbf{V}_{\mathbf{h}}^N \right\}_+, \\ II &= \left\{ 12 \left| \int_{[0, T] \times \mathbb{R}^d} \phi(t, x) \xi_t^N(x) \mu_t^N(dx) dt \right|^2 - \frac{1}{3} \mathbf{V}_{\mathbf{h}}^N \right\}_+, \\ III &= \left\{ 12 \left| \mathcal{M}_T^N(\phi) \right|^2 - \frac{1}{3} \mathbf{V}_{\mathbf{h}}^N \right\}_+. \end{aligned}$$

For the term I , writing $b = (b^1, \dots, b^d)$ in components, we have

$$I \leq 12T^2 \sum_{k=1}^d \left\{ \left(\int_{[0, T] \times \mathbb{R}^d} \phi(t, x) b^k(t, x, \mu_t) (\mu_t^N(dx) - \mu_t(x) dx) T^{-1} dt \right)^2 - \frac{1}{36dT^2} \mathbf{V}_{\mathbf{h}}^N \right\}_+.$$

Applying Theorem 2.18 to $(\mu_t^N(dx) - \mu_t(x) dx) T^{-1} dt$ and using the estimates (2.55) of Lemma 2.26 above, we infer

$$\begin{aligned} &\mathbb{E}_{\mathbb{P}^N} \left[\left\{ \left(\int_{[0, T] \times \mathbb{R}^d} \phi(t, x) b^k(t, x, \mu_t) (\mu_t^N(dx) - \mu_t(x) dx) T^{-1} dt \right)^2 - \frac{1}{36dT^2} \mathbf{V}_{\mathbf{h}}^N \right\}_+ \right] \\ &\lesssim \int_0^\infty \mathbb{P}^N \left(\left| \int_{[0, T] \times \mathbb{R}^d} \phi(t, x) b^k(t, x, \mu_t) (\mu_t^N(dx) - \mu_t(x) dx) T^{-1} dt \right| \geq \left(z + \frac{1}{36dT^2} \mathbf{V}_{\mathbf{h}}^N \right)^{1/2} \right) dz \\ &\lesssim 2\kappa_1 \int_{\mathbf{V}_{\mathbf{h}}^N / (36dT^2)}^\infty \exp \left(- \frac{\kappa_2 N h_1 h_2^d z}{\kappa_3 \kappa_5^2 T^{-1} |H \otimes K|_2^2 + \kappa_5 |H \otimes K|_\infty z^{1/2}} \right) dz \\ &\lesssim (N h_1 h_2^d)^{-1} N^{-(72dT)^{-1} \kappa_2 \kappa_3^{-1} \kappa_5^{-2} \varpi_2} \\ &\quad + (\log N)^{1/2} (N h_1 h_2^d)^{-3/2} \exp \left(- \frac{\kappa_2 \kappa_5^{-1} \varpi_2^{1/2} |H \otimes K|_2}{12d^{1/2} T |H \otimes K|_\infty} (\log N)^{1/2} (N h_1 h_2^d)^{1/2} \right) \\ &\lesssim N^{-2} \end{aligned}$$

as soon as $\varpi_2 \geq 144dT\kappa_2^{-1}\kappa_3\kappa_5^2$, thanks to $\max_{(h_1, h_2) \in \mathcal{H}_2^N} (N h_1 h_2^d)^{-1} \lesssim 1$. We also use the assumption $\min_{(h_1, h_2) \in \mathcal{H}_2^N} N h_1 h_2^d \geq (\log N)^2$ given by (2.15) to show that the second term is negligible in front of N^{-2} . We conclude

$$\mathbb{E}_{\mathbb{P}^N} [I] \lesssim N^{-2}. \quad (2.64)$$

We next consider the term II . For $\tau > 0$, introduce the event

$$\mathcal{B}_\tau^N = \left\{ \int_{[0,T] \times \mathbb{R}^d} |\phi(t, x)| \mu_t^N(dx) dt \leq \tau \right\}.$$

By Cauchy-Schwarz's inequality, on $\mathcal{B}_\tau^N$, we have

$$\begin{aligned} & \left| \int_{[0,T] \times \mathbb{R}^d} \phi(t, x) \xi_t^N(x) \mu_t^N(dx) dt \right|^2 \\ & \leq \int_{[0,T] \times \mathbb{R}^d} |\phi(t, x)| \mu_t^N(dx) dt \int_{[0,T] \times \mathbb{R}^d} |\phi(t, x)| |\xi_t^N(x)|^2 \mu_t^N(dx) dt \\ & \leq \tau \int_{[0,T] \times \mathbb{R}^d} |\phi(t, x)| |\xi_t^N(x)|^2 \mu_t^N(dx) dt. \end{aligned}$$

By Lemma 2.27 and exchangeability, we also have the rough bound

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^N} \left[\left| \int_{[0,T] \times \mathbb{R}^d} \phi(t, x) \xi_t^N(x) \mu_t^N(dx) dt \right|^4 \right] & \leq |\phi|_\infty^4 T^3 \sup_{0 \leq t \leq T} \mathbb{E}_{\mathbb{P}^N} [|\xi_t^N(X_t^N)|^4] \\ & \lesssim (h_1 h_2^d)^{-4} N^{-2}, \end{aligned} \quad (2.65)$$

where we estimate $\mathbb{E}_{\mathbb{P}^N} [|\xi_t^N(X_t^N)|^4]$ as in (2.58) above. It follows that $II \leq IV + V$, with

$$\begin{aligned} IV & = 12\tau \left\{ \int_{[0,T] \times \mathbb{R}^d} |\phi(t, x)| |\xi_t^N(x)|^2 \mu_t^N(dx) dt - \frac{1}{36\tau} \mathbf{V}_{\mathbf{h}}^N \right\}_+ \mathbf{1}_{\mathcal{B}_\tau^N} \\ & \leq 12\tau |K_{h_2}|_\infty \int_0^T |H_{h_1}(t_0 - t)| \left\{ \int_{\mathbb{R}^d} |\xi_t^N(x)|^2 \mu_t^N(dx) - \frac{1}{36\tau |H|_1 |K_{h_2}|_\infty} \mathbf{V}_{\mathbf{h}}^N \right\}_+ dt \mathbf{1}_{\mathcal{B}_\tau^N}, \\ V & = 12 \left| \int_{[0,T] \times \mathbb{R}^d} \phi(t, x) \xi_t^N(x) \mu_t^N(dx) dt \right|^2 \mathbf{1}_{(\mathcal{B}_\tau^N)^c}. \end{aligned}$$

Taking expectation and using exchangeability, we further have

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^N} [IV] & \lesssim |K_{h_2}|_\infty \int_0^T |H_{h_1}(t_0 - t)| \mathbb{E}_{\mathbb{P}^N} \left[\left\{ |\xi_t^N(X_t^N)|^2 - \frac{1}{36\tau |H|_1 |K_{h_2}|_\infty} \mathbf{V}_{\mathbf{h}}^N \right\}_+ \right] dt \\ & \lesssim |K_{h_2}|_\infty \int_0^T |H_{h_1}(t_0 - t)| \int_{\frac{1}{36\tau |H|_1 |K_{h_2}|_\infty} \mathbf{V}_{\mathbf{h}}^N}^\infty \mathbb{P}^N(|\xi_t^N(X_t^N)| \geq z^{1/2}) dz dt \\ & \lesssim |K_{h_2}|_\infty \int_{\frac{1}{36\tau |H|_1 |K_{h_2}|_\infty} \mathbf{V}_{\mathbf{h}}^N}^\infty \exp \left(- \frac{\kappa_7 N z}{1 + N^{1/2} z^{1/2}} \right) dz \\ & \lesssim N^{-1} (h_2^d)^{-1} \exp \left(- \frac{\kappa_7 \varpi_2 |H \otimes K|_2^2}{72\tau |H|_1 |K|_\infty} h_1^{-1} \log N \right) \\ & \quad + N^{-1} (h_2^d)^{-1} h_1^{-1/2} (\log N)^{1/2} \exp \left(- \frac{\kappa_7 \varpi_2^{1/2} |H \otimes K|_2}{12\tau^{1/2} |H|_1^{1/2} |K|_\infty^{1/2}} h_1^{-1/2} (\log N)^{1/2} \right), \end{aligned}$$

by Lemma 2.27, using in particular the fact that $\frac{1}{36\tau |H|_1 |K_{h_2}|_\infty} \mathbf{V}_{\mathbf{h}}^N \gtrsim N^{-1} (\log N)^3 \geq \kappa_8^2 N^{-1}$ for large enough N . Since $\max_{(h_1, h_2) \in \mathcal{H}_2^N} h_1 \leq (\log N)^{-2}$ by assumption (2.15), both terms are negligible in front of N^{-2} and we conclude $\mathbb{E}_{\mathbb{P}^N} [IV] \lesssim N^{-2}$.

We turn to the term V . By Cauchy-Schwarz's inequality and (2.65), we have

$$\begin{aligned}\mathbb{E}_{\mathbb{P}^N}[V] &\leq 12 \mathbb{E}_{\mathbb{P}^N} \left[\left| \int_{[0,T] \times \mathbb{R}^d} \phi(t,x) \xi_t^N(x) \mu_t^N(dx) dt \right|^4 \right]^{1/2} \mathbb{P}^N((\mathcal{B}_\tau^N)^c)^{1/2} \\ &\lesssim N^{-1} (h_1 h_2^d)^{-2} \mathbb{P}^N \left(\int_{[0,T] \times \mathbb{R}^d} |\phi(t,x)| \mu_t^N(dx) dt > \tau \right)^{1/2}.\end{aligned}$$

Note that $\int_{[0,T] \times \mathbb{R}^d} |\phi(t,x)| \mu_t(dx) dt \leq \kappa_3 |\phi|_1$ hence, for the choice $\tau \geq 2\kappa_3 |H \otimes K|_1$, that we make from now on and that does not depend on N , we have

$$\int_{[0,T] \times \mathbb{R}^d} |\phi(t,x)| \mu_t(dx) dt \leq \frac{1}{2} \tau.$$

By triangle inequality and a union bound argument, it follows that

$$\begin{aligned}\mathbb{P}^N \left(\int_{[0,T] \times \mathbb{R}^d} |\phi(t,x)| \mu_t^N(dx) dt > \tau \right)^{1/2} \\ \leq \mathbb{P}^N \left(\left| \int_{[0,T] \times \mathbb{R}^d} |\phi(t,x)| (\mu_t^N(dx) - \mu_t(dx)) T^{-1} dt \right| \geq \frac{1}{2} T^{-1} \tau \right)^{1/2} \\ \leq (2\kappa_1)^{1/2} \exp \left(-\frac{1}{2} \frac{\kappa_2 N h_1 h_2^d \frac{1}{4} T^{-2} \tau^2}{\kappa_3 T^{-1} |H \otimes K|_2^2 + |H \otimes K|_\infty \frac{1}{2} T^{-1} \tau} \right),\end{aligned}$$

where we applied Theorem 2.18. Using $\min_{(h_1, h_2) \in \mathcal{H}_2^N} N h_1 h_2^d \geq (\log N)^2$ granted by (2.15), we obtain $\mathbb{E}_{\mathbb{P}^N}[V] \lesssim N^{-2}$. Putting together our estimates for IV and V , we conclude

$$\mathbb{E}_{\mathbb{P}^N}[II] \lesssim N^{-2}. \quad (2.66)$$

Finally, we consider the term III . The classical following deviation bound holds for continuous martingales :

$$\mathbb{P}^N(\mathcal{M}_T^N(\phi)^k \geq u, \langle \mathcal{M}_T^N(\phi)^k \rangle_T \leq v) \leq \exp \left(-\frac{u^2}{2v} \right)$$

for every $u, v \geq 0$, see *e.g.* [165]. Let $\kappa > 0$ to be tuned below. The choice $v = \kappa (\mathbf{V}_{\mathbf{h}}^N)^{1/2} (\log N)^{-1} u$ entails

$$\mathbb{P}^N(\mathcal{M}_T^N(\phi)^k \geq u) \leq \exp \left(-\frac{1}{2} \kappa^{-1} (\mathbf{V}_{\mathbf{h}}^N)^{-1/2} (\log N) u \right) + \mathbb{P}^N(\langle \mathcal{M}_T^N(\phi)^k \rangle_T \geq \kappa (\mathbf{V}_{\mathbf{h}}^N)^{1/2} (\log N)^{-1} u).$$

It follows that

$$\begin{aligned}\mathbb{E}_{\mathbb{P}^N}[III] &\leq 12 \sum_{k=1}^d \mathbb{E}_{\mathbb{P}^N} \left[\left\{ (\mathcal{M}_T^N(\phi)^k)^2 - \frac{1}{36d} \mathbf{V}_{\mathbf{h}}^N \right\}_+ \right] \\ &\lesssim \sum_{k=1}^d \int_0^\infty \mathbb{P}^N(|\mathcal{M}_T^N(\phi)^k| \geq (z + \frac{1}{36d} \mathbf{V}_{\mathbf{h}}^N)^{1/2}) dz \lesssim VI + VII,\end{aligned}$$

with

$$VI = \int_{\frac{1}{36d} \mathbf{V}_{\mathbf{h}}^N}^\infty \exp \left(-\frac{1}{2} \kappa^{-1} (\mathbf{V}_{\mathbf{h}}^N)^{-1/2} (\log N) z^{1/2} \right) dz,$$

$$VII = \sum_{k=1}^d \int_{\frac{1}{36d} \mathbf{V}_{\mathbf{h}}^N}^{\infty} \mathbb{P}^N(\langle \mathcal{M}^N(\phi)^k \rangle_T \geq \kappa (\mathbf{V}_{\mathbf{h}}^N)^{1/2} (\log N)^{-1} z^{1/2}) dz.$$

Taking for instance $\kappa = (25\sqrt{d})^{-1} < (24\sqrt{d})^{-1}$, we obtain

$$VI \lesssim \mathbf{V}_{\mathbf{h}}^N (\log N)^{-1} N^{-1/(12\kappa d^{1/2})} \lesssim N^{-2}.$$

In order to bound the term VII , we first notice that by Lemma 2.25, we have

$$\begin{aligned} \langle \mathcal{M}^N(\phi)^k \rangle_T &\leq N^{-1} |\text{Tr}(c)|_{\infty} \int_{[0,T] \times \mathbb{R}^d} \phi(t, x)^2 \mu_t^N(dx) dt \\ &\leq N^{-1} |\text{Tr}(c)|_{\infty} \left(\kappa_3 (h_1 h_2^d)^{-1} |H \otimes K|_2^2 + \left| \int_{[0,T] \times \mathbb{R}^d} \phi(t, x)^2 (\mu_t^N(dx) - \mu_t(dx)) dt \right| \right). \end{aligned}$$

Next, the condition

$$N^{-1} |\text{Tr}(c)|_{\infty} \kappa_3 (h_1 h_2^d)^{-1} |H \otimes K|_2^2 \geq \frac{1}{2} \kappa (\mathbf{V}_{\mathbf{h}}^N)^{1/2} (\log N)^{-1} z^{1/2}$$

is equivalent to

$$\begin{aligned} z &\leq 4\kappa^{-2} |\text{Tr}(c)|_{\infty}^2 \kappa_3^2 |H \otimes K|_2^4 N^{-2} (h_1 h_2^d)^{-2} (\mathbf{V}_{\mathbf{h}}^N)^{-1} (\log N)^2 \\ &\leq 4\kappa^{-2} |\text{Tr}(c)|_{\infty}^2 \kappa_3^2 \varpi_2^{-2} \mathbf{V}_{\mathbf{h}}^N \\ &< \frac{1}{36d} \mathbf{V}_{\mathbf{h}}^N \end{aligned}$$

as soon as $\varpi_2 \geq 300d |\text{Tr}(c)|_{\infty} \kappa_3$. It follows that VII is of order

$$\begin{aligned} &\int_{\frac{1}{36d} \mathbf{V}_{\mathbf{h}}^N}^{\infty} \mathbb{P}^N \left(\left| \int_{[0,T] \times \mathbb{R}^d} \phi(t, x)^2 (\mu_t^N(dx) - \mu_t(dx)) dt \right| \geq \frac{1}{2} N |\text{Tr}(c)|_{\infty}^{-1} \kappa (\mathbf{V}_{\mathbf{h}}^N)^{1/2} (\log N)^{-1} z^{1/2} \right) dz \\ &\lesssim \int_{\frac{1}{36d} \mathbf{V}_{\mathbf{h}}^N}^{\infty} \exp \left(- \frac{\kappa_2 \frac{1}{4} N^3 |\text{Tr}(c)|_{\infty}^{-2} \kappa^2 (\mathbf{V}_{\mathbf{h}}^N) (\log N)^{-2} T^{-2} z}{\kappa_3 T^{-1} |\phi|_4^4 + |\phi|_{\infty}^2 \frac{1}{2} N |\text{Tr}(c)|_{\infty}^{-1} \kappa (\mathbf{V}_{\mathbf{h}}^N)^{1/2} (\log N)^{-1} T^{-1} z^{1/2}} \right) du \\ &\lesssim (N h_1 h_2^d)^{-2} (\log N) \exp \left(- \frac{\kappa^2 \kappa_2 |\text{Tr}(c)|_{\infty}^{-2} \varpi_2^2 |H \otimes K|_2^4}{288 \kappa_3 d T |H \otimes K|_4^4} N h_1 h_2^d \right) \\ &\quad + (N h_1 h_2^d)^{-2} (\log N) \exp \left(- \frac{\kappa_2 \kappa |\text{Tr}(c)|_{\infty}^{-1} \varpi_2 |H \otimes K|_2^2}{24 d^{1/2} T |H \otimes K|_{\infty}^2} N h_1 h_2^d \right) \\ &\lesssim N^{-2} \end{aligned}$$

where we applied again Theorem 2.18 and used $\min_{(h_1, h_2) \in \mathcal{H}_2^N} N h_1 h_2^d \geq (\log N)^2$ granted by (2.15). We infer $VII \lesssim N^{-2}$ and conclude

$$\mathbb{E}_{\mathbb{P}^N} [III] \lesssim N^{-2} \quad (2.67)$$

and putting together (2.64), (2.66) and (2.67), we have established

$$\mathbb{E}_{\mathbb{P}^N} \left[\left\{ 4 \left| \widehat{\pi}_{\mathbf{h}}^N(t_0, x_0) - \pi_{\mathbf{h}}(t_0, x_0) \right|^2 - \mathbf{V}_{\mathbf{h}}^N \right\}_+ \right] \lesssim N^{-2}.$$

Step 4 : The control of the second term in the right-hand side of (2.63) is done in the same way as in Step 4 of the proof of Theorem 2.7 and only inflates the previous bound by a factor of order

$\text{Card}(\mathcal{H}_2^N) \lesssim N$. In turn $\mathbb{E}_{\mathbb{P}^N} [A_{\mathbf{h}}^N] \lesssim N^{-1} + \mathcal{B}_{\mathbf{h}}^N(\pi)(t_0, x_0)^2$ and we have established by Step 2 that for any $\mathbf{h} \in \mathcal{H}_2^N$,

$$\mathbb{E}_{\mathbb{P}^N} [|\widehat{\pi}_{\mathbf{h}}^N(t_0, x_0) - \pi(t_0, x_0)|^2] \lesssim \mathcal{B}_{\mathbf{h}}^N(\pi)(t_0, x_0)^2 + V_{\mathbf{h}}^N + N^{-1}$$

holds true. Putting together Step 1 and Theorem 2.7 and using $N^{-1} \lesssim V_{\mathbf{h}}^N$ completes the proof of Theorem 2.9.

7.3 Proof of Theorem 2.14

Proof of the lower bound (2.24)

Step 1

Pick an infinitely many times differentiable function $V_1 : \mathbb{R}^d \rightarrow \mathbb{R}$ such that

- (i) ∇V_1 is Lipschitz continuous,
- (ii) $\limsup_{|x| \rightarrow \infty} -\nabla V_1(x)^\top (x/|x|^2) < 0$,
- (iii) $V_1 = 0$ in a neighbourhood of x_0 .

Let $C_{V_1} = \int_{\mathbb{R}^d} \exp(-2V_1(x))dx$ and define

$$\nu_1(x) = C_{V_1}^{-1} \exp(-2V_1(x)), \quad x \in \mathbb{R}^d.$$

From the classical theory of multidimensional diffusion processes (see *e.g.* the classical textbook of Stroock and Varadhan [?]), properties (i) and (ii) imply that ν_1 is the unique invariant measure of the diffusion process $d\xi_t = -\nabla V_1(\xi_t)dt + dW_t$ for some Brownian motion W on $\mathbb{R}^d$. In turn $\nu_1(t, x) = \nu_1(x)$ as a function defined on $[0, T] \times \mathbb{R}^d$ satisfies

$$\nu_1 = \mathcal{S}(b_1, \text{Id}, \nu_1) \quad \text{with} \quad b_1(t, x, \mu) = -\nabla V_1(x),$$

assuming moreover that V_1 is such that ν_1 satisfies Assumption 2.1, a choice which is obviously possible. Since ν_1 is constant in a neighbourhood of (t_0, x_0) we may (and will) assume that $(b_1, \text{Id}, \nu_1) \in \mathcal{S}_{L/2}^{\alpha, \beta}(t_0, x_0)$. Next, we set

$$V_2^N(x) = V_1(x) + \varpi C_{V_1} N^{-1/2} \tau_N^{d/2} \psi(\tau_N(x - x_0)), \quad 0 < \tau_N \rightarrow \infty,$$

for some $0 < \varpi \leq 1$, where $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$ is infinitely many times differentiable, compactly supported and satisfies

$$\psi(0) = 1, \quad |\psi|_\infty \leq 1, \quad \int_{\mathbb{R}^d} \psi(x)dx = 0, \quad |\psi|_2 = 1,$$

hence ∇V_2^N satisfies (i) and (ii) and (iii). It defines in turn a solution

$$\nu_2^N = \mathcal{S}(b_2^N, \text{Id}, \nu_2^N) \quad \text{with} \quad b_2^N(t, x, \mu) = -\nabla V_2^N(x),$$

having

$$\nu_2^N(x) = C_{V_2^N}^{-1} \exp(-2V_2^N(x)) \quad \text{with} \quad C_{V_2^N} = \int_{\mathbb{R}^d} \exp(-2V_2^N(x))dx$$

and $\nu_2^N(t, x) = \nu_2^N(x)$ is understood as a function defined on $[0, T] \times \mathbb{R}^d$.

Step 2 : We claim that for every $\beta > 0$, setting $\tau_N = N^{1/(2\beta+d)}$ and taking ϖ sufficiently small, we have $(b_2^N, \text{Id}, \nu_2^N) \in \mathcal{S}_L^{\alpha,\beta}(t_0, x_0)$ for large enough N . Indeed, in a neighbourhood of x_0 , we have $V_1 = 0$ hence for x in such a neighbourhood, we have

$$\nu_2^N(x) = C_{V_2^N}^{-1} \exp(-2\varpi C_{V_1} N^{-1/2} \tau_N^{d/2} \psi(\tau_N(x - x_0))).$$

On the one hand, $\varpi N^{-1/2} \tau_N^{d/2} |\psi(\tau_N(\cdot - x_0))|_{\mathcal{H}^\beta(x_0)} \lesssim \varpi |\psi|_{\mathcal{H}^\beta(x_0)}$ which can be taken arbitrarily small. On the other hand, we also have $C_{V_2^N} \rightarrow C_{V_1}$ as $N \rightarrow \infty$, see in particular (2.68) below, and the claim follows.

Step 3

For $(b, c, \mu_0) \in \mathcal{P}$, we write $\mathbb{P}_{b,c,\mu_0}^N$ for $\mathbb{P}^N$ to emphasise the model parameter (b, c, μ_0) . For data extracted from $\mu_{t_0}^N$ solely, we restrict the model to

$$\mathbb{P}_{b,c,\mu_0}^N(t_0) = (X_{t_0}^1, \dots, X_{t_0}^N) \circ \mathbb{P}_{b,c,\mu_0}^N,$$

the law of $(X_{t_0}^1, \dots, X_{t_0}^N)$ under $\mathbb{P}_{b,c,\mu_0}^N$. Note that for a drift $b(t, x, \mu) = b(t, x)$ independent of an interaction measure term μ , we have

$$\mathbb{P}_{b,c,\mu_0}^N = \bar{\mathbb{P}}_{b,c,\mu_0}^N \quad \text{hence} \quad \mathbb{P}_{b,c,\mu_0}^N(t_0) = \mu_{t_0}^{\otimes N}.$$

By Pinsker's inequality, it follows that

$$\begin{aligned} \|\mathbb{P}_{b_1, \text{Id}, \nu_1}^N(t_0) - \mathbb{P}_{b_2^N, \text{Id}, \nu_2^N}^N(t_0)\|_{TV}^2 &= \|\nu_1^{\otimes N} - (\nu_2^N)^{\otimes N}\|_{TV}^2 \\ &\leq \frac{N}{2} \int_{\mathbb{R}^d} \nu_1(x) \log \frac{\nu_1(x)}{\nu_2^N(x)} dx \\ &= N \int_{\mathbb{R}^d} \nu_1(x) (V_2^N(x) - V_1(x)) dx + \frac{N}{2} \log \frac{C_{V_2^N}}{C_{V_1}} \\ &= \frac{N}{2} \log \frac{C_{V_2^N}}{C_{V_1}}, \end{aligned}$$

for large enough N , where $\|\cdot\|_{TV}$ denotes the total variation distance, using successively

$$V_2^N(x) - V_1(x) = \varpi C_{V_1} N^{-1/2} \tau_N^{d/2} \psi(\tau_N(x - x_0)) = \varpi \nu_1(x)^{-1} N^{-1/2} \tau_N^{d/2} \psi(\tau_N(x - x_0)),$$

since $\nu_1(x)^{-1} = C_{V_1}$ in a neighbourhood of x_0 and the fact that $\psi(\tau_N(x - x_0)) = 0$ outside this neighbourhood, for large enough N , thanks to the fact that $\text{Supp}(\psi)$ is compact, together with the cancellation $\int_{\mathbb{R}^d} \psi(x) dx = 0$. Moreover, a Taylor's expansion yields

$$\begin{aligned} \frac{C_{V_2^N}}{C_{V_1}} - 1 &= \int_{\mathbb{R}^d} \nu_1(x) \exp(-2\varpi \nu_1(x)^{-1} N^{-1/2} \tau_N^{d/2} \psi(\tau_N(x - x_0))) dx - 1 \\ &= 2\varpi^2 N^{-1} \tau_N^d \int_{\mathbb{R}^d} \nu_1(x)^{-1} \psi(\tau_N(x - x_0))^2 \vartheta^N(x) dx, \end{aligned}$$

thanks to the cancellation property of ψ again, with a remainder term satisfying

$$0 \leq \vartheta^N(x) \leq \exp(2\varpi N^{-1/2} \tau_N^{d/2} |\psi|_\infty \sup_{x \in \text{Supp}(\psi)} \nu_1(x)^{-1}) \leq 2$$

for large enough N . It follows that

$$\left| \frac{C_{V_2^N}}{C_{V_1}} - 1 \right| \leq 4\varpi^2 N^{-1} \sup_{x \in \text{Supp}(\psi)} \nu_1(x)^{-1} \lesssim N^{-1}. \quad (2.68)$$

The inequality $\log(1+x) \leq x$ for $x \geq -1$ enables us to conclude

$$\|\mathbb{P}_{b_1, c, \nu_1}^N(t_0) - \mathbb{P}_{b_2^N, c, \nu_2^N}^N(t_0)\|_{TV}^2 \leq 2\varpi^2 \sup_{x \in \text{Supp}(\psi)} \nu_1(x)^{-1} \leq \frac{1}{2} \quad (2.69)$$

for large enough N by taking $\varpi > 0$ sufficiently small.

Step 4

We conclude by a classical two-point lower bound argument using Le Cam's lemma : if $\mathbb{P}_i$, $i = 1, 2$ are two probability measures defined on the same probability space and $\Psi(\mathbb{P}_i) \in \mathbb{R}$ is a functional of $\mathbb{P}_i$, we have

$$\inf_{\widehat{\Psi}} \max_{i=1,2} \mathbb{E}_{\mathbb{P}_i} [|\widehat{\Psi} - \Psi(\mathbb{P}_i)|] \geq \frac{1}{2} |\Psi(\mathbb{P}_1) - \Psi(\mathbb{P}_2)| (1 - \|\mathbb{P}_1 - \mathbb{P}_2\|_{TV}), \quad (2.70)$$

where the infimum is taken over all estimators of $\Psi(\mathbb{P}_i)$, see *e.g.* [123] among many other references. We let

$$\Psi(\mathbb{P}_{b_1, \text{Id}, \nu_1}^N(t_0)) = \nu_1(t_0, x_0), \quad \Psi(\mathbb{P}_{b_2^N, \text{Id}, \nu_2^N}^N(t_0)) = \nu_2^N(t_0, x_0),$$

so that

$$\begin{aligned} |\Psi(\mathbb{P}_{b_1, \text{Id}, \nu_1}^N(t_0)) - \Psi(\mathbb{P}_{b_2^N, \text{Id}, \nu_2^N}^N(t_0))| &\gtrsim \nu_1(x_0) \left(\frac{C_{V_2^N}}{C_{V_1}} - \exp(2\varpi \nu_1(x)^{-1} N^{-1/2} \tau_N^{d/2} \psi(0)) \right) \\ &\gtrsim N^{-1/2} \tau_N^{d/2} = N^{-\beta/(2\beta+d)} \end{aligned}$$

in the same way as before, using the properties of ψ and (2.68). We conclude by applying Le Cam's lemma together with (2.69). The proof of the lower bound (2.24) is complete.

Proof of the upper bound (2.23)

The argument is classical (see *e.g.* [87, 88, 89]). Pick

$$\mathcal{H}_1^N = \{e^{-k}, 1 \leq k \leq d^{-1} \log N - 2d^{-1} \log \log N\}.$$

We have $\text{Card } \mathcal{H}_1^N \lesssim N$ holds as well (and is actually much smaller). Moreover, for every $h \in \mathcal{H}_1^N$:

$$\mathcal{B}_h^N(\mu)(t_0, x_0)^2 \lesssim h^{2\beta \wedge \ell} \quad \text{and} \quad V_h^N \lesssim N^{-1} h^{-d} (\log N).$$

Applying Theorem 2.7, we obtain

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^N} [(\widehat{\mu}_{\text{GL}}^N(t_0, x_0) - \mu_{t_0}(x_0))^2] &\lesssim \min_{h \in \mathcal{H}_1^N} (h^{2\beta \wedge \ell} + N^{-1} h^{-d} (\log N)) \\ &\lesssim \left(\frac{\log N}{N} \right)^{2\beta \wedge \ell / (2\beta \wedge \ell + d)}, \end{aligned}$$

for large enough N , since for every $\beta \in (0, \ell]$, we have $e^{-(k^N+1)} \leq (N/\log N)^{-1/(2\beta+d)} \leq e^{-k^N}$ with $k^N = \lfloor \frac{1}{2\beta+d} (\log N - \log \log N) \rfloor$. This proves (2.23) and completes the proof of Theorem 2.14.

7.4 Proof of Theorem 2.15

Proof of the lower bound (2.26)

We apply the same strategy as for Theorem 2.14, establishing a two-point inequality and applying Le Cam's lemma for two drift functions that have no interaction. We write $\mathbb{P}_{b,c,\mu_0}^N$ for the law of $(X_t^1, \dots, X_t^N)_{0 \leq t \leq T}$ parametrised by (b, c, μ_0) . We start with the following simple consequence of Girsanov's theorem :

Lemma 2.28. *For $i = 1, 2$, let $b_i(t, x, \mu) = b_i(t, x)$ be two drift functions that satisfy Assumption 2.4 with no interaction. Set $\Delta(t, x) = b_2(t, x) - b_1(t, x)$. We have*

$$\|\mathbb{P}_{b_2, \text{Id}, \mu_0}^N - \mathbb{P}_{b_1, \text{Id}, \mu_0}^N\|_{TV}^2 \leq \frac{N}{4} \int_0^T |\Delta(t, \cdot)|_{L^2(\mu_t)}^2 dt,$$

where $\mu = \mathcal{S}(b_1, \text{Id}, \mu_0)$ is a solution of (2.3) with parameter (b_1, Id, μ_0) , up to an explicitly computable constant that only depends on μ_0 and b_1 .

Démonstration. With the notation of Section 6.2, by Girsanov's theorem,

$$\frac{d\mathbb{P}_{b_2, \text{Id}, \mu_0}^N}{d\mathbb{P}_{b_1, \text{Id}, \mu_0}^N} = \mathcal{E}_T(M^N(\Delta)), \text{ where } M_t^N(\Delta) = \sum_{i=1}^N \int_0^t \Delta(s, X_s^i)^\top dB_s^i$$

is a $\mathbb{P}_{b_1, \text{Id}, \mu_0}^N$ -martingale, with $B_t^i = \int_0^t (dX_s^i - b_1(s, X_s^i)ds)$. Moreover, since there is no interaction term in the drift b_1 , we have

$$\mathbb{E}_{\mathbb{P}_{b_1, \text{Id}, \mu_0}^N} [\langle M^N(\Delta) \rangle_T] = N \int_0^T \mathbb{E}_{\mathbb{P}_{b_1, \text{Id}, \mu_0}^N} [|\Delta(t, X_t^1)|^2] dt = N \int_0^T |\Delta(t, \cdot)|_{L^2(\mu_t)}^2 dt.$$

By Pinsker's inequality $\|\mathbb{P}_{b_1, \text{Id}, \mu_0}^N - \mathbb{P}_{b_2, \text{Id}, \mu_0}^N\|_{TV}^2 \leq \frac{1}{2} \mathbb{E}_{\mathbb{P}_{b_1, \text{Id}, \mu_0}^N} \left[\log \frac{d\mathbb{P}_{b_1, \text{Id}, \mu_0}^N}{d\mathbb{P}_{b_2, \text{Id}, \mu_0}^N} \right]$. The conclusion follows from

$$\mathbb{E}_{\mathbb{P}_{b_1, \text{Id}, \mu_0}^N} \left[\log \frac{d\mathbb{P}_{b_1, \text{Id}, \mu_0}^N}{d\mathbb{P}_{b_2, \text{Id}, \mu_0}^N} \right] = -\mathbb{E}_{\mathbb{P}_{b_1, \text{Id}, \mu_0}^N} \left[\log \frac{d\mathbb{P}_{b_2, \text{Id}, \mu_0}^N}{d\mathbb{P}_{b_1, \text{Id}, \mu_0}^N} \right] = \frac{1}{2} \mathbb{E}_{\mathbb{P}_{b_1, \text{Id}, \mu_0}^N} [\langle M^N(\Delta) \rangle_T]$$

and the previous estimate on $\langle M^N(\Delta) \rangle_T$ under $\mathbb{P}_{b_1, \text{Id}, \mu_0}^N$. $\square$

Step 1

Pick now a function $b_1 : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ satisfying Assumption 2.4 and any initial condition μ_0 such that $(b_1, \text{Id}, \mu_0) \in \mathcal{S}_{L/2}^{\alpha, \beta}(t_0, x_0) \cap \mathcal{D}_{L/2}^{\alpha, \beta}(t_0, x_0)$.

Let $\psi = (\psi^1, \dots, \psi^d) : (0, T) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ be infinitely many times differentiable, compactly supported and such that for every $1 \leq k \leq d$, we have

$$\psi^k(0) = 1, \quad |\psi^k|_\infty \leq 1, \quad \int_{[0, T] \times \mathbb{R}^d} \psi(t, x) dt dx = 0, \quad |\psi^k|_2 = 1.$$

For $N \geq 1$ and some $0 < \varpi \leq 1$, we define

$$b_2^N(t, x) = b_1(t, x) + \varpi N^{-1/2} \tau_N^{1/2} (\tilde{\tau}_N)^{1/2} \psi(\tau_N(t - t_0), \tilde{\tau}_N(x - x_0)),$$

where τ_N and $\tilde{\tau}_N$ are defined via

$$\tau_N^\alpha = (\tilde{\tau}_N)^\beta = N^{s_d(\alpha, \beta)/(2s_d(\alpha, \beta) + 1)}.$$

By accomodating b_1 and $\varpi > 0$, we may (and will) assume that $(b_2^N, \text{Id}, \mu_0) \in \mathcal{S}_L^{\alpha, \beta}(t_0, x_0) \cap \mathcal{D}_L^{\alpha, \beta}(t_0, x_0)$ for every $N \geq 1$. Setting $\Delta^N(t, x) = b_2^N(t, x) - b_1(t, x)$ and noting that

$$\int_0^T |\Delta^N(t, \cdot)|_{L^2(\mu_t)}^2 dt \lesssim \varpi^2 N^{-1},$$

thanks to the compactness of the support of ψ and Lemma 2.23, we obtain

$$\|\mathbb{P}_{b_1, \text{Id}, \mu_0}^N - \mathbb{P}_{b_2^N, \text{Id}, \mu_0}^N\|_{TV}^2 \lesssim \varpi^2 \leq \frac{1}{2}$$

by Lemma 2.28, for a suitable choice of $\varpi > 0$.

Step 2

We conclude in the same way as in the proof of the lower bound of Theorem 2.14 : by Le Cam's lemma (2.70), with

$$\Psi(\mathbb{P}_{b_1, \text{Id}, \mu_0}^N) = b_1(t_0, x_0) \text{ and } \Psi(\mathbb{P}_{b_2^N, \text{Id}, \mu_0}^N) = b_2^N(t_0, x_0),$$

we have

$$|\Psi(\mathbb{P}_{b_1, \text{Id}, \mu_0}^N) - \Psi(\mathbb{P}_{b_2^N, \text{Id}, \mu_0}^N)| \gtrsim N^{-1/2} \tau_N^{1/2} (\tilde{\tau}_N)^{1/2} = N^{-s_d(\alpha, \beta)/(2s_d(\alpha, \beta) + 1)}$$

and the conclusion follows. The proof of the lower bound (2.26) is complete.

Proof of the upper bound (2.25)

Let

$$\delta^N(k) = \frac{g^{-1}(k/\log N)}{\alpha(g^{-1}(k/\log N))} k,$$

where g^{-1} is the inverse of the function $g(\beta) = \frac{1}{\beta} \frac{s_d(\alpha(\beta), \beta)}{2s_d(\alpha(\beta), \beta) + 1}$, which is non-increasing for $\beta > 0$ thanks to the assumption that $\beta \mapsto \alpha(\beta)$ is non-decreasing and $\beta \mapsto \alpha(\beta)/\beta$ is non-increasing. Pick

$$\begin{aligned} \mathcal{H}_2^N = \Big\{ (e^{-k_1}, e^{-k_2}), k_1 = \delta^N(k_2), 1 \leq k_2 \leq (d+1)^{-1}(\log N - 2 \log \log N), \\ 2 \log \log N \leq \delta^N(k_2) \leq (d+1)^{-1}(\log N - 2 \log \log N) \Big\}. \end{aligned}$$

The condition (2.15) for the grid $\mathcal{H}_2^N$ is satisfied and $\text{Card } \mathcal{H}_2^N \lesssim N$ holds as well (and is actually much smaller). Now, set $\beta_{k_2} = g^{-1}(k_2/\log N)$. We define an ordering $\preceq$ on $\mathcal{H}_2^N$ that has the right behaviour with respect to the bias of π at scale $\mathbf{h}$. We say that $\mathbf{h} = (h_1, h_2) \preceq \mathbf{h}' = (h'_1, h'_2)$ if

$$h_1(k_2)^{\alpha(\beta_{k_2})} + h_2(k_2)^{\beta_{k_2}} \leq h'_1(k'_2)^{\alpha(\beta_{k'_2})} + h'_2(k'_2)^{\beta_{k'_2}}, \quad (2.71)$$

where we write $\mathbf{h} = (h_1, h_2) = (h_1(k_2), h_2(k_2)) = (e^{-\delta^N(k_2)}, e^{-k_2})$ and likewise for $\mathbf{h}'$. We have that $\mathbf{h} \preceq \mathbf{h}'$ is equivalent to $k_2 \geq k'_2$ since

$$h_1(k_2)^{\alpha(\beta_{k_2})} + h_2(k_2)^{\beta_{k_2}} = 2N^{-s_d(\alpha(\beta_{k_2}), \beta_{k_2})/(2s_d(\alpha(\beta_{k_2}), \beta_{k_2})+1)}$$

and the fact that $\beta \mapsto s_d(\alpha(\beta), \beta)$ is non-decreasing, following from the assumption that $\beta \mapsto \alpha(\beta)$ is non-decreasing. Hence $\mathbf{h} \preceq \mathbf{h}'$ or $\mathbf{h}' \preceq \mathbf{h}$. Moreover,

$$(b, c, \mu_0) \in \mathcal{S}_L^{\alpha, \beta}(t_0, x_0) \cap \mathcal{D}_L^{\alpha, \beta}(t_0, x_0) \text{ implies } \pi \in \mathcal{H}^{\alpha, \beta}(t_0, x_0).$$

Therefore, for every $\mathbf{h} = (h_1, h_2) \in \mathcal{H}_2^N$:

$$\mathcal{B}_{\mathbf{h}}^N(\pi)(t_0, x_0)^2 \lesssim h_1^{2\alpha \wedge \ell} + h_2^{2\beta \wedge \ell} \text{ and } \mathbf{V}_{\mathbf{h}}^N \lesssim N^{-1} h_1^{-1} h_2^{-d} (\log N)$$

thanks to the definition of the ordering $\preceq$ in (2.71). It follows that for every $s_d(\alpha, \beta) \in (0, \ell/d]$, we have

$$e^{-(k_1^N+1)} \leq N^{-\alpha^{-1}s_d(\alpha, \beta)/(2s_d(\alpha, \beta)+1)} \leq e^{-k_1^N}$$

with $k_1^N = \lfloor \frac{\alpha^{-1}s_d(\alpha, \beta)}{2s_d(\alpha, \beta)+1} \log N \rfloor$ and

$$e^{-(k_2^N+1)} \leq N^{-\beta^{-1}s_d(\alpha, \beta)/(2s_d(\alpha, \beta)+1)} \leq e^{-k_2^N}$$

with $k_2^N = \lfloor \frac{\beta^{-1}s_d(\alpha, \beta)}{2s_d(\alpha, \beta)+1} \log N \rfloor$. Applying Theorem 2.9, we obtain

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^N} [|\hat{\pi}_{\text{GL}}^N(t_0, x_0) - \pi_{t_0}(x_0)|^2] &\lesssim \min_{(h_1, h_2) \in \mathcal{H}_1^N} (h_1^{2\alpha \wedge \ell} + h_2^{2\beta \wedge \ell} + N^{-1} h_1^{-1} h_2^{-d} (\log N)) + N^{-1} \\ &\lesssim \left(\frac{\log N}{N} \right)^{2s_d(\alpha, \beta) \wedge \ell_d / (2s_d(\alpha, \beta) \wedge \ell_d + 1)}. \end{aligned}$$

This proves (2.23) and completes the proof of Theorem 2.14.

7.5 Proof of Theorem 2.17

Preliminary results

Lemma 2.29. *Work under Assumptions 2.1, 2.2, and 2.4. We have*

$$\sup_{\xi \in \mathbb{R}^d} \mathbb{E}_{\mathbb{P}^N} [|\mathcal{F}(\mathcal{L}(\mu^N - \mu))(\xi)|^2] \lesssim N^{-1}.$$

Démonstration. Writing $\varphi_1(x) = \cos(2\pi\xi^\top x)$ and $\varphi_2(x) = \sin(2\pi\xi^\top x)$, we have by Jensen's inequality and Theorem 2.18

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^N} [|\int_{\mathbb{R}^d} e^{-2i\pi\xi^\top x} \mathcal{L}(\mu^N - \mu)(dx)|^2] &= \mathbb{E}_{\mathbb{P}^N} [|\int_{[0, T]} \int_{\mathbb{R}^d} e^{-2i\pi\xi^\top x} (\mu_t^N - \mu_t)(dx) w(t) \rho(dt)|^2] \\ &\leq \int_{[0, T]} \mathbb{E}_{\mathbb{P}^N} [|\int_{\mathbb{R}^d} e^{-2i\pi\xi^\top x} (\mu_t^N - \mu_t)(dx)|^2] w(t)^2 \rho(dt) \\ &\lesssim \sup_{t \in [0, T]} \max_{k=1, 2} \int_0^\infty \mathbb{P}^N (|\int_{\mathbb{R}^d} \varphi_k d(\mu_t^N - \mu_t)| \geq z^{1/2}) dz \\ &\lesssim 2\kappa_1 \int_0^\infty \exp\left(-\kappa_2 \frac{Nz}{1+z^{1/2}}\right) dz \lesssim N^{-1} \end{aligned}$$

using $|\varphi_k|_{L^2(\mu_t)} \leq 1$ that stems from $|\varphi_k|_\infty \leq 1$. Since this bound is uniform in $\xi \in \mathbb{R}^d$, the result follows. $\square$

Completion of proof of Theorem 2.17

We have

$$|\widehat{F}_{\varpi, \varpi'}^N - F|_2^2 \lesssim |\mathcal{F}(\widehat{F}_{\varpi, \varpi'}^N) - \mathcal{F}(F)|_2^2 = I + II,$$

with

$$\begin{aligned} I &= \left| \left(\frac{\mathcal{F}(\mathcal{L}((\widehat{b}_{h, \mathbf{h}}^N)_{\varpi'}^r)) \overline{\mathcal{F}(\mathcal{L}\mu^N)}}{|\mathcal{F}(\mathcal{L}\mu^N)|^2} - \mathcal{F}(F) \right) \mathbf{1}_{\{|\mathcal{F}(\mathcal{L}\mu^N)|^2 \geq \varpi\}} \right|_2^2, \\ II &= \left| \mathcal{F}(F) \mathbf{1}_{\{|\mathcal{F}(\mathcal{L}\mu^N)|^2 < \varpi\}} \right|_2^2. \end{aligned}$$

On $\{|\mathcal{F}(\mathcal{L}\mu^N)|^2 \geq \varpi\}$ and using the fact that $|\mathcal{F}(\mathcal{L}\mu)(\xi)| > 0$ almost everywhere, we write

$$\begin{aligned} & \frac{\mathcal{F}(\mathcal{L}(\widehat{b}_{h, \mathbf{h}}^N)_{\varpi'}^r) \overline{\mathcal{F}(\mathcal{L}\mu^N)}}{|\mathcal{F}(\mathcal{L}\mu^N)|^2} - \mathcal{F}(F) \\ &= \frac{(\mathcal{F}(\mathcal{L}(\widehat{b}_{h, \mathbf{h}}^N)_{\varpi'}^r) - \mathcal{F}(\mathcal{L}b)) \overline{\mathcal{F}(\mathcal{L}\mu^N)}}{|\mathcal{F}(\mathcal{L}\mu^N)|^2} + \frac{\mathcal{F}(\mathcal{L}b) \overline{\mathcal{F}(\mathcal{L}\mu^N)}}{|\mathcal{F}(\mathcal{L}\mu^N)|^2} - \mathcal{F}(F) \\ &= \frac{(\mathcal{F}(\mathcal{L}(\widehat{b}_{h, \mathbf{h}}^N)_{\varpi'}^r) - \mathcal{F}(\mathcal{L}b)) \overline{\mathcal{F}(\mathcal{L}\mu^N)}}{|\mathcal{F}(\mathcal{L}\mu^N)|^2} + \mathcal{F}(\mathcal{L}b) \left(\frac{\overline{\mathcal{F}(\mathcal{L}\mu^N)}}{|\mathcal{F}(\mathcal{L}\mu^N)|^2} - \frac{\overline{\mathcal{F}(\mathcal{L}\mu)}}{|\mathcal{F}(\mathcal{L}\mu)|^2} \right), \end{aligned}$$

thanks to $\mathcal{F}(\mathcal{L}b) = \mathcal{F}(F) \cdot \mathcal{F}(\mathcal{L}\mu)$. It follows that for $r > 0$, we have

$$\mathbb{E}_{\mathbb{P}^N}[I] \lesssim III + IV,$$

with

$$\begin{aligned} III &= \varpi^{-2} \mathbb{E}_{\mathbb{P}^N} \left[\left| \mathcal{F}(\mathcal{L}((\widehat{b}_{h, \mathbf{h}}^N)_{\varpi'}^r - b)) \right|_2^2 \right], \\ IV &= \varpi^{-2} \mathbb{E}_{\mathbb{P}^N} \left[\left| \mathcal{F}(F) \overline{\mathcal{F}(\mathcal{L}\mu^N)} \mathcal{F}(\mathcal{L}(\mu^N - \mu)) \right|_2^2 \right]. \end{aligned}$$

By Parseval's identity, and the boundedness of $\mathcal{L}$, we have

$$III \lesssim \varpi^{-2} \sup_{t \in \text{Supp}(w), |x| \leq r} \mathbb{E}_{\mathbb{P}^N} \left[\left| \widehat{b}_{h, \mathbf{h}}^N(t, x)_{\varpi'} - b(x, \mu_t) \right|^2 \right] + \varpi^{-2} \sup_{t \in \text{Supp}(w)} \int_{|x| \geq r} |b(x, \mu_t)|^2 dx.$$

By Lemma 2.24 and Lemma 2.26, we have

$$\sup_{t \in \text{Supp}(w), |x| \leq r} \mathbb{E}_{\mathbb{P}^N} \left[\left| \widehat{b}_{h, \mathbf{h}}^N(t, x)_{\varpi'} - b(x, \mu_t) \right|^2 \right] \lesssim \varpi'(r)^{-2} (\mathcal{B}_h^N(\mu)(t, x) + \mathcal{B}_{\mathbf{h}}^N(\pi)(t, x) + \mathbf{V}_h^N + \mathbf{V}_{\mathbf{h}}^N).$$

Moreover, by Proposition 2.13 the smoothness of F , G and μ_0 entails some Hölder smoothness on μ and b (hence on π) that implies in turn the estimate

$$\sup_{t \in \text{Supp}(w), |x| \leq r} \mathcal{B}_h^N(\mu)(t, x) + \mathcal{B}_{\mathbf{h}}^N(\pi)(t, x) \lesssim C(r)(h^\gamma + h_1^{\gamma'} + h_2^{\gamma''}),$$

for some $\gamma, \gamma', \gamma'' > 0$ and $C(r)$ a locally bounded function in r . We infer

$$\varpi^{-2} \sup_{t \in \text{Supp}(w), |x| \leq r} \mathbb{E}_{\mathbb{P}^N} \left[\left| \widehat{b}_{h, \mathbf{h}}^N(t, x)_{\varpi'} - b(x, \mu_t) \right|^2 \right] \lesssim \varpi^{-2} \varpi'(r)^{-2} C(r) u_N,$$

with $u_N \rightarrow 0$, for a choice $(h, \mathbf{h}) = (h_N, \mathbf{h}_N) \rightarrow 0$ as $N \rightarrow \infty$. Since $b(x, \mu_t) = G(x) + F \star \mu_t(x)$, the second term in III can be bounded as follows :

$$\int_{|x| \geq r} |b(x, \mu_t)|^2 dx \lesssim \int_{|x| \geq r} |G(x)|^2 dx + \int_{\mathbb{R}^d} \left(\int_{|x| \geq r} |F(x-y)|^2 dx \right) \mu_t(dy)$$

by Fubini's theorem, and both terms converge to 0 at some rate $\tilde{C}(r) \rightarrow 0$ as $r \rightarrow \infty$ by dominated convergence since F and G are in $L^2(\mathbb{R}^d)$. We conclude

$$III \lesssim \varpi^{-2} (\varpi'(r)^{-2} C(r) u_N + \tilde{C}(r)).$$

By Parseval's identity, the boundedness of $\overline{\mathcal{F}(\mathcal{L}\mu^N)}$ and Lemma 2.29, we also have

$$IV \lesssim \varpi^{-2} \sup_{\xi \in \mathbb{R}^d} \mathbb{E}_{\mathbb{P}^N} [|\mathcal{F}(\mathcal{L}(\mu^N - \mu))(\xi)|^2] |F|_2^2 \lesssim \varpi^{-2} N^{-1}.$$

We finally turn to the term II . We have

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^N} [II] &\lesssim \mathbb{E}_{\mathbb{P}^N} [|\mathcal{F}(F) \mathbf{1}_{\{|\mathcal{F}(\mathcal{L}(\mu^N - \mu))|^2 \geq \varpi\}}|_2^2] + |\mathcal{F}(F) \mathbf{1}_{\{|\mathcal{F}(\mathcal{L}\mu)|^2 \leq 2\varpi\}}|_2^2 \\ &\lesssim \varpi^{-2} \sup_{\xi \in \mathbb{R}^d} \mathbb{E}_{\mathbb{P}^N} [|\mathcal{F}(\mathcal{L}(\mu^N - \mu))(\xi)|^2] |F|_2^2 + H(\varpi) \\ &\lesssim \varpi^{-2} N^{-1} + H(\varpi), \end{aligned}$$

where $H(\varpi) \rightarrow 0$ as $\varpi \rightarrow 0$ by dominated convergence thanks to the property $|\mathcal{F}(\mathcal{L}\mu)(\xi)| > 0$ almost everywhere of Assumption 2.16. We conclude

$$\mathbb{E}_{\mathbb{P}^N} [|\hat{F}_{\varpi, \varpi'}^N - F|_2^2] \lesssim \varpi^{-2} (\varpi'(r)^{-2} C(r) u_N + \tilde{C}(r)) + N^{-1} + H(\varpi).$$

Let $r_N \rightarrow \infty$ slowly enough so that $\varpi'(r_N)^{-2} C(r_N) u_N \rightarrow 0$. This yields $v_N = C(r_N) u_N + \tilde{C}(r_N) + N^{-1} \rightarrow 0$. Pick now $\varpi_N \rightarrow 0$ slowly enough so that $\varpi_N^{-2} v_N \rightarrow 0$. The proof of Theorem 2.17 follows.

8 Appendix

Characterisation of sub-Gaussian random variables

We recall a classical definition of a sub-Gaussian random variable. Recommended reference is [31].

Definition 2.30. *A real-valued random variable Z such that $\mathbb{E}[Z] = 0$ is λ^2 sub-Gaussian if one of the following conditions is satisfied, each statement implying the next :*

(i) *Laplace transform condition*

$$\mathbb{E}[\exp(zZ)] \leq \exp\left(\frac{1}{2}\lambda^2 z^2\right) \text{ for every } z \in \mathbb{R}.$$

(ii) *Moment condition*

$$\mathbb{E}[Z^{2p}] \leq p!(4\lambda^2)^p \text{ for every integer } p \geq 1.$$

(iii) *Orlicz condition*

$$\mathbb{E}\left[\exp\left(\frac{1}{8\lambda^2} Z^2\right)\right] \leq 2.$$

(iv) *Laplace transform condition (bis)*

$$\mathbb{E}[\exp(zZ)] \leq \exp\left(\frac{24}{2}\lambda^2 z^2\right) \text{ for every } z \in \mathbb{R}.$$

We will also use the following additive property of sub-Gaussian random variables : if the random variables Z_i are independent and λ_i^2 sub-Gaussian, then $\rho(Z_1 + Z_2)$ is $|\rho|^2(\lambda_1^2 + \lambda_2^2)$ sub-Gaussian for every $\rho \in \mathbb{R}$.

8.1 Proof of Lemma 2.20

By Assumption 2.4, the estimate

$$|b(t, x, \mu_t)| \leq b_0 + |b|_{\text{Lip}}(|x| + \mathbb{E}_{\mathbb{P}^N}[|X_t^i|])$$

holds for every $(t, x) \in [0, T] \times \mathbb{R}^d$, where $b_0 = \sup_{t \in [0, T]} |b(t, 0, \delta_0)|$. Remember that

$$\bar{B}_t^i = \int_0^t c(s, X_s^i)^{-1/2} (dX_s^i - b(s, X_s^i, \mu_s) ds), \quad 1 \leq i \leq N,$$

are independent d -dimensional $\bar{\mathbb{P}}^N$ -Brownian motions. By Minkowski's and Jensen's inequality, we have

$$\begin{aligned} |X_t^i| &\leq |X_0^i| + \int_0^t |b(t, X_s^i, \mu_s)| ds + \left| \int_0^t \sigma(s, X_s^i) d\bar{B}_s^i \right| \\ &\leq |X_0^i| + b_0 t + |b|_{\text{Lip}} \int_0^t (|X_s^i| + \mathbb{E}_{\bar{\mathbb{P}}^N}[|X_s^i|]) ds + \left| \int_0^t \sigma(s, X_s^i) d\bar{B}_s^i \right| \\ &\leq |X_0^i| + b_0 T + |b|_{\text{Lip}} \int_0^t (|X_s^i| + \mathbb{E}_{\bar{\mathbb{P}}^N}[|X_s^i|]) ds + \zeta_T^i, \end{aligned} \quad (2.72)$$

where $\zeta_T^i = \sup_{0 \leq t \leq T} \left| \int_0^t \sigma(s, X_s^i) d\bar{B}_s^i \right|$. Integrating w.r.t. $\bar{\mathbb{P}}^N$, we also have

$$\mathbb{E}_{\bar{\mathbb{P}}^N}[|X_t^i|] \leq \mathbb{E}_{\bar{\mathbb{P}}^N}[|X_0^i|] + b_0 T + 2|b|_{\text{Lip}} \int_0^t \mathbb{E}_{\bar{\mathbb{P}}^N}[|X_s^i|] ds + \mathbb{E}_{\bar{\mathbb{P}}^N}[\zeta_T^i].$$

We infer by Grönwall's lemma

$$\mathbb{E}_{\bar{\mathbb{P}}^N}[|X_t^i|] \leq (\mathbb{E}_{\bar{\mathbb{P}}^N}[|X_0^i|] + b_0 T + \mathbb{E}_{\bar{\mathbb{P}}^N}[\zeta_T^i]) e^{2|b|_{\text{Lip}} t}$$

and plugging this estimate in (2.72) we infer

$$|X_t^i| \leq |X_0^i| + b_0 T + |b|_{\text{Lip}} \int_0^t |X_s^i| ds + (\mathbb{E}_{\bar{\mathbb{P}}^N}[|X_0^i|] + b_0 T + \mathbb{E}_{\bar{\mathbb{P}}^N}[\zeta_T^i]) e^{2|b|_{\text{Lip}} T} + \zeta_T^i.$$

Applying Grönwall's lemma again, we derive

$$\begin{aligned} |X_t^i| &\leq (|X_0^i| + b_0 T + (\mathbb{E}_{\bar{\mathbb{P}}^N}[|X_0^i|] + b_0 T + \mathbb{E}_{\bar{\mathbb{P}}^N}[\zeta_T^i]) e^{2|b|_{\text{Lip}} T} + \zeta_T^i) e^{|b|_{\text{Lip}} t} \\ &\leq (|X_0^i| + \mathbb{E}_{\bar{\mathbb{P}}^N}[|X_0^i|] + 2b_0 T + \zeta_T^i + \mathbb{E}_{\bar{\mathbb{P}}^N}[\zeta_T^i]) e^{3|b|_{\text{Lip}} t}. \end{aligned}$$

Taking the exponent $2p$ and expectation w.r.t. $\bar{\mathbb{P}}^N$, we further obtain

$$\begin{aligned}\mathbb{E}_{\bar{\mathbb{P}}^N} [|X_t^i|^{2p}] &\leq 5^{2p-1} (2\mathbb{E}_{\bar{\mathbb{P}}^N} [|X_0^i|^{2p}] + (2b_0T)^p + 2\mathbb{E}_{\bar{\mathbb{P}}^N} [(\zeta_T^i)^{2p}]) e^{3|b|_{\text{Lip}}T} \\ &\leq C_6^p (\mathbb{E}_{\bar{\mathbb{P}}^N} [|X_0^i|^{2p}] + (b_0T)^p + \mathbb{E}_{\bar{\mathbb{P}}^N} [(\zeta_T^i)^{2p}])\end{aligned}$$

with $C_6 = 50 e^{3|b|_{\text{Lip}}T}$. By Assumption 2.1, the initial condition $|X_0^i|$ satisfies

$$\mathbb{E}_{\bar{\mathbb{P}}^N} [\exp(\gamma_0 |X_0^i|^2)] = 1 + \sum_{p \geq 1} \frac{\gamma_0^p}{p!} \mathbb{E}_{\bar{\mathbb{P}}^N} [|X_0^i|^{2p}] \leq \gamma_1$$

hence for every $p \geq 1$, we obtain

$$\mathbb{E}_{\bar{\mathbb{P}}^N} [|X_0^i|^{2p}] \leq p! \left(\frac{\gamma_1}{\gamma_0} \right)^p$$

since $\gamma_1 \geq 1$. By Burkholder-Davis-Gundy's inequality with constant $(C^*)^{p/2} p^{p/2}$ for some numerical constant C^* , see *e.g.* Barlow and Yor [7], we also have

$$\begin{aligned}\mathbb{E}_{\bar{\mathbb{P}}^N} [(\zeta_T^i)^{2p}] &\leq \left(\frac{2p}{2p-1} \right)^{2p} \mathbb{E}_{\bar{\mathbb{P}}^N} \left[\left| \int_0^T \sigma(t, X_t^i) d\bar{B}_t^i \right|^{2p} \right] \\ &\leq \left(\frac{2p}{2p-1} \right)^{2p} (2C^*)^p p^p \mathbb{E}_{\bar{\mathbb{P}}^N} \left[\left(\int_0^T \text{Tr}(c(t, X_t^i)) dt \right)^p \right] \\ &\leq p^p (8C^*T |\text{Tr}(c)|_\infty)^p \leq p! (8C^*eT |\text{Tr}(c)|_\infty)^p.\end{aligned}$$

Putting these estimates together, we conclude

$$\mathbb{E}_{\bar{\mathbb{P}}^N} [|X_t^i|^{2p}] \leq p! C_6^p \left(\frac{\gamma_1}{\gamma_0} + T(b_0 + 8C^*e |\text{Tr}(c)|_\infty) \right)^p$$

and Lemma 2.20 is established with $C_2 = C_6 \left(\frac{\gamma_1}{\gamma_0} + T(b_0 + 8C^*e |\text{Tr}(c)|_\infty) \right)$.

8.2 Proof of Lemma 2.22

Fix $\mathcal{J}_k = \{i_1, \dots, i_k\} \subset \{1, \dots, N\}$. For $g : [0, T] \times (\mathbb{R}^d)^k \times (\mathbb{R}^d)^\ell \rightarrow \mathbb{R}^d$, we define

$$g_{\mathcal{J}_k}(t, y^\ell) = g(t, X_t^{i_1}, \dots, X_t^{i_k}, y^\ell).$$

For technical convenience, we establish a slightly stronger, replacing $\mathcal{V}_{2p}^N(f(t, \cdot))$ in (2.46) by

$$\mathcal{V}_{2p, \ell}^N(g_{\mathcal{J}_{k-\ell+1}}(t, \cdot)) = \mathbb{E}_{\bar{\mathbb{P}}^N} \left[\left| \int_{(\mathbb{R}^d)^\ell} g(t, X_t^{i_1}, X_t^{i_2}, \dots, X_t^{i_{k-\ell+1}}, y^\ell) (\mu_t^N - \mu_t)^{\otimes \ell} (dy^\ell) \right|^{2p} \right]$$

for every $\mathcal{J}_{k-\ell+1} \subset \{1, \dots, N\}$ with cardinality $k - \ell + 1$ and every function $g : [0, T] \times (\mathbb{R}^d)^{k-\ell+1} \times (\mathbb{R}^d)^\ell \rightarrow \mathbb{R}^d$, Lipschitz continuous in the space variables, that defines in turn a class $\mathcal{G}_{k-\ell+1, \ell}$. In particular $\mathcal{V}_{2p, \ell}^N(f(t, \cdot))$ and $\mathcal{V}_{2p, \ell}^N(g_{\mathcal{J}_{k-\ell+1}}(t, \cdot))$ agree for $\ell = k$ in which case the class $\mathcal{G}_{1, k}$ coincide with $\mathcal{G}_k$ and we obtain Lemma 2.22. We prove the result by induction.

Step 1 :

The case $\ell = 1$. For $g \in \mathcal{G}_{k,1}$, $x^k \in (\mathbb{R}^d)^k$ and $\mathcal{J} \subset \{1, \dots, N\}$, let

$$\Lambda_t^{\mathcal{J}}(g, x^k) = \int_{\mathbb{R}^d} g(t, x^k, y) (\mu_t^{\mathcal{J}} - \mu_t)(dy),$$

where we write $\mu_t^{\mathcal{J}}(dx) = |\mathcal{J}|^{-1} \sum_{i \in \mathcal{J}} \delta_{X_t^i}(dx)$ for the empirical measure in restriction to $\mathcal{J}$. Observe that $\Lambda_t^{\mathcal{J}}(g, x^k)$ is a sum of independent and centred random variables under $\bar{\mathbb{P}}^N$. We write

$$\begin{aligned} \Lambda_t^{\{1, \dots, N\}}(g, X_t^{i_1}, \dots, X_t^{i_k}) &= N^{-1} \sum_{i \in \mathcal{J}_k} \left(g(t, X_t^{i_1}, \dots, X_t^{i_k}, X_t^i) - \int_{\mathbb{R}^d} g(t, X_t^{i_1}, \dots, X_t^{i_k}, y) \mu_t(dy) \right) \\ &\quad + \frac{N-k}{N} \Lambda_t^{\mathcal{J}_k^c}(g, X_t^{i_1}, \dots, X_t^{i_k}), \end{aligned}$$

since $|\mathcal{J}_k| = k$. We obtain the decomposition

$$\mathcal{V}_{2p,1}^N(g_{\mathcal{J}_k}(t, \cdot)) = \mathbb{E}_{\bar{\mathbb{P}}^N} \left[\left| \Lambda_t^{\{1, \dots, N\}}(g, X_t^{i_1}, \dots, X_t^{i_k}) \right|^{2p} \right] \leq 2^{2p-1} (I + II),$$

with

$$\begin{aligned} I &= \frac{k^{2p-1}}{N^{2p}} \sum_{i \in \mathcal{J}_k} \left(\mathbb{E}_{\bar{\mathbb{P}}^N} \left[\left| g(t, X_t^{i_1}, \dots, X_t^{i_k}, X_t^i) - \int_{\mathbb{R}^d} g(t, X_t^{i_1}, \dots, X_t^{i_k}, y) \mu_t(dy) \right|^{2p} \right] \right), \\ II &= \left(\frac{N-k}{N} \right)^{2p} \mathbb{E}_{\bar{\mathbb{P}}^N} \left[\left| \Lambda_t^{\mathcal{J}_k^c}(g, X_t^{i_1}, \dots, X_t^{i_k}) \right|^{2p} \right]. \end{aligned}$$

The term I is controlled by the smoothness of g :

$$I \leq \frac{k^{2p-1}}{N^{2p}} |g(t, \cdot)|_{\text{Lip}}^{2p} \sum_{i \in \mathcal{J}_k} \mathbb{E}_{\bar{\mathbb{P}}^N} \left[\int_{\mathbb{R}^d} |X_t^i - y|^{2p} \mu_t(dy) \right] \leq N^{-2p} p! (k^2 4C_2)^p |g(t, \cdot)|_{\text{Lip}}^{2p},$$

where the last estimate stems from Lemma 2.20. For the term II , writing $g = (g^1, \dots, g^d)$ where the functions g^j are real-valued, we further have

$$II \leq \left(\frac{N-k}{N} \right)^{2p} d^{2p-1} \sum_{j=1}^d \mathbb{E}_{\bar{\mathbb{P}}^N} \left[\Lambda_t^{\mathcal{J}_k^c}(g^j, X_t^{i_1}, \dots, X_t^{i_k})^{2p} \right]. \quad (2.73)$$

Moreover, for every $x \in \mathbb{R}^d$, the term

$$\Lambda_t^{\mathcal{J}_k^c}(g^j, x^k) = \frac{1}{N-k} \sum_{i \in \mathcal{J}_k^c} (g^j(t, x^k, X_t^i) - \mathbb{E}_{\bar{\mathbb{P}}^N} [g^j(t, x^k, X_t^i)])$$

is the sum of independent centred random variables that are independent of $(X_t^{i_1}, \dots, X_t^{i_k})$ and

$$g^j(t, x^k, X_t^i) - \mathbb{E}_{\bar{\mathbb{P}}^N} [g^j(t, x^k, X_t^i)]$$

is λ^2 sub-Gaussian with $\lambda^2 = 24C_2 |g^j(t, \cdot)|_{\text{Lip}}^2$ via the same estimate as for I and the fact that (ii) implies (iv) in Definition 2.30. Thanks to the additivity property of independent sub-Gaussian random variables, we further infer that $\Lambda_t^{\mathcal{J}_k^c}(g^j, x^k)$ is $\tilde{\lambda}^2$ sub-Gaussian with

$$\tilde{\lambda}^2 = \frac{1}{N-k} \lambda^2 = \frac{1}{N-k} 24C_2 |g^j(t, \cdot)|_{\text{Lip}}^2.$$

Conditioning on $(X_t^{i_1}, \dots, X_t^{i_k})$, we derive

$$\mathbb{E}_{\mathbb{P}^N} [\Lambda_t^{\mathcal{J}_k^c}(g^j, X_t^{i_1}, \dots, X_t^{i_k})^{2p}] \leq \frac{p!(96C_2)^p}{(N-k)^p} |g^j(t, \cdot)|_{\text{Lip}}^{2p}$$

by (ii) of Definition 2.30. Plugging this estimate in (2.73), we obtain

$$II \leq \frac{p!(96C_2d^2)^p}{(N-k)^p} |g(t, \cdot)|_{\text{Lip}}^{2p}$$

and putting together our estimates for I and II , we conclude

$$\mathcal{V}_{2p,1}^N(g(t, \cdot)) \leq \frac{p!K_1^p}{(N-k)^p} |g(t, \cdot)|_{\text{Lip}}^{2p}$$

with $K_1 = 16(k^2 + 24d^2)C_2$. This establishes Lemma 2.22 for g in the case $\ell = 1$.

Step 2 :

We assume that (2.46) holds for $\mathcal{V}_{2p,\ell}^N(g_{\mathcal{J}_{k-\ell+1}}(t, \cdot))$, for every $\mathcal{J}_{k-\ell+1} \subset \{1, \dots, N\}$ with cardinality $k - \ell + 1$ and every $g \in \mathcal{G}_{k-\ell+1,\ell}$ with $\ell < k$. Let $g \in \mathcal{G}_{k-\ell,\ell+1}$ and $\mathcal{J}_{k-\ell} \subset \{1, \dots, N\}$. We have :

$$\begin{aligned} \mathcal{V}_{2p,\ell+1}^N(g_{\mathcal{J}_{k-\ell}}(t, \cdot)) &= \mathbb{E}_{\mathbb{P}^N} \left[\left| \int_{(\mathbb{R}^d)^{\ell+1}} g_{\mathcal{J}_{k-\ell}}(t, y^{\ell+1}) (\mu_t^N - \mu_t)^{\otimes(\ell+1)}(dy^{\ell+1}) \right|^{2p} \right] \\ &\leq 2^{2p-1}(III + IV), \end{aligned}$$

with

$$\begin{aligned} III &= N^{-1} \sum_{i=1}^N \mathbb{E}_{\mathbb{P}^N} \left[\left| \int_{(\mathbb{R}^d)^\ell} g(t, X_t^{i_1}, \dots, X_t^{i_{k-\ell}}, (X_t^i, y^\ell)) (\mu_t^N - \mu_t)^{\otimes \ell}(dy^\ell) \right|^{2p} \right], \\ IV &= \int_{\mathbb{R}^d} \mathbb{E}_{\mathbb{P}^N} \left[\left| \int_{(\mathbb{R}^d)^\ell} g(t, X_t^{i_1}, \dots, X_t^{i_{k-\ell}}, (y, y^\ell)) (\mu_t^N - \mu_t)^{\otimes \ell}(dy^\ell) \right|^{2p} \right] \mu_t(dy). \end{aligned}$$

Let $i_0 \in \mathcal{J}_{k-\ell}^c$ and put $\mathcal{J}_{k-\ell+1} = \mathcal{J}_{k-\ell} \cup \{i_0\}$. The term IV can be rewritten as

$$IV = \int_{\mathbb{R}^d} \mathcal{V}_{2p,\ell}^N(g'_{\mathcal{J}_{k-\ell+1}}(t, \cdot)(y)) \mu_t(dy),$$

where, for fixed $y \in \mathbb{R}^d$, the function $g'(t, x_{i_1}, \dots, x_{i_{k-\ell}}, x_{i_0}, y^\ell)(y) = g(t, x_{i_1}, \dots, x_{i_{k-\ell}}, (y, y^\ell))$ with the artificial variable x_{i_0} belongs to $\mathcal{G}_{k-\ell+1,\ell}$. By the induction hypothesis and noting that $\sup_{y \in \mathbb{R}^d} |g'(t, \cdot, y)|_{\text{Lip}} \leq |g(t, \cdot)|_{\text{Lip}}$, we infer

$$IV \leq \frac{p!K_\ell^p}{(N-k)^p} |g(t, \cdot)|_{\text{Lip}}^{2p}.$$

We split the sum in III over indices in $\mathcal{J}_{k-\ell}$ and $\mathcal{J}_{k-\ell}^c$. If $i \in \mathcal{J}_{k-\ell}$, in the same way as for IV , we write

$$g(t, X_t^{i_1}, \dots, X_t^{i_{k-\ell}}, (X_t^i, y^\ell)) = g''_{\mathcal{J}_{k-\ell+1}}(t, y^\ell)$$

with $\mathcal{J}_{k-\ell+1} = \mathcal{J}_{k-\ell} \cup \{i_0\}$ for some arbitrary $i_0 \in \mathcal{J}_{k-\ell}^c$ and with $g''(t, x_{i_1}, \dots, x_{i_{k-\ell}}, x_{i_0}, y^\ell) = g(t, x_{i_1}, \dots, x_{i_{k-\ell}}, (x_i, y^\ell))$, where i coincides with one of the $i_j \in \mathcal{J}_{k-\ell}$. Also, g'' belongs to $\mathcal{G}_{k-\ell+1, \ell}$. If $i \in \mathcal{J}_{k-\ell}^c$, we write

$$g(t, X_t^{i_1}, \dots, X_t^{i_{k-\ell}}, (X_t^i, y^\ell)) = g'''_{\mathcal{J}_{k-\ell} \cup \{i\}}(t, y^\ell)$$

with $g'''(t, x_{i_1}, \dots, x_{i_{k-\ell}}, x_i, y^\ell) = g(t, x_{i_1}, \dots, x_{i_{k-\ell}}, (x_i, y^\ell))$ and g''' belongs to $\mathcal{G}_{k-\ell+1, \ell}$ as well. We infer

$$III \leq (k-\ell)N^{-1}\mathcal{V}_{2p, \ell}^N(g''_{\mathcal{J}_{k-\ell+1}}(t, \cdot)) + N^{-1} \sum_{i \in \mathcal{J}_{k-\ell}^c} \mathcal{V}_{2p, \ell}^N(g'''_{\mathcal{J}_{k-\ell} \cup \{i\}}(t, \cdot)) \leq \frac{p!K_\ell^p}{(N-k)^p} |g(t, \cdot)|_{\text{Lip}}^{2p}$$

by the induction hypothesis and noting again that $|g''(t, \cdot)|_{\text{Lip}}$ and $|g'''(t, \cdot)|_{\text{Lip}}$ are controlled by $|g(t, \cdot)|_{\text{Lip}}$. We conclude

$$\mathcal{V}_{2p, \ell+1}^N(g_{\mathcal{J}_{k-\ell}}(t, \cdot)) \leq 2^{2p} \frac{p!K_\ell^p}{(N-k)^p} |g(t, \cdot)|_{\text{Lip}}^{2p} = \frac{p!K_{\ell+1}^p}{(N-k)^p} |g(t, \cdot)|_{\text{Lip}}^{2p}$$

with $K_{\ell+1} = 4K_\ell$. The proof of Lemma 2.22 is complete.

8.3 (Sketch of) proof of Proposition 2.13

Step 1 : Thanks to Chapters 6 and 9 of [17], since F, G are bounded and μ_0 satisfies Assumption 2.1, it can be shown that (2.3) admits a unique probability solution μ in the sense of [17], absolutely continuous w.r.t. the Lebesgue measure, that we still denote $\mu(t, x) = \mu_t(x)$. Moreover $\mu \in \mathcal{H}_{\text{loc}}^{\delta/2, \delta} = \cap_{(t_0, x_0) \in (0, T) \times \mathbb{R}^d} \mathcal{H}^{\delta/2, \delta}(t_0, x_0)$ for every $0 < \delta < \frac{1}{2}$. The main arguments of these properties rely on the existence of a suitable Lyapunov function associated to (2.3), following the terminology of [143] and [17] (for instance $x \mapsto 1 + |x|^2$) together with Sobolev embeddings.

Step 2 : Define

$$\tilde{a}_k(t, x) = G^k(x) + \int_{\mathbb{R}^d} F^k(x-y) \mu_t(y) dy, \quad k = 1, \dots, d,$$

and

$$\tilde{a}(t, x) = \text{div}(G(x) + \int_{\mathbb{R}^d} F(x-y) \mu_t(y) dy),$$

which are well defined since $\beta, \beta' > 1$. Consider next the Cauchy problem associated to (2.3) in its strong form :

$$\begin{cases} \partial_t \tilde{\mu}_t = \frac{1}{2} \sigma^2 \Delta \tilde{\mu}_t - \sum_{k=1}^d \tilde{a}_k(t, \cdot) \partial_k \tilde{\mu}_t - \tilde{a}(t, \cdot) \tilde{\mu}_t \\ \tilde{\mu}_{t=0} = \mu_0. \end{cases} \quad (2.74)$$

Taking $\delta = \beta - \lfloor \beta \rfloor$ we obtain $\tilde{a}_i, \tilde{a} \in \mathcal{C}_{\text{loc}}^{(\beta - \lfloor \beta \rfloor)/2, \beta - \lfloor \beta \rfloor}$ by Step 1.

Step 3 : Using $\inf \tilde{a} > -\infty$ and the existence of a Lyapunov function associated to the problem, by Theorem 2.3 of [4], there exists a unique solution $\tilde{\mu}$ of (2.74). Moreover, $\tilde{\mu}$ is continuous on $(0, T) \times \mathbb{R}^d$ and

$$\tilde{\mu} \in \mathcal{C}_{\text{loc}}^{1+(\beta - \lfloor \beta \rfloor)/2, 2+\beta - \lfloor \beta \rfloor}.$$

It is also the unique solution defined in Theorem 12 of Chapter 1 of [75], therefore the unique integrable solution of the problem (2.3). By uniqueness, $\mu = \tilde{\mu}$.

Step 4 : If $\lfloor \beta \rfloor = 1$, we obtain $\mu \in \mathcal{H}^{(1+\beta)/2, 1+\beta}(t_0, x_0)$ for every $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$. Otherwise, we can iterate the process thanks to results of Section 8.12 in [113] : successively :

- Since $\partial_{x_k} \tilde{a}_k$ and $\partial_{x_k} \tilde{a}$ are in $\mathcal{C}_{\text{loc}}^{(\beta - \lfloor \beta \rfloor)/2, \beta - \lfloor \beta \rfloor}$, we have

$$\partial_{x_k} \mu \in \mathcal{C}_{\text{loc}}^{1 + (\beta - \lfloor \beta \rfloor)/2, 2 + \beta - \lfloor \beta \rfloor}.$$

- Since $\partial_t \tilde{a}_k$ and $\partial_t \tilde{a}$ are now in $\mathcal{C}_{\text{loc}}^{(\beta - \lfloor \beta \rfloor)/2, \beta - \lfloor \beta \rfloor}$, we have

$$\partial_t \mu \in \mathcal{C}_{\text{loc}}^{1 + (\beta - \lfloor \beta \rfloor)/2, 2 + \beta - \lfloor \beta \rfloor}.$$

Therefore, if $\lfloor \beta \rfloor = 2$, we obtain $\mu \in \mathcal{H}^{(1+\beta)/2, 1+\beta}(t_0, x_0)$ for every $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$. Otherwise, we can iterate again the process and so on. The result follows.

Acknowledgements

Informal discussions with colleagues at CEREMADE are gratefully acknowledged; we thank in particular, Pierre Cardaliaguet, Djalil Chafaï and Stéphane Mischler. We also thank Denis Belomestny and Nicolas Fournier for insightful comments. This work partially answers a problem that was posed to us by Sylvie Méléard almost two decades ago (at a time we did not have the proper tools to address it!).

1 Proof of Theorem 25 of the Introduction

In order to prove this Theorem, we are going to *juggle* between the probabilistic point of view on the McKean-Vlasov equation like in the textbook of Carmona and Delarue [37] or the lecture notes of Lacker [115], and the analytic point of view like in the textbook of Bogachev, Krylov, Röckner and Shaposhnikov [17].

1.1 Existence, Unicity & Uniformly bounded moments of μ^b

We follow here the probabilistic point of view of [37] in order to prove the points (i) and (ii) of Theorem 25.

Theorem 2.31. *Let $(b, c, \mu_0) \in \mathfrak{B}_{T,d,\mathfrak{h},\mathfrak{n},\mathfrak{c}_2,\mathfrak{c}_3} \times \mathfrak{C}_{T,d,\mathfrak{s}_+,\mathfrak{s}_-,\mathfrak{c}_1} \times \mathfrak{M}_d$. Let $(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{t \in [0,T]}, \tilde{\mathbb{P}})$ a (complete) filtered probability space endowed with a Brownian motion $\tilde{B}$. Let $\tilde{X}_0 \sim \mu_0$, measurable with respect to $\tilde{\mathcal{F}}_0$ and independent from $\tilde{B}$. Then, there exists a unique strong solution $(\bar{X}_t^b)_{t \in [0,T]}$ of the McKean-Vlasov SDE*

$$\begin{cases} d\bar{X}_t^b = b(t, \bar{X}_t^b, \mathcal{L}_{\tilde{\mathbb{P}}}(\bar{X}_t^b))dt + \sigma(t, \bar{X}_t^b)d\tilde{B}_t, & t \in [0, T], \\ \bar{X}_{|t=0}^b = \tilde{X}_0. \end{cases} \quad (2.75)$$

Moreover, $\bar{X}^b$ has continuous trajectories and, if we denote $\mu^b := (\mu_t^b)_{t \in [0,T]}$ where for every $t \in [0, T]$, $\mu_t^b = \mathcal{L}_{\tilde{\mathbb{P}}}(\bar{X}_t^b)$, we have $\mu^b \in \mathcal{C}([0, T]; \mathcal{P}_1(\mathbb{R}^d))$ and, for every $r \geq 1$, there exists some constant $\kappa_{r,b} > 0$ such that

$$\sup_{t \in [0, T]} \mathfrak{m}_r(\mu_t^b) = \sup_{t \in [0, T]} \mathbb{E}_{\tilde{\mathbb{P}}} [|\bar{X}_t^b|^r] \leq \mathbb{E}_{\tilde{\mathbb{P}}} \left[\sup_{t \in [0, T]} |\bar{X}_t^b|^r \right] \leq \kappa_{r,b}.$$

Finally, considering $\bar{X}^b$ as the random function $\bar{X}^b : \omega \in (\tilde{\Omega}, \tilde{\mathcal{F}}_T, \tilde{\mathbb{P}}) \mapsto \bar{X}^b(\omega) = (\bar{X}_t^b(\omega))_{t \in [0,T]} \in \mathcal{C}$, we denote by $\tilde{\mathbb{P}}_b$ the pushforward measure on $(\mathcal{C}, \mathcal{F}_T)$ image of $\tilde{\mathbb{P}}$ by $\bar{X}^b$. We then have that $\tilde{\mathbb{P}}_b$ is the unique probability law on $(\mathcal{C}, \mathcal{F}_T)$ under which $X_0 \sim \mu_0$ and the process

$$\bar{B}_t^b := \int_0^t c^{-1/2}(s, X_s)(dX_s - b(s, X_s, \mu_s^b)ds), \quad t \in [0, T],$$

is a standard Brownian motion on $\mathbb{R}^d$ (where X is the canonical process on $\mathcal{C}$).

Proof. This fact is well-known (it corresponds for instance to the Theorem 4.21 of [37]) and its proof is based on a traditionnal fixed-point argument. For sake of completeness, we give it thereafter (we follow here the proof given in [37]).

Step 1. Let $\nu \in \mathcal{C}([0, T]; \mathcal{P}_2(\mathbb{R}^d))$. Thanks to Assumptions 6, 11, 16, and Remark 20, the *linear* SDE

$$\begin{cases} d\tilde{X}_t = \tilde{\beta}(t, \tilde{X}_t)dt + \sigma(t, \tilde{X}_t)d\tilde{B}_t, & t \in [0, T], \\ \tilde{X}_{|t=0} = \tilde{X}_0, \end{cases} \quad (2.76)$$

where $\tilde{\beta} : (t, x) \in [0, T] \times \mathbb{R}^d \mapsto b(t, x, \nu_t)$, has a unique strong solution that we denote $\tilde{X}^\nu = (\tilde{X}_t^\nu)_{t \in [0, T]}$ (see *e.g.* Theorem 5.2.9. of [110]).

Step 2. By assumption $\mu_0 \in \mathfrak{M}_d$, which means that $\tilde{X}_0 \in L^r$ for every integer $r \geq 1$. Therefore, we also have that $\tilde{X}_t^\nu \in L^r$ for every $t \in [0, T]$, and for every integer $r \geq 1$. Indeed, as $\mathcal{P}_2(\mathbb{R}^d) \subset \mathcal{P}_1(\mathbb{R}^d)$, we have $\nu \in \mathcal{C}([0, T]; \mathcal{P}_1(\mathbb{R}^d))$, and so $\sup_{s \in [0, T]} \mathbf{m}_1(\nu_s) < \infty$, and then there exists some $C_{r, \mathbf{b}} > 0$ such that

$$\begin{aligned} |\tilde{X}_t^\nu|^r &\leq |\tilde{X}_0|^r + \int_0^t |b(s, \tilde{X}_s^\nu, \nu_s)|^r ds + \left| \int_0^t \sigma(s, \tilde{X}_s^\nu) d\tilde{B}_s \right|^r \\ &\leq |\tilde{X}_0|^r + C_{r, \mathbf{b}} T \left(1 + \sup_{s \in [0, T]} \mathbf{m}_1(\nu_s)^r \right) + C_{r, \mathbf{b}} \int_0^t |\tilde{X}_s^\nu|^r ds + \left| \int_0^t \sigma(s, \tilde{X}_s^\nu) d\tilde{B}_s \right|^r, \end{aligned}$$

which implies thanks to Burkholder-Davis-Gundy inequality and Remark 5, that there exists some constant $C'_{r, \mathbf{b}} > 0$ such that, for every $t \in [0, T]$,

$$\mathbb{E}_{\tilde{\mathbb{P}}} [|\tilde{X}_t^\nu|^r] \leq C'_{r, \mathbf{b}} \left(1 + \mathbb{E}_{\tilde{\mathbb{P}}} [|\tilde{X}_0|^r] + \sup_{s \in [0, T]} \mathbf{m}_1(\nu_s)^r + \int_0^t \mathbb{E}_{\tilde{\mathbb{P}}} [|\tilde{X}_s^\nu|^r] ds \right).$$

So, thanks to Grönwall Lemma, for every $t \in [0, T]$,

$$\mathbb{E}_{\tilde{\mathbb{P}}} [|\tilde{X}_t^\nu|^r] \leq C'_{r, \mathbf{b}} \left(1 + \mathbf{m}_r(\mu_0) + \sup_{s \in [0, T]} \mathbf{m}_1(\nu_s)^r \right) e^{C'_{r, \mathbf{b}} t},$$

and then

$$\sup_{t \in [0, T]} \mathbb{E}_{\tilde{\mathbb{P}}} [|\tilde{X}_t^\nu|^r] < \infty.$$

Actually, we even have

$$\mathbb{E}_{\tilde{\mathbb{P}}} \left[\sup_{t \in [0, T]} |\tilde{X}_t^\nu|^r \right] \leq C'_{r, \mathbf{b}} \left(1 + \mathbb{E}_{\tilde{\mathbb{P}}} [|\tilde{X}_0|^r] + \sup_{s \in [0, T]} \mathbf{m}_1(\nu_s)^r + \int_0^t \mathbb{E}_{\tilde{\mathbb{P}}} \left[\sup_{u \in [0, s]} |\tilde{X}_u^\nu|^r \right] ds \right),$$

and so, thanks again to Grönwall Lemma, for every $t \in [0, T]$,

$$\mathbb{E}_{\tilde{\mathbb{P}}} \left[\sup_{t \in [0, T]} |\tilde{X}_t^\nu|^r \right] \leq C'_{r, \mathbf{b}} \left(1 + \mathbf{m}_r(\mu_0) + \sup_{s \in [0, T]} \mathbf{m}_1(\nu_s)^r \right) e^{C'_{r, \mathbf{b}} T}.$$

Step 3. Let us now define the application

$$\Phi : \nu \in \mathcal{C}([0, T]; \mathcal{P}_2(\mathbb{R}^d)) \mapsto (\mathcal{L}(\tilde{X}_t^\nu))_{t \in [0, T]}.$$

We want to show that Φ has a *unique fixed point* in $\mathcal{C}([0, T]; \mathcal{P}_2(\mathbb{R}^d))$. As this metric space is complete, we plan to apply the Fixed-Point Theorem of Banach-Picard. In order to do this, it is sufficient to show that some iteration of Φ is a *contraction*, that is to say there exists some integer $k \geq 1$, and some constant $0 < \kappa < 1$, such that for every $\nu, \nu' \in \mathcal{C}([0, T]; \mathcal{P}_2(\mathbb{R}^d))$,

$$\sup_{t \in [0, T]} \mathcal{W}_2(\Phi^k(\nu')_t, \Phi^k(\nu)_t) \leq \kappa \sup_{t \in [0, T]} \mathcal{W}_2(\nu'_t, \nu_t) .$$

So, let $\nu, \nu' \in \mathcal{C}([0, T]; \mathcal{P}_2(\mathbb{R}^d))$. We have :

$$|\tilde{X}_t^{\nu'} - \tilde{X}_t^\nu|^2 \leq 2T \int_0^t |b(s, \tilde{X}_s^{\nu'}, \nu'_s) - b(s, \tilde{X}_s^\nu, \nu_s)|^2 ds + 2 \left| \int_0^t (\sigma(s, \tilde{X}_s^{\nu'}) - \sigma(s, \tilde{X}_s^\nu)) d\tilde{B}_s \right|^2 .$$

Thanks to Doob's inequality and Assumptions 4 and 6, there exists $C''_{r,b} > 0$ such that

$$\mathbb{E}_{\tilde{\mathbb{P}}} \left[\sup_{s \in [0, t]} |\tilde{X}_s^{\nu'} - \tilde{X}_s^\nu|^2 \right] \leq C''_{r,b} \int_0^t (\mathbb{E}_{\tilde{\mathbb{P}}} \left[\sup_{u \in [0, s]} |\tilde{X}_u^{\nu'} - \tilde{X}_u^\nu|^2 \right] + \mathcal{W}_1(\nu'_s, \nu_s)^2) ds .$$

Then, thanks again to Grönwall Lemma, for every $t \in [0, T]$,

$$\mathbb{E}_{\tilde{\mathbb{P}}} \left[\sup_{s \in [0, t]} |\tilde{X}_s^{\nu'} - \tilde{X}_s^\nu|^2 \right] \leq C''_{r,b} e^{C''_{r,b} T} \int_0^t \mathcal{W}_2(\nu'_s, \nu_s)^2 ds ,$$

and so

$$\sup_{s \in [0, t]} \mathcal{W}_2(\Phi(\nu')_s, \Phi(\nu)_s)^2 \leq C''_{r,b} e^{C''_{r,b} T} \int_0^t \mathcal{W}_2(\nu'_s, \nu_s)^2 ds .$$

Iterating this, we find that for every $k \geq 1$,

$$\sup_{s \in [0, T]} \mathcal{W}_2(\Phi^k(\nu')_s, \Phi^k(\nu)_s)^2 \leq \frac{(C''_{r,b} e^{C''_{r,b} T})^k T^k}{k!} \sup_{s \in [0, T]} \mathcal{W}_2(\nu'_s, \nu_s)^2 ds ,$$

and so, for k large enough, Φ^k is a contraction. We denote by μ^b the *unique fixed-point* of Φ , and then $\bar{X}^b := \tilde{X}^{\mu^b}$ is the *unique strong solution* of the SDE (2.76). Finally, we have that $\mu^b \in \mathcal{C}([0, T]; \mathcal{P}_2(\mathbb{R}^d))$, but as $\mathcal{W}_1 \leq \mathcal{W}_2$, we also have $\mu^b \in \mathcal{C}([0, T]; \mathcal{P}_1(\mathbb{R}^d))$.

Defining $\tilde{\mathbb{P}}_b$ as the pushforward measure on $(\mathcal{C}, \mathcal{F}_T)$ image of $\tilde{\mathbb{P}}$ by $\bar{X}^b$ implies that $\mathcal{L}_{\tilde{\mathbb{P}}_b}(X_0) = \mathcal{L}_{\tilde{\mathbb{P}}}(\bar{X}_0^b)$ by definition and that, for every $0 \leq t_1 \leq \dots \leq t_n \leq T$,

$$\mathcal{L}_{\tilde{\mathbb{P}}_b}(\bar{B}_{t_1}^b, \dots, \bar{B}_{t_n}^b) = \mathcal{L}_{\tilde{\mathbb{P}}}(\tilde{B}_{t_1}, \dots, \tilde{B}_{t_n}) ,$$

so that $\bar{B}^b$ is a d -dimensional standard Brownian motion under $\tilde{\mathbb{P}}_b$.

In the next section, we take back the canonical space point of view in order to prove the point (iv) of Theorem 25.

1.2 Regularity in time of μ^b

Let $(b, c, \mu_0) \in \mathfrak{B}_{T,d,h,n,c_2,c_3} \times \mathfrak{C}_{T,d,s+,s-,c_1} \times \mathfrak{M}_{d,m}$. We have, for every $0 \leq t \leq t' \leq T$,

$$\mathbb{E}_{\tilde{\mathbb{P}}_b} [|X_{t'} - X_t|] \leq \int_t^{t'} \mathbb{E}_{\tilde{\mathbb{P}}_b} [|b(s, X_s, \mu_s^b)|] ds + \mathbb{E}_{\tilde{\mathbb{P}}_b} \left[\int_t^{t'} |c(s, X_s)| ds \right]^{1/2}$$

$$\begin{aligned} &\lesssim \int_t^{t'} \mathbb{E}_{\bar{\mathbb{P}}_b} [1 + |X_s| + \mathbf{m}_1(\mu_s^b)] ds + (t' - t)^{1/2} \\ &\lesssim (t' - t) \left(1 + \sup_{s \in [0, T]} \mathbf{m}_1(\mu_s^b)\right) + (t' - t)^{1/2}, \end{aligned}$$

and so,

$$\mathcal{W}_1(\mu_{t'}^b, \mu_t^b) \leq \mathbb{E}_{\bar{\mathbb{P}}_b} [|X_{t'} - X_t|] \lesssim (t' - t)^{1/2}.$$

Remark 2.32. In the same way, we even have that, for every $r \geq 1$, for every $t, t' \in [0, T]$,

$$\mathbb{E}_{\bar{\mathbb{P}}_b} [|X_{t'} - X_t|^r] \lesssim |t' - t|^{r/2}.$$

1.3 μ^b is the unique probability solution of the Cauchy problem (2)

Let $(b, c, \mu_0) \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3} \times \mathfrak{C}_{T, d, \mathfrak{s}_+, \mathfrak{s}_-, \mathfrak{c}_1} \times \mathfrak{M}_{d, \mathfrak{m}}$. By Itô's formula, for every $\Phi \in \mathcal{C}_c^\infty(\mathbb{R}^d; \mathbb{R})$,

$$\Phi(X_t) = \Phi(X_0) + \sum_{i=1}^d \int_0^t \partial_i \Phi(X_s) dX_s^i + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \partial_{ij}^2 \Phi(X_s) c_{ij}(s, X_s) ds, \quad t \in [0, T],$$

So, taking the expectation $\mathbb{E}_{\bar{\mathbb{P}}_b}$ gives

$$\mathbb{E}_{\bar{\mathbb{P}}_b} [\Phi(X_t)] = \mathbb{E}_{\bar{\mathbb{P}}_b} [\Phi(X_0)] + \int_0^t \mathbb{E}_{\bar{\mathbb{P}}_b} \left[\sum_{i=1}^d \partial_i \Phi(X_s) b^i(s, X_s, \mu_s^b) + \frac{1}{2} \sum_{i,j=1}^d \partial_{ij}^2 \Phi(X_s) c_{ij}(s, X_s) \right] ds,$$

which means that

$$\int_{\mathbb{R}^d} \Phi d\mu_t^b = \int_{\mathbb{R}^d} \Phi d\mu_0 + \int_0^t \int_{\mathbb{R}^d} L^{\mu^b} \Phi(s, x) \mu_s^b(dx) ds.$$

In the same way, for every $0 \leq t' \leq t \leq T$, we have

$$\int_{\mathbb{R}^d} \Phi d\mu_t^b = \int_{\mathbb{R}^d} \Phi d\mu_{t'}^b + \int_{t'}^t \int_{\mathbb{R}^d} L^{\mu^b} \Phi(s, x) \mu_s^b(dx) ds,$$

and so,

$$\int_{\mathbb{R}^d} \Phi d\mu_{t'}^b \xrightarrow{t' \rightarrow 0} \int_{\mathbb{R}^d} \Phi d\mu_0,$$

as

$$\left| \int_{\mathbb{R}^d} \Phi d\mu_{t'}^b - \int_{\mathbb{R}^d} \Phi d\mu_0 \right| \leq \mathbb{E}_{\bar{\mathbb{P}}_b} [|X_{t'} - X_0|] \lesssim \sqrt{t'}.$$

Finally,

$$\begin{aligned} \int_{t'}^t \int_{\mathbb{R}^d} L^{\mu^b} \phi(s, x) \mu_s^b(dx) ds &= \int_{\mathbb{R}^d} \phi d\mu_t^b - \int_{\mathbb{R}^d} \phi d\mu_{t'}^b \\ &\xrightarrow{t' \rightarrow 0} \int_{\mathbb{R}^d} \phi d\mu_t^b - \int_{\mathbb{R}^d} \phi d\mu_0 = \int_0^t \int_{\mathbb{R}^d} L^{\mu^b} \phi(s, x) \mu_s^b(dx) ds. \end{aligned}$$

As, for every $t \in [0, T]$, μ_t^b is the probability law of X_t under $\bar{\mathbb{P}}_b$, and as, following Remarks 5 and 20, $(t, x) \mapsto c(t, x)$ and $(t, x) \mapsto b(t, x, \mu_t^b)$ are locally bounded, these functions are in $L^1([0, T] \times U, \mu^b)$ for every U ball of $\mathbb{R}^d$, and so μ^b is the unique probability solution of the Cauchy problem.

In the next section, we prove the point (v) of Theorem 25.

1.4 Regularity of $b \mapsto \mu^b$

Let $b_1, b_2 \in \mathfrak{B}_{T,d,\mathfrak{h},\mathfrak{n},\mathfrak{c}_2,\mathfrak{c}_3}, (c, \mu_0) \in \mathfrak{C}_{T,d,\mathfrak{s}_+,\mathfrak{s}_-,\mathfrak{c}_1} \times \mathfrak{M}_{d,\mathfrak{m}}$.

Regularity in Wasserstein distance

Let $(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{t \in [0,T]}, \tilde{\mathbb{P}})$ a (complete) filtered probability space endowed with a Brownian motion $\tilde{B}$. Let $\tilde{X}_0 \sim \mu_0$, measurable with respect to $\tilde{\mathcal{F}}_0$ and independent from $\tilde{B}$. We denote by $\tilde{X}^{b_1}$ (resp. $\tilde{X}^{b_2}$) the unique strong solution of the McKean-Vlasov SDE 2.75 with b_1 (resp. b_2) as drift function. We have, for every $t \in [0, T]$,

$$\mathcal{W}_1(\mu_t^{b_2}, \mu_t^{b_1}) \leq \mathbb{E}_{\tilde{\mathbb{P}}} [|\tilde{X}_t^{b_2} - \tilde{X}_t^{b_1}|] ,$$

where

$$\tilde{X}_t^{b_2} - \tilde{X}_t^{b_1} = \int_0^t (b_2(s, \tilde{X}_s^{b_2}, \mu_s^{b_2}) - b_1(s, \tilde{X}_s^{b_1}, \mu_s^{b_1})) ds + \int_0^t (\sigma(s, \tilde{X}_s^{b_2}) - \sigma(s, \tilde{X}_s^{b_1})) d\tilde{B}_s .$$

So, using $|b_2(s, x', \nu') - b_1(s, x, \nu)| \leq |b_2(s, x', \nu') - b_2(s, x, \nu)| + |b_2(s, x, \nu) - b_1(s, x, \nu)|$, we obtain

$$\begin{aligned} \mathbb{E}_{\tilde{\mathbb{P}}} [|\tilde{X}_t^{b_2} - \tilde{X}_t^{b_1}|^2] &\lesssim \mathbb{E}_{\tilde{\mathbb{P}}} \left[\int_0^t |b_2(s, \tilde{X}_s^{b_2}, \mu_s^{b_2}) - b_1(s, \tilde{X}_s^{b_1}, \mu_s^{b_1})|^2 ds \right] \\ &\quad + \mathbb{E}_{\tilde{\mathbb{P}}} \left[\left| \int_0^t (\sigma(s, \tilde{X}_s^{b_2}) - \sigma(s, \tilde{X}_s^{b_1})) d\tilde{B}_s \right|^2 \right] \\ &\lesssim \int_0^t \mathbb{E}_{\tilde{\mathbb{P}}} [|\tilde{X}_s^{b_2} - \tilde{X}_s^{b_1}|^2] ds + \int_0^t \int_{\mathbb{R}^d} |b_2(s, x, \mu_s^{b_1}) - b_1(s, x, \mu_s^{b_1})|^2 \mu_s^{b_1}(dx) ds , \end{aligned}$$

which, by Grönwall Lemma, implies that

$$\mathbb{E}_{\tilde{\mathbb{P}}} [|\tilde{X}_t^{b_2} - \tilde{X}_t^{b_1}|^2] \lesssim \int_0^t \int_{\mathbb{R}^d} |b_2(s, x, \mu_s^{b_1}) - b_1(s, x, \mu_s^{b_1})|^2 \mu_s^{b_1}(dx) ds ,$$

and so finally

$$\mathcal{W}_1(\mu_t^{b_2}, \mu_t^{b_1}) \lesssim \left(\int_0^t \int_{\mathbb{R}^d} |b_2(s, x, \mu_s^{b_1}) - b_1(s, x, \mu_s^{b_1})|^2 \mu_s^{b_1}(dx) ds \right)^{1/2} .$$

As we also have $|b_2(s, x', \nu') - b_1(s, x, \nu)| \leq |b_2(s, x', \nu') - b_1(s, x', \nu')| + |b_1(s, x', \nu') - b_1(s, x, \nu)|$, we obtain, swapping the roles of μ^{b_1} and μ^{b_2} ,

$$\mathcal{W}_1(\mu_t^{b_2}, \mu_t^{b_1}) \lesssim \left(\int_0^t \int_{\mathbb{R}^d} |b_2(s, x, \mu_s^{b_2}) - b_1(s, x, \mu_s^{b_2})|^2 \mu_s^{b_2}(dx) ds \right)^{1/2} .$$

In the parametric case, if $b_i = b_0(\vartheta_i; \cdot)$, $i = 1, 2$, we have

$$\mathcal{W}_1(\mu_t^{\vartheta_2}, \mu_t^{\vartheta_1}) \lesssim \left(\int_0^t \int_{\mathbb{R}^d} |b_0(\vartheta_2; s, x, \mu_s^{\vartheta_1}) - b_0(\vartheta_1; s, x, \mu_s^{\vartheta_1})|^2 \mu_s^{\vartheta_1}(dx) ds \right)^{1/2} \lesssim |\vartheta_2 - \vartheta_1| ,$$

as

$$|b_0(\vartheta_2; s, x, \mu_s^{\vartheta_1}) - b_0(\vartheta_1; s, x, \mu_s^{\vartheta_1})| \lesssim |\vartheta_2 - \vartheta_1| (1 + |x|^{r_1(b_0)} + \mathfrak{m}_{r_2(b_0)}(\mu_s^{\vartheta_1})) .$$

Regularity in total variation distance

We have, for every $t \in (0, T)$, by Pinsker inequality (see *e.g.* Lemma 2.5 in [182])

$$\begin{aligned} \|\mu_t^{b_2} - \mu_t^{b_1}\|_{\text{TV}}^2 &\leq \frac{1}{2} \mathbb{E}_{\bar{\mathbb{P}}_{b_1}} \left[\log \frac{d\bar{\mathbb{P}}_{b_1}}{d\bar{\mathbb{P}}_{b_2}}(X) \Big|_t \right] \\ &= \frac{1}{2} \mathbb{E}_{\bar{\mathbb{P}}_{b_1}} \left[\int_0^t \left((c^{-1}b_1)(s, X_s, \mu_s^{b_1}) - (c^{-1}b_2)(s, X_s, \mu_s^{b_2}) \right)^\top dX_s \right. \\ &\quad \left. - \frac{1}{2} \int_0^t \left(|(c^{-1/2}b_1)(s, X_s, \mu_s^{b_1})|^2 - |(c^{-1/2}b_2)(s, X_s, \mu_s^{b_2})|^2 \right) ds \right]. \end{aligned}$$

And, as under $\bar{\mathbb{P}}_{b_1}$,

$$dX_s = b_1(s, X_s, \mu_s^{b_1})ds + \sigma(s, X_s)dB_s^{b_1},$$

we have

$$\begin{aligned} \|\mu_t^{b_2} - \mu_t^{b_1}\|_{\text{TV}}^2 &\leq \frac{1}{4} \mathbb{E}_{\bar{\mathbb{P}}_{b_1}} \left[\int_0^t \left| (c^{-1/2}b_2)(s, X_s, \mu_s^{b_2}) - (c^{-1/2}b_1)(s, X_s, \mu_s^{b_1}) \right|^2 ds \right] \\ &= \frac{1}{4} \int_0^t \int_{\mathbb{R}^d} \left| (c^{-1/2}b_2)(s, x, \mu_s^{b_2}) - (c^{-1/2}b_1)(s, x, \mu_s^{b_1}) \right|^2 \mu_s^{b_1}(dx) ds \\ &\lesssim \int_0^t \int_{\mathbb{R}^d} \left| b_2(s, x, \mu_s^{b_2}) - b_1(s, x, \mu_s^{b_1}) \right|^2 \mu_s^{b_1}(dx) ds. \end{aligned}$$

And so, finally,

$$\|\mu_t^{b_2} - \mu_t^{b_1}\|_{\text{TV}} \lesssim \left(\int_0^t \int_{\mathbb{R}^d} \left| b_2(s, x, \mu_s^{b_2}) - b_1(s, x, \mu_s^{b_1}) \right|^2 \mu_s^{b_1}(dx) ds \right)^{1/2}.$$

We can also swap the roles of μ^{b_1} and μ^{b_2} in Pinsker's inequality, which gives :

$$\|\mu_t^{b_2} - \mu_t^{b_1}\|_{\text{TV}} \lesssim \left(\int_0^t \int_{\mathbb{R}^d} \left| b_2(s, x, \mu_s^{b_2}) - b_1(s, x, \mu_s^{b_1}) \right|^2 \mu_s^{b_2}(dx) ds \right)^{1/2}.$$

If $b_i = b_0(\vartheta_i; \cdot)$, $i = 1, 2$, it implies that

$$\|\mu_t^{\vartheta_2} - \mu_t^{\vartheta_1}\|_{\text{TV}} \lesssim |\vartheta_2 - \vartheta_1|,$$

as, thanks to $\mathcal{W}_1(\mu_s^{\vartheta_2}, \mu_s^{\vartheta_1}) \lesssim |\vartheta_2 - \vartheta_1|$ and Remark 22

$$\left| b_0(\vartheta_2; s, x, \mu_s^{\vartheta_2}) - b_0(\vartheta_1; s, x, \mu_s^{\vartheta_1}) \right| \lesssim |\vartheta_2 - \vartheta_1| (1 + |x|^{r_1(b_0)} + \mathbf{m}_{r_2(b_0)}(\mu_s^{\vartheta_1})).$$

In the next section, we prove the point (vi) of Theorem 25.

1.5 Existence of a regular density of μ^b

Let $b_1, b_2 \in \mathfrak{B}_{T,d,\mathfrak{h},\mathfrak{n},\mathfrak{c}_2,\mathfrak{c}_3}$, $(c, \mu_0) \in \mathfrak{C}_{T,d,\mathfrak{s}_+, \mathfrak{s}_-, \mathfrak{c}_1} \times \mathfrak{M}_{d,\mathfrak{m}}$. Thanks to Theorem 6.3.1., Corollaries 6.3.2. and 6.4.3. of [17], we know that μ^b admits a *density*, in the sense that there exists some function, still denoted by μ^b , such that $\mu^b : (t, x) \in (0, T) \times \mathbb{R}^d \mapsto \mu_t^b(x)$ and for every $t \in (0, T)$, $\mu_t^b(dx) = \mu_t^b(x)dx$.

Moreover, for every $0 < R'_1 < R_1 < R_2 < R'_2 < T$, $t_0 \in [R_1, R_2]$, and $R' > R > 0$, $x_0 \in \mathbb{R}^d$,

- for every $\delta \in [1, \frac{d+2}{d+1}]$, there exists some constant $C_{R_1, R_2, x_0, R, R'_1, R'_2, R', \mathfrak{b}} > 0$ such that

$$\|\mu^b\|_{L^\delta([R_1, R_2] \times B(x_0, R))} \leq C_{R_1, R_2, x_0, R, R'_1, R'_2, R', \mathfrak{b}} ;$$

- for every $\delta > d + 2$, $\mu^b \in \mathfrak{H}^{1, \delta}([R_1, R_2] \times B(x_0, R))$ (which definition is in the Notations section, see (1)), which implies, thanks to the Parabolic Sobolev Embeddings (reminded *e.g.* in Theorem 6.2.2. of [17]), that μ^b is locally Hölder on $(0, T) \times \mathbb{R}^d$: for every $\alpha, \beta \in (0, \frac{1}{2})$,

$$\mu^b \in \mathcal{C}^{\alpha, \beta}([R_1, R_2] \times B(x_0, R)) \subset \mathcal{H}^{\alpha, \beta}(t_0, x_0) ;$$

and in particular, for every $\beta \in (0, \frac{1}{2})$,

$$\mu^b \in \mathcal{C}^{\frac{\beta}{2}, \beta}([R_1, R_2] \times B(x_0, R)) \subset \mathcal{H}^{\frac{\beta}{2}, \beta}(t_0, x_0) .$$

It also implies that μ^b has a local supremum bound on $(0, T) \times \mathbb{R}^d$: for every $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$, for every $0 < R_1 < t_0 < R_2 < T$, $R > 0$, for every $(t, x) \in [R_1, R_2] \times B(x_0, R)$, for every $\beta \in (0, 1)$,

$$\begin{aligned} |\mu^b(t, x) - \mu^b(t_0, x_0)| &\leq [\mu^b]_{\frac{\beta}{2}, \beta; [R_1, R_2] \times B(x_0, R)} (|t - t_0|^{\beta/2} + |x - x_0|^\beta) \\ &\leq [\mu^b]_{\frac{\beta}{2}, \beta; [R_1, R_2] \times B(x_0, R)} ((R_2 - R_1)^{\beta/2} + R^\beta) , \end{aligned}$$

and so, there exists some constant $\kappa_3 = \kappa_3(t_0, x_0, R, R_1, R_2, \mathfrak{b}, \beta)$

$$|\mu^b(t, x)| \leq |\mu^b(t_0, x_0)| + [\mu^b]_{\frac{\beta}{2}, \beta; [R_1, R_2] \times B(x_0, R)} ((R_2 - R_1)^{\beta/2} + R^\beta) \leq \kappa_3 .$$

In the next section, we prove the point (viii) of Theorem 25.

1.6 Positivity & Local bounds of μ^b

We plan to apply the Theorem 8.2.1. and the Corollary 8.2.2. of [17], that is to say consequences of Parabolic Harnack type inequalities. Let us rewrite our equation in divergence form. We have μ^b solution of $\partial_t \mu^b = \mathcal{L}_{\mu^b}^* \mu^b$, where for every smooth test function $\Phi \in \mathcal{C}_c^\infty((0, T) \times \mathbb{R}^d; \mathbb{R})$,

$$\mathcal{L}_{\mu^b} \Phi(t, x) = \sum_{i, j=1}^d \partial_{x_i} (\tilde{c}_{ij}(t, x) \partial_{x_j} \Phi(t, x)) + \sum_{i=1}^d \tilde{b}^i(t, x) \partial_{x_i} \Phi(t, x) ,$$

where for every $1 \leq i, j \leq d$,

$$\tilde{c}_{ij}(t, x) := \frac{1}{2} c_{ij}(t, x) ,$$

$$\tilde{b}^i(t, x) := b^i(t, x, \mu_t^b) - \frac{1}{2} \sum_{j=1}^d \partial_{x_j} c_{ij}(t, x) .$$

We remark that $\partial_{x_j} c_{ij}$ is well defined in the generalized sense, and bounded, as $c_{ij}(t, \cdot)$ is Lipschitz by assumption. It then implies that $\tilde{c}_{ij}, \tilde{b}^i$ are locally bounded on $(0, T) \times \mathbb{R}^d$, and so integrable with respect to μ^b . Let us check the assumptions of Theorem 8.2.1. and Corollary 8.2.2. of [17]. Let

$J = [R_1, R_2] \subset (0, T)$ and U some ball of $\mathbb{R}^d$. For every $t \in J$, $\mu_t^b \in W^{1,1}(U)$ as $\mu_t^b \in W^{1,\delta}(U)$ for every $\delta > d + 2$ by the precedent subsection, and so, for every $1 \leq i \leq d$, $\partial_{x_i} \mu_t^b \in L^\delta(U) \subset L^1(U)$ and

$$\int_J \int_U |\partial_{x_i} \mu^b(t, x)| dx dt \leq |U|^{\frac{\delta-1}{\delta}} \int_J \|\partial_{x_i} \mu_t^b\|_{L^\delta(U)} dt \leq |U|^{\frac{\delta-1}{\delta}} (|J| + \int_J \|\partial_{x_i} \mu_t^b\|_{L^\delta(U)}^\delta dt) < \infty .$$

It then implies that, for every $1 \leq i, j \leq d$, $\tilde{c}_{ij} \partial_{x_i} \mu^b \in L^1(J \times U)$ as $\tilde{c}_{ij}$ is bounded on $J \times U$, and $\tilde{b}^i \mu^b \in L^1(J \times U)$ as these functions are locally bounded on $(0, T) \times \mathbb{R}^d$. Finally, for every $t \in (0, T)$, μ_t^b does not vanish identically as $\mu_t^b(x) dx \in \mathcal{P}^1(\mathbb{R}^d)$. It remains to check the *Lyapunov condition* type of Theorem 8.2.1. of [17]. We have

$$|\tilde{b}(t, x)| \leq |b(t, x, \mu_t^b)| + \frac{1}{2} \sum_{j=1}^d |\partial_{x_j} c_{ij}(t, x)| ,$$

and so, thanks to Assumption 6 and Remark 20, there exists some constant $\tilde{C}_b > 0$ such that

$$\sup_{t \in (0, T)} |\tilde{b}(t, x)| \leq \tilde{C}_b \left(1 + \frac{|x|}{2}\right) .$$

We then take, keeping the notations in [17], $\theta = 2$, and $V : z \in \mathbb{R}_+ \mapsto \tilde{C}_b(1 + z)$ which is well increasing and positive. We conclude by applying Corollary 8.2.2. in [17] : there exists some constant $K_{b, R_1, R_2} > 0$ such that for every $t \in [R_1, R_2]$ and $x \in \mathbb{R}^d$,

$$e^{-K_{b, R_1, R_2}(1+V(|x|)^2+|x|^2)} \leq \mu_t^b(x) \leq e^{K_{b, R_1, R_2}(1+V(|x|)^2+|x|^2)} .$$

We finally deduce from it that there exists some constant $\tilde{K}_{b, R_1, R_2} > 0$ such that for every $t \in [R_1, R_2]$ and $x \in \mathbb{R}^d$,

$$e^{-\tilde{K}_{b, R_1, R_2}(1+|x|^2)} \leq \mu_t^b(x) \leq e^{\tilde{K}_{b, R_1, R_2}(1+|x|^2)} ,$$

which gives *upper and lower local bounds* on μ^b , and its *positivity* :

$$\mu_t^b(x) > 0 \text{ for every } (t, x) \in (0, T) \times \mathbb{R}^d .$$

In the next section, we prove the point (ix) of Theorem 25.

1.7 Injectivity of $b \mapsto \mu^b$

Let $b_1, b_2 \in \mathfrak{B}_{T, d, \mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3} \cap \mathcal{C}([0, T] \times \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d); \mathbb{R}^d)$ such that $\mu^{b_1} = \mu^{b_2}$. Then, for every smooth test function $\Phi \in \mathcal{C}_c^\infty((0, T) \times \mathbb{R}^d \times \mathbb{R})$, for $k = 1, 2$,

$$\int_0^T \int_{\mathbb{R}^d} \left(\partial_t \Phi(t, x) + \frac{1}{2} \sum_{i,j=1}^d \partial_{x_i x_j}^2 \Phi(t, x) c_{ij}(t, x) + \sum_{i=1}^d \partial_{x_i} \Phi(t, x) b_k^i(t, x, \mu_t^{b_k}) \right) \mu_t^{b_k}(x) dx dt = 0 ,$$

and so

$$\int_0^T \int_{\mathbb{R}^d} \nabla_x \Phi(t, x)^\top (b_2(t, x, \mu_t^{b_1}) - b_1(t, x, \mu_t^{b_1})) \mu_t^{b_1}(x) dx dt = 0 .$$

Let $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$. We now consider some sequence $(\Phi_n)_n$ of smooth test functions with compact support in some neighbourhood of (t_0, x_0) , and such that $(\nabla \Phi_n)_n$ converges uniformly to $(t, x) \mapsto b_2(t, x, \mu_t^{b_1}) - b_1(t, x, \mu_t^{b_1})$ in $[r_1, r_2] \times B(x_0; r)$, for some $0 < r_1 < t_0 < r_2 < T$, $r > 0$. Then, by dominated convergence theorem,

$$\int_{[r_1, r_2] \times B(x_0; r)} |b_2(t, x, \mu_t^{b_1}) - b_1(t, x, \mu_t^{b_1})|^2 \mu_t^{b_1}(x) dx dt = 0 ,$$

and so, by continuity of $(t, x) \mapsto |b_2(t, x, \mu_t^{b_1}) - b_1(t, x, \mu_t^{b_1})|^2 \mu_t^{b_1}(x)$ and positivity of $\mu_t^{b_1}(x)$, for every $(t, x) \in [r_1, r_2] \times B(x_0; r)$,

$$b_1(t, x, \mu_t^{b_1}) = b_2(t, x, \mu_t^{b_1}) .$$

Finally, as it is true for every $(t_0, x_0) \in (0, T) \times \mathbb{R}^d$, by continuity of b_1 and b_2 ,

$$b_1(t, x, \mu_t^{b_1}) = b_2(t, x, \mu_t^{b_1}) \text{ for every } (t, x) \in [0, T] \times \mathbb{R}^d .$$

In the next section, we prove the Theorem 29 given in the Introduction.

1.8 More regularity in the Vlasov case

Let us denote, for every integer $n \geq 1$,

$$\delta_{G;n} := \mathbf{1}_{\delta_G > n+1} + (\delta_G - \lfloor \delta_G \rfloor) \mathbf{1}_{\delta_G \in (n, n+1]} , \quad \delta_{F;n} := \mathbf{1}_{\delta_F > n+1} + (\delta_F - \lfloor \delta_F \rfloor) \mathbf{1}_{\delta_F \in (n, n+1]} ,$$

and $\delta_{G,F} := \min(\delta_{G;1}, \delta_{F;1})$.

Step 1. Let $1 \leq i \leq d$, $t, t' \in [0, T]$ and $x, x' \in \mathbb{R}^d$, we have

$$\begin{aligned} |\beta^i(t', x') - \beta^i(t, x)| &\leq |G^i(x') - G^i(x)| + \left| \int_{\mathbb{R}^d} (F^i(x' - y) - F^i(x - y)) \mu_{t'}^b(dy) \right| \\ &\quad + \left| \int_{\mathbb{R}^d} F^i(x - y) (\mu_{t'}^b - \mu_t^b)(dy) \right| . \end{aligned}$$

On the one hand, G^i is Lipschitz by assumption. On the other hand, F^i is Lipschitz, as it is in $\mathcal{C}^1$ and with compact support. So

$$|\beta^i(t', x') - \beta^i(t, x)| \leq |G|_{\text{Lip}} |x' - x| + |D_x F|_{\infty} (|x' - x| + \mathcal{W}_1(\mu_{t'}^b, \mu_t^b)) \lesssim |t' - t|^{1/2} + |x' - x| ,$$

and $\beta^i \in \mathcal{H}^{\frac{1}{2}, 1}([0, T] \times \mathbb{R}^d)$. Moreover, for every $(t, x) \in [0, T] \times \mathbb{R}^d$,

$$\partial_{x_i} \beta^i(t, x) = \partial_{x_i} G^i(x) + \int_{\mathbb{R}^d} \partial_{x_i} F^i(x - y) \mu_t^b(dy) .$$

Let $t, t' \in (0, T)$, $x_0 \in \mathbb{R}^d$, $R > 0$, and $x, x' \in B(x_0; R)$. As $\partial_{x_i} F^i \in \mathcal{H}_{\text{loc}}^{\delta_{F;1}}(\mathbb{R}^d)$,

$$\begin{aligned} \left| \int_{\mathbb{R}^d} \partial_{x_i} F^i(x - y) (\mu_{t'}^b - \mu_t^b)(dy) \right| &\leq \mathbb{E}_{\mathbb{P}_b} [|\partial_{x_i} F^i(x - X_{t'}) - \partial_{x_i} F^i(x - X_t)|] \\ &\lesssim \mathbb{E}_{\mathbb{P}_b} [|X_{t'} - X_t|^{\delta_{F;1}}] \lesssim |t' - t|^{\delta_{F;1}/2} , \end{aligned}$$

and, as $\partial_{x_i} G^i \in \mathcal{H}_{\text{loc}}^{\delta_{G;1}}(\mathbb{R}^d)$, we finally have

$$\begin{aligned} |\partial_{x_i} \beta^i(t', x') - \partial_{x_i} \beta^i(t, x)| &\leq |\partial_{x_i} G^i(x') - \partial_{x_i} G^i(x)| + \left| \int_{\mathbb{R}^d} (\partial_{x_i} F^i(x' - y) - \partial_{x_i} F^i(x - y)) \mu_{t'}^b(dy) \right| \\ &\quad + \left| \int_{\mathbb{R}^d} \partial_{x_i} F^i(x - y) (\mu_{t'}^b - \mu_t^b)(dy) \right| \\ &\lesssim |x' - x|^{\delta_{G;1}} + |t' - t|^{\delta_{F;1}/2} + |x' - x|^{\delta_{F;1}} \lesssim |t' - t|^{\delta_{G,F}/2} + |x' - x|^{\delta_{G,F}}. \end{aligned}$$

Therefore, as $\beta^i, \partial_{x_i} \beta^i \in \mathcal{H}_{\text{loc}}^{\frac{\delta_{G,F}}{2}, \delta_{G,F}}((0, T) \times \mathbb{R}^d)$, $1 \leq i \leq d$, by Theorem 28, $\mu^b \in \mathcal{C}^{1,2}((0, T) \times \mathbb{R}^d) \cap \mathcal{C}_b([0, T] \times \mathbb{R}^d)$ and is a classical solution of the problem (18) (it is of course the unique one, by uniqueness of the probability solution). Moreover, for every $\delta \in (0, 1) \cap (0, \delta_{G,F})$,

$$\mu^b \in \mathcal{C}_{\text{loc}}^{1+\frac{\delta}{2}, 2+\delta}((0, T) \times \mathbb{R}^d),$$

so taking back the part $\partial_{x_i} \beta^i$, we have for every $(t, x) \in (0, T) \times \mathbb{R}^d$,

$$\partial_{x_i} \beta^i(t, x) = \partial_{x_i} G^i(x) + \int_{\mathbb{R}^d} \partial_{x_i} F^i(x - y) \mu_t^b(dy) = \partial_{x_i} G^i(x) + \int_{\mathbb{R}^d} F^i(z) \partial_{x_i} \mu_t^b(x - z) dz.$$

and, so for every $t, t' \in [R_1, R_2] \subset (0, T)$, $R > 0$, $x_0 \in \mathbb{R}^d$, $x, x' \in B(x_0; R)$,

$$\begin{aligned} |\partial_{x_i} \beta^i(t', x') - \partial_{x_i} \beta^i(t, x)| &\lesssim |x' - x|^{\delta_{G;1}} + \left| \int_{\mathbb{R}^d} (F^i(x' - y) - F^i(x - y)) \partial_{x_i} \mu_{t'}^b(y) dy \right| \\ &\quad + \left| \int_{\mathbb{R}^d} \partial_{x_i} F^i(x - y) (\mu_{t'}^b(y) - \mu_t^b(y)) dy \right| \\ &\lesssim |x' - x|^{\delta_{G;1}} + |D_x F|_{\infty} |x' - x| \sup_{(t'', y'') \in [R_1, R_2] \times B_{x_0, R, F}} |\partial_{x_i} \mu_{t''}^b(y'')| \\ &\quad + |D_x F|_{\infty} |\text{supp}(F)| \sup_{(t'', y'') \in [R_1, R_2] \times B_{x_0, R, F}} |\partial_t \mu_{t''}^b(y'')| |t' - t| \\ &\lesssim |x' - x|^{\delta_{G;1}} + |x' - x| + |t' - t|, \end{aligned}$$

where $B_{x_0, R, F} := B(x_0; R) - \text{supp}(F)$. So, $\partial_{x_i} \beta^i$ is actually in $\mathcal{H}_{\text{loc}}^{\frac{\delta_{G;1}}{2}, \delta_{G;1}}((0, T) \times \mathbb{R}^d)$, and so, for every $\delta_1 \in (0, 1) \cap (0, \delta_{G;1})$, $\mu^b \in \mathcal{C}_{\text{loc}}^{1+\frac{\delta_1}{2}, 2+\delta_1}((0, T) \times \mathbb{R}^d) \subset \mathcal{H}_{\text{loc}}^{1+\frac{\delta_1}{2}, 2+\delta_1}((0, T) \times \mathbb{R}^d)$.

Step 2. If $\delta_G \in (1, 2]$, it ends the proof, otherwise we can iterate the previous method, thanks to the results of Section 8.12 of [113]. So, let us suppose now that $\delta_G > 2$, we then have $\delta_{G;1} = 1$, and so, for every $1 \leq i \leq d$, $\partial_{x_i} \beta^i \in \mathcal{H}_{\text{loc}}^{\frac{1}{2}, 1}((0, T) \times \mathbb{R}^d)$. Moreover, for every $1 \leq i, j \leq d$, and $(t, x) \in (0, T) \times \mathbb{R}^d$,

$$\partial_{x_j x_i}^2 \beta^i(t, x) = \partial_{x_j x_i}^2 G^i(x) + \int_{\mathbb{R}^d} F^i(z) \partial_{x_j x_i}^2 \mu_t^b(x - z) dz.$$

As $\partial_{x_j x_i}^2 G^i \in \mathcal{H}_{\text{loc}}^{\delta_{G;2}}(\mathbb{R}^d)$, and $\partial_{x_j x_i}^2 \mu^b \in \mathcal{H}_{\text{loc}}^{\frac{\delta_1}{2}, \delta_1}((0, T) \times \mathbb{R}^d)$ for every $\delta_1 \in (0, 1)$, we have, for every $\delta_2 \in (0, 1) \cap (0, \delta_{G;2})$, $\partial_{x_j x_i}^2 \beta^i \in \mathcal{H}_{\text{loc}}^{\frac{\delta_2}{2}, \delta_2}((0, T) \times \mathbb{R}^d)$, and so, thanks to Theorem 8.12.1 of [113],

$$\partial_{x_j} \mu^b \in \mathcal{C}_{\text{loc}}^{1+\frac{\delta_2}{2}, 2+\delta_2}((0, T) \times \mathbb{R}^d),$$

and then

$$\mu^b \in \mathcal{H}_{\text{loc}}^{1+\frac{\delta_2}{2}, 3+\delta_2}((0, T) \times \mathbb{R}^d).$$

In the same way, for every $1 \leq i \leq d$, and $(t, x) \in (0, T) \times \mathbb{R}^d$,

$$\partial_t \beta^i(t, x) = \int_{\mathbb{R}^d} F^i(x - y) \partial_t \mu_t^b(y) dy,$$

$$\partial_t \partial_{x_i} \beta^i(t, x) = \int_{\mathbb{R}^d} \partial_{x_i} F^i(x - y) \partial_t \mu_t^b(y) dy,$$

and so, as $\partial_t \mu^b \in \mathcal{H}_{\text{loc}}^{\frac{\delta_1}{2}, \delta_1}((0, T) \times \mathbb{R}^d)$ for every $\delta_1 \in (0, 1)$, $\partial_t \beta^i, \partial_t \partial_{x_i} \beta^i \in \mathcal{H}_{\text{loc}}^{\frac{\delta_1}{2}, \delta_1}((0, T) \times \mathbb{R}^d)$ for every $\delta_1 \in (0, 1)$. Thanks to the Remark 8.12.4 of [113], for every $\delta_1 \in (0, 1)$,

$$\partial_t \mu^b \in \mathcal{C}_{\text{loc}}^{1+\frac{\delta_1}{2}, 2+\delta_1}((0, T) \times \mathbb{R}^d),$$

and so

$$\mu^b \in \mathcal{H}_{\text{loc}}^{2+\frac{\delta_1}{2}, 2+\delta_1}((0, T) \times \mathbb{R}^d).$$

To conclude this step, we can say that for every $\delta_2 \in (0, 1) \cap (0, \delta_{G;2}]$,

$$\mu^b \in \mathcal{H}_{\text{loc}}^{2+\frac{\delta_2}{2}, 3+\delta_2}((0, T) \times \mathbb{R}^d) \subset \mathcal{H}_{\text{loc}}^{\frac{3+\delta_2}{2}, 3+\delta_2}((0, T) \times \mathbb{R}^d).$$

Step 3. If $\delta_G \in (2, 3]$, $[\delta_G] = 2$, and so it ends the proof. Otherwise we can again iterate again the method, and so on.

In the next section, we prove the point (vii) of Theorem 25.

1.9 Continuity in the parameter ϑ

We mimick here the proofs given in [18]. Let $(\vartheta_n)_n, \vartheta \in \Theta$ such that $\vartheta_n \rightarrow \vartheta$. We denote by $\mu^{\vartheta_n}(dtdx) = \rho^{\vartheta_n}(t, x) dtdx$. Let $\phi : \mathbb{N} \rightarrow \mathbb{N}$ increasing, let us show that there exists some $\psi : \mathbb{N} \rightarrow \mathbb{N}$ increasing such that $\rho^{\vartheta_{\phi(\psi(n))}}$ converges locally uniformly on $(0, T) \times \mathbb{R}^d$. Let us take a countable covering of $(0, T) \times \mathbb{R}^d$ by a sequence of increasing compacts $(K_q)_q$ where $K_q := [s_{q,1}, s_{q,2}] \times U_q$ with $s_{q+1,1} < s_{q,1} < s_{q,2} < s_{q+1,2}$ and U_q a strict subset of U_{q+1} .

Let us first show that for every q , $\{\rho_{|K_q}^{\vartheta_{\phi(n)}}\}_n$ is relatively compact. In order to do that we will apply Arzela-Ascoli Theorem.

First step. Let us show that $\{\rho_{|K_q}^{\vartheta_{\phi(n)}}\}_n$ is bounded. In order to do that, we want to apply Theorem 6.5.4 in [17]. We just have to verify that

- $\mu^{\vartheta_{\phi(n)}}$ is a locally finite Borel measure on $(0, T) \times \mathbb{R}^d$;
- c is uniformly elliptic;
- $((t, x) \mapsto c_{ij}(t, x)) \in L_{\text{loc}}^1((0, T) \times \mathbb{R}^d, \mu^{\vartheta_{\phi(n)}})$;
- $((t, x) \mapsto b^i(\theta_{\phi(n)}; t, x, \mu_t^{\theta_{\phi(n)}})) \in L_{\text{loc}}^1((0, T) \times \mathbb{R}^d, \mu^{\theta_{\phi(n)}})$;
- for every test function $\Phi \in \mathcal{C}_c^\infty((0, T) \times \mathbb{R}^d; \mathbb{R})$,

$$\int_{(0, T) \times \mathbb{R}^d} (\partial_t \Phi + L_{c,b}^{\mu^{\theta_{\phi(n)}}} \Phi) d\mu^{\theta_{\phi(n)}} = 0.$$

All these items are well verified by definition of $\mu^{\vartheta_{\phi(n)}}$, and so the conclusion of Theorem 6.5.4 in [17] applies : for every $\gamma > \frac{d+2}{2}$, there exists some constant

$$C = C(d, T, \gamma, U_q, U_{q+1}, s_{q,1}, s_{q,2}, s_{q+1,1}, s_{q+1,2}, \sigma_{\pm})$$

such that

$$\|\rho^{\vartheta_{\phi(n)}}\|_{L^\infty(K_q)} \leq C \int_{s_{q+1,1}}^{s_{q+1,2}} \int_{U_{q+1}} (1 + \sigma_+^\gamma + |c^{-1/2} g_{\phi(n)}|^{2\gamma}) \mu^{\vartheta_{\phi(n)}}(dt, dx)$$

where

$$g_{\phi(n)}^i(t, x) := b^i(\vartheta_{\phi(n)}; t, x, \mu_t^{\vartheta_{\phi(n)}}) - \frac{1}{2} \sum_{j=1}^d \partial_j c_{ij}(t, x),$$

which is well-defined almost surely as c is Lipschitz in x by Assumption. By the Assumptions on b and by 2.31, we easily show that

$$\int_{K_{q+1}} |b(\vartheta_{\phi(n)}; t, x, \mu_t^{\vartheta_{\phi(n)}})|^{2\gamma} \mu^{\vartheta_{\phi(n)}}(dt dx) \lesssim 1 + \sup_{\vartheta' \in \Theta} \sup_{t \in [0, T]} \mathbf{m}_{2 \max(r_1, r_2) \gamma}(\mu_t^{\vartheta'}),$$

and so finally, by the uniform ellipticity of c , there exists some

$$C' = C'(d, T, \Theta, \gamma, K_q, K_{q+1}, \sigma_{\pm}, r_1, r_2)$$

such that, for every n ,

$$\|\rho^{\vartheta_{\phi(n)}}\|_{L^\infty(K_q)} \leq C'.$$

Second step. Now let us show that $\{\rho^{\vartheta_{\phi(n)}}\}_n$ is uniformly equicontinuous. In order to do that, we plan to apply Theorem 6.2.7 in [17]. Let us take $q' > q$, $r > 0$ and $\tau \in (0, T)$ such that

$$K_q \subset Q(r) \subset Q(3r) \subset K_{q'},$$

where

$$Q(r) := (\tau - r^2, \tau) \times B_{|\cdot|_\infty}(0, r).$$

The assumptions of this theorem

- c uniformly elliptic,
- there exist $q_1 > 2$, $q_2 > 1$ such that $\frac{d}{2q_1} + \frac{1}{q_2} < \frac{1}{2}$ and $g_{\phi(n)} \in L^{q_1, q_2}((0, T) \times U_{q'})$, where

$$\|g_{\phi(n)}^i\|_{L^{q_1, q_2}((0, T) \times U_{q'})} := \left(\int_0^T \left(\int_{U_{q'}} |g_{\phi(n)}^i(t, x)|^{q_1} dx \right)^{q_2/q_1} dt \right)^{1/q_2},$$

- $\rho^{\vartheta_{\phi(n)}} \in \mathbb{H}^{1,2}((0, T) \times U_{q'})$,

are easily verified, and so, there exists some constants

$$\gamma' = \gamma'(d, \sigma_{\pm}, \sup_{n'} \|g_{\phi(n')}^i\|_{L^{q_1, q_2}((0, T) \times U_{q'})}),$$

$$C'' = C''(d, \sigma_{\pm}, \sup_{n'} \|g_{\phi(n')}^i\|_{L^{q_1, q_2}((0, T) \times U_{q'})}),$$

such that for every $(t, x), (s, y) \in K_q$,

$$|\rho^{\theta_{\phi(n)}}(t, x) - \rho^{\theta_{\phi(n)}}(s, y)| \leq C'' r^{-\gamma'} (|x - y| + |t - s|^{1/2})^{\gamma'} \times \|\rho^{\theta_{\phi(n)}}\|_{L^\infty(Q(3r))},$$

and by what precedes, we already know that, for every n ,

$$\|\rho^{\theta_{\phi(n)}}\|_{L^\infty(Q(3r))} \leq C',$$

which ends the second step.

Third step. By a diagonalisation argument, there exists an increasing function $\psi : \mathbb{N} \rightarrow \mathbb{N}$ such that, for every q , $(\rho_{K_q}^{\theta_{\phi(\psi(n))}})_n$ converges uniformly in K_q . We will denote by $\tilde{\rho}_q$ on K_q . As $K_q \subset K_{q+1}$, by unicity of the limit on K_q we have $\tilde{\rho}_q = \tilde{\rho}_{q+1}|_{K_q}$, and so we can define for $(t, x) \in K_q$

$$\rho(t, x) := \tilde{\rho}_q(t, x),$$

and we have $\rho^{\theta_{\phi(\psi(n))}}$ converges locally uniformly on $(0, T) \times \mathbb{R}^d$ to ρ , which is then non-negative and continuous on $(0, T) \times \mathbb{R}^d$.

Fourth step. Let us denote $m(dt, dx) := \rho(t, x) dx dt$ and $m_t(dx) := \rho(t, x) dx$. As ρ is continuous, m and m_t are Borel measures, and it remains to show that for almost every $t \in (0, T)$, $m_t \in \mathcal{P}(\mathbb{R}^d)$ which is implied by $\int_{\mathbb{R}^d} m_t(dx) = 1$.

By the Prokhorov theorem, the family $\{\mu_t^{\theta_{\phi(\psi(n))}}\}_n$ is tight if and only if it is relatively compact for the weak topology. Let us show that the family $\{\mu_t^{\theta'}\}_{\theta' \in \Theta}$ is uniformly tight. In order to do so, we are going to find a Borel function $W : \mathbb{R}^d \mapsto \mathbb{R}$ positive non-decreasing and locally bounded such that $W(x) \rightarrow \infty$ as $|x| \rightarrow \infty$ and such that

$$C_W := \sup_{\theta' \in \Theta} \sup_{t \in [0, T]} \int_{\mathbb{R}^d} W d\mu_t^{\theta'} < \infty.$$

We can simply take $W(x) = 1 + |x|$ which verifies these assumptions.

Let us show that it implies that $\{\mu^{\theta'}\}_{\theta' \in \Theta}$ is uniformly tight. Let $\epsilon > 0$,

$$\begin{aligned} \mu^{\theta'}((0, T) \times \mathbb{R}^d \setminus ([k_1^\epsilon, k_2^\epsilon] \times \bar{B}(0, R_\epsilon))) &= \int_{(0, T) \times \mathbb{R}^d} \mathbf{1}_{t \notin [k_1^\epsilon, k_2^\epsilon]} \mathbf{1}_{|x| > R_\epsilon} \mu^{\theta'}(dt dx) \\ &\leq \int_{(0, T) \times \mathbb{R}^d} \mathbf{1}_{t \notin [k_1^\epsilon, k_2^\epsilon]} \frac{W(|x|)}{W(R_\epsilon)} \mu^{\theta'}(dt dx) \frac{1}{W(R_\epsilon)} \int_0^T \mathbf{1}_{t \notin [k_1^\epsilon, k_2^\epsilon]} \int_{\mathbb{R}^d} W(x) \mu_t^{\theta'}(dx) dt \leq \frac{TC_W}{W(R_\epsilon)}. \end{aligned}$$

So it is sufficient to take $\frac{TC_W}{W(R_\epsilon)} < \epsilon$, and this is possible as $W(x) \rightarrow \infty$ when $|x| \rightarrow \infty$.

So the family $\{\mu_t^{\theta_{\phi(\psi(n))}}\}_n$ is tight, and then there exists an increasing function $\chi : \mathbb{N} \rightarrow \mathbb{N}$ such that $(\mu_t^{\theta_{\phi(\chi(n))}})_n$ converges weakly. We denote by $\tilde{m}_t$ its weak limit. So, we have for every test function $\Phi \in \mathcal{C}_b(\mathbb{R}^d; \mathbb{R})$,

$$\int_{\mathbb{R}^d} \Phi(x) \mu_t^{\theta_{\phi(\psi(n))}}(dx) \rightarrow \int_{\mathbb{R}^d} \Phi(x) \tilde{m}_t(dx)$$

as $n \rightarrow \infty$, and taking $\Phi \equiv 1$, we obtain that $\int_{\mathbb{R}^d} \tilde{m}_t(dx) = 1$.

Let us now show that $\tilde{m}_t = m_t$. First, we will show that for every compact $A \in \mathcal{B}(\mathbb{R}^d)$, we have that $\tilde{m}_t(A) = m_t(A)$. Let us take a sequence of non-negative test functions $(\Phi_k)_k \subset \mathcal{C}_c^\infty(\mathbb{R}^d; \mathbb{R})$

such that Φ_k converges uniformly to $\mathbf{1}_A$. Then, successively by dominated convergence theorem, weak convergence of $(\mu_t^{\vartheta_{\phi(\psi(\chi(n)))}})_n$, convergence as $k \rightarrow \infty$ uniformly in n , and again dominated convergence theorem as $\rho^{\vartheta_{\phi(\psi(\chi(n)))}}$ converges locally uniformly on $(0, T) \times \mathbb{R}^d$ to ρ ,

$$\begin{aligned} \tilde{m}_t(A) &= \int_{\mathbb{R}^d} \mathbf{1}_A(x) \tilde{m}_t(dx) = \lim_{k \rightarrow \infty} \int_{\mathbb{R}^d} \Phi_k(x) \tilde{m}_t(dx) = \lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} \Phi_k(x) \mu_t^{\vartheta_{\phi(\psi(\chi(n)))}}(dx) , \\ &= \lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \int_{\mathbb{R}^d} \Phi_k(x) \mu_t^{\vartheta_{\phi(\psi(\chi(n)))}}(dx) = \lim_{n \rightarrow \infty} \int_A \rho^{\vartheta_{\phi(\psi(\chi(n)))}}(t, x) dx = \int_A \rho(t, x) dx = m_t(A) . \end{aligned}$$

Now, we will show this equality for all $A \in \mathcal{B}(\mathbb{R}^d)$. As $\tilde{m}_t, m_t$ are Borel measures, on $\mathbb{R}^d$ a locally compact space in which every open set is σ -compact, they are regular, and so

$$\tilde{m}_t(A) = \sup\{\tilde{m}_t(K); K \text{ compact} \subset A\} = \sup\{m_t(K); K \text{ compact} \subset A\} = m_t(A) .$$

We conclude that $\tilde{m}_t = m_t$ and so $\int_{\mathbb{R}^d} m_t(dx) = 1$.

Fifth step. By Scheffé's lemma

$$\|\mu^{\vartheta_{\phi(\phi(n))}} - m\|_{TV} \rightarrow 0$$

as $n \rightarrow \infty$, because

$$\|\mu^{\vartheta_{\phi(\phi(n))}} - m\|_{TV} = \int_0^T \int_{\mathbb{R}^d} |\rho^{\vartheta_{\phi(\phi(n))}}(t, x) - \rho(t, x)| dx dt .$$

It remains to show that $L_{\vartheta}^* m = \partial_t m$, where for some test function Φ ,

$$L_{\vartheta} \Phi(t, x) := \sum_{i,j=1}^d \frac{1}{2} c_{ij}(t, x) \partial_{ij}^2 \Phi(x) + \sum_{i=1}^d b^i(\vartheta; t, x, \mu_t^{\vartheta}) \partial_i \Phi(x) ,$$

as then by unicity of the solution of this equation on $(0, T) \times \mathbb{R}^d$, we will have $m = \mu^{\vartheta}$. We remind that, by definition, it consists in showing that

- for every $1 \leq i, j \leq d$, $((t, x) \mapsto c_{ij}(t, x)) \in L_{\text{loc}}^1((0, T) \times \mathbb{R}^d, m)$,
- for every $1 \leq i \leq d$, $((t, x) \mapsto b^i(\vartheta; t, x, \mu_t^{\vartheta})) \in L_{\text{loc}}^1((0, T) \times \mathbb{R}^d, m)$,
- for every test function $\Phi \in \mathcal{C}_c^\infty((0, T) \times \mathbb{R}^d; \mathbb{R})$,

$$\int_{(0, T) \times \mathbb{R}^d} (\partial_t \Phi + L_{\vartheta} \Phi) m(dt dx) = 0 .$$

Let $1 \leq i, j \leq d$, and $J \subset (0, T), U \subset \mathbb{R}^d$ such that $J \times U$ is compact. We have

$$\int_{J \times U} |c_{ij}(t, x)| m(dt dx) \leq |c|_{\infty} T ,$$

$$\begin{aligned} \int_{J \times U} |b^i(\vartheta; t, x, \mu_t^{\vartheta})| m(dt dx) &\lesssim \int_{J \times U} (1 + |x|^{r_1} + \mathbf{m}_{r_2}(\mu_t^{\vartheta})) \rho(t, x) dx dt , \\ &\lesssim 1 + \sup_{t \in [0, T]} \mathbf{m}_{r_2}(\mu_t^{\vartheta}) . \end{aligned}$$

We first have, as $\partial_t \Phi, \partial_{ij}^2 \in \mathcal{C}_c((0, T) \times \mathbb{R}^d)$ by dominated convergence theorem

$$\begin{aligned} \int_{(0, T) \times \mathbb{R}^d} \partial_t \Phi(t, x) m(dt dx) &= \lim_{n \rightarrow \infty} \int_{(0, T) \times \mathbb{R}^d} \partial_t \Phi(t, x) \rho^{\vartheta_{\phi(\psi(n))}}(t, x) dx dt, \\ \int_{(0, T) \times \mathbb{R}^d} \sum_{i, j=1}^d \frac{1}{2} c_{ij}(t, x) \partial_{ij}^2 \Phi(t, x) m(dt, dx) &= \lim_{n \rightarrow \infty} \int_{(0, T) \times \mathbb{R}^d} \sum_{i, j=1}^d \frac{1}{2} c_{ij}(t, x) \partial_{ij}^2 \Phi(t, x) \rho^{\vartheta_{\phi(\psi(n))}}(t, x) dx dt. \end{aligned}$$

As for the term

$$\int_{(0, T) \times \mathbb{R}^d} \sum_{i=1}^d b^i(\vartheta; t, x, \mu_t^{\vartheta}) \partial_i \Phi(t, x) \rho(t, x) dx dt,$$

let us show that it is equal to

$$\lim_{n \rightarrow \infty} \int_{(0, T) \times \mathbb{R}^d} \sum_{i=1}^d b^i(\vartheta_{\phi(\psi(n))}; t, x, \mu_t^{\vartheta_{\phi(\psi(n))}}) \partial_i \Phi(t, x) \rho^{\vartheta_{\phi(\psi(n))}}(t, x) dx dt.$$

As $\partial_i \Phi \in C_c((0, T) \times \mathbb{R}^d)$ by dominated convergence theorem, we have

$$\begin{aligned} \int_{(0, T) \times \mathbb{R}^d} \sum_{i=1}^d b^i(\vartheta; t, x, \mu_t^{\vartheta}) \partial_i \Phi(t, x) \rho(t, x) dx dt \\ = \lim_{n \rightarrow \infty} \int_{(0, T) \times \mathbb{R}^d} \sum_{i=1}^d b^i(\vartheta; t, x, \mu_t^{\vartheta}) \partial_i \Phi(t, x) \rho^{\vartheta_{\phi(\psi(n))}}(t, x) dx dt, \end{aligned}$$

and so it remains to show that for every $1 \leq i \leq d$,

$$\int_{(0, T) \times \mathbb{R}^d} |b^i(\vartheta_{\phi(\psi(n))}; t, x, \mu_t^{\vartheta_{\phi(\psi(n))}}) - b^i(\vartheta; t, x, \mu_t^{\vartheta})| |\partial_i \Phi(t, x)| \rho^{\vartheta_{\phi(\psi(n))}}(t, x) dx dt \rightarrow 0$$

as $n \rightarrow \infty$. As $\rho^{\vartheta_{\phi(\psi(n))}}$ converges locally uniformly to ρ , and $\partial_i \Phi \in C_c((0, T) \times \mathbb{R}^d)$, we have that for n large enough, for (t, x) in the support of $\partial_i \Phi$, $\rho^{\vartheta_{\phi(\psi(n))}}(t, x) \leq 1 + \rho(t, x)$, and as

$$|b^i(\vartheta_{\phi(\psi(n))}; t, x, \mu_t^{\vartheta_{\phi(\psi(n))}}) - b^i(\vartheta; t, x, \mu_t^{\vartheta})| \lesssim |\vartheta_{\phi(\psi(n))} - \vartheta| (1 + |x|^{r_1} + \mathbf{m}_{r_2}(\mu_t^{\vartheta_{\phi(\psi(n))}})),$$

we have for n large enough

$$\begin{aligned} \int_{(0, T) \times \mathbb{R}^d} |b^i(\vartheta_{\phi(\psi(n))}; t, x, \mu_t^{\vartheta_{\phi(\psi(n))}}) - b^i(\vartheta; t, x, \mu_t^{\vartheta})| |\partial_i \Phi(t, x)| \rho^{\vartheta_{\phi(\psi(n))}}(t, x) dx dt \\ \lesssim |\vartheta_{\phi(\psi(n))} - \vartheta| \int_{(0, T) \times \mathbb{R}^d} (1 + |x|^{r_1} + \mathbf{m}_{r_2}(\mu_t^{\vartheta_{\phi(\psi(n))}})) |\partial_i \Phi(t, x)| (1 + \rho(t, x)) dx dt, \end{aligned}$$

which tends to 0 as $n \rightarrow \infty$. It ends the proof of $\partial_t m = L_{\vartheta} m$, and so $m = \mu^{\vartheta}$, and $\rho = \rho^{\vartheta}$ almost everywhere.

Final step. Let $(\vartheta, t, x) \in \Theta \times (0, T) \times \mathbb{R}^d$ and $(\vartheta_n, t_n, x_n) \rightarrow (\vartheta, t, x)$ as $n \rightarrow \infty$. There exists some compact $K_q \subset (0, T) \times \mathbb{R}^d$ such that $(\vartheta, t, x) \in K_q$ and, for every n , $(\vartheta_n, t_n, x_n) \in K_q$. Following the proof of the equicontinuity of $\{\rho_{\theta_{\phi(n)}}|_{K_q}\}_n$, we have

$$\begin{aligned} |\rho^{\vartheta_n}(t_n, x_n) - \rho^{\vartheta}(t, x)| &\leq |\rho^{\vartheta_n}(t_n, x_n) - \rho^{\vartheta_n}(t, x)| + |\rho^{\vartheta_n}(t, x) - \rho^{\vartheta}(t, x)| \\ &\lesssim (|x_n - x| + |t_n - t|^{1/2})^{\gamma'} + \|\rho^{\vartheta_n} - \rho^{\vartheta}\|_{L^\infty(K_q)} \end{aligned}$$

which tends to 0 as $n \rightarrow \infty$.

2 Proof of Theorem 31

2.1 Proof of Theorem 31(i) : $\mathbb{P}_b^N$ is well-defined

This result corresponds to the Lemma 3.2. of [115] and we remind here its proof for sake of completeness. We use the formulation (12) of the equation (2.77), that is to say

$$\begin{cases} d\mathbb{X}_t^N = b^{(N)}(t, \mathbb{X}_t^N)dt + \sigma^{(N)}(t, \mathbb{X}_t^N)d\mathbb{B}_t^{N,b}, & t \in [0, T], \\ \mathcal{L}(\mathbb{X}_0^N) = \mu_0^{\otimes N}, \end{cases}$$

$$b^{(N)} : (t, (x^1, \dots, x^N)) \in [0, T] \times (\mathbb{R}^d)^N \mapsto \left(b(t, x^1, \mu^N(x^1, \dots, x^N)), \dots, b(t, x^N, \mu^N(x^1, \dots, x^N)) \right)^\top,$$

$$\sigma^{(N)} : (t, (x^1, \dots, x^N)) \in [0, T] \times (\mathbb{R}^d)^N \mapsto \begin{pmatrix} \sigma(t, x^1) & & (0) \\ & \ddots & \\ (0) & & \sigma(t, x^N) \end{pmatrix}.$$

Then, $b^{(N)}, \sigma^{(N)}$ are Lipschitz in $(x^1, \dots, x^N)$, uniformly with respect to t . Indeed for every $t \in [0, T]$, $x^{(N)} = (x^1, \dots, x^N), y^{(N)} = (y^1, \dots, y^N) \in (\mathbb{R}^d)^N$, on the one hand

$$\begin{aligned} |b^{(N)}(t, y^{(N)}) - b^{(N)}(t, x^{(N)})|^2 &= \sum_{i=1}^N |b(t, y^i, \mu^N(y^{(N)})) - b(t, x^i, \mu^N(x^{(N)}))|^2 \\ &\leq 2|b|_{\text{Lip}}^2 \sum_{i=1}^N \left(|y^i - x^i|^2 + \mathcal{W}_1^2(\mu^N(y^{(N)}), \mu^N(x^{(N)})) \right) \\ &\leq 2|b|_{\text{Lip}}^2 \left(|y^{(N)} - x^{(N)}|^2 + N\mathcal{W}_1^2(\mu^N(y^{(N)}), \mu^N(x^{(N)})) \right), \end{aligned}$$

and so, as

$$\mathcal{W}_1^2(\mu^N(y^{(N)}), \mu^N(x^{(N)})) = \left(\inf_{|\Phi|_{\text{Lip}} \leq 1} \frac{1}{N} \sum_{i=1}^N (\Phi(y^i) - \Phi(x^i)) \right)^2 \leq \frac{1}{N} \sum_{i=1}^N |y^i - x^i|^2 = \frac{1}{N} |y^{(N)} - x^{(N)}|^2,$$

we finally have

$$|b^{(N)}(t, y^{(N)}) - b^{(N)}(t, x^{(N)})| \leq 2|b|_{\text{Lip}} |y^{(N)} - x^{(N)}|,$$

and then

$$|b^{(N)}(t, x^{(N)})| \leq |b^{(N)}(t, (0, \dots, 0))| + 2|b|_{\text{Lip}} |x^{(N)}| \leq \sqrt{N} \sup_{t \in [0, T]} |b(t, 0, \delta_0)| + 2|b|_{\text{Lip}} |x^{(N)}|.$$

On the other hand,

$$|\sigma^{(N)}(t, y^{(N)}) - \sigma^{(N)}(t, x^{(N)})| \leq |\sigma|_{\text{Lip}} |y^{(N)} - x^{(N)}|,$$

and therefore, thanks to Remark 5,

$$|\sigma^{(N)}(t, x^{(N)})| \leq |\sigma^{(N)}(t, (0, \dots, 0))| + |\sigma|_{\text{Lip}} |x^{(N)}| \leq \sqrt{Nd}\sigma_+ + |\sigma|_{\text{Lip}} |x^{(N)}|.$$

So, thanks to usual SDE's results (see *e.g.* Theorem 5.2.9. of [110]), for some fixed integer $N \geq 1$, for a (complete) filtered given probability space $(\tilde{\Omega}^N, \tilde{\mathcal{F}}^N, (\tilde{\mathcal{F}}_t^N)_{t \in [0, T]}, \tilde{\mathbb{P}}^N)$, endowed with

N independent Brownian motions $\tilde{B}^1, \dots, \tilde{B}^N$ of dimension d , and a vector $(\tilde{X}_0^1, \dots, \tilde{X}_0^N)$ of law $\mu_0^{\otimes N}$, measurable with respect to $\tilde{\mathcal{F}}_0$, and independent of $\tilde{B}^1, \dots, \tilde{B}^N$, the system of SDE's

$$\begin{cases} d\tilde{X}_t^i = b(t, \tilde{X}_t^i, \mu^N(\tilde{X}_t^1, \dots, \tilde{X}_t^N))dt + \sigma(t, \tilde{X}_t^i)d\tilde{B}_t^i, & t \in [0, T], \\ (\tilde{X}_t^1, \dots, \tilde{X}_t^N)|_{t=0} = (\tilde{X}_0^1, \dots, \tilde{X}_0^N), \end{cases} \quad (2.77)$$

has a unique strong solution $(\tilde{X}^1, \dots, \tilde{X}^N)$, and so its law under $\tilde{\mathbb{P}}^N$, which is denoted by $\mathbb{P}_b^N$, is well defined.

2.2 Proof of Theorem 31 (iv)

Let $(c, \mu_0) \in \mathfrak{C}_{T,d,\mathfrak{s}+, \mathfrak{s}-, \mathfrak{c}_1} \times \mathfrak{M}_d$, and let $(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{t \in [0, T]}, \tilde{\mathbb{P}})$ a (complete) filtered probability space endowed with a Brownian motion $\tilde{B}$. Let $\tilde{X}_0 \sim \mu_0$, measurable with respect to $\tilde{\mathcal{F}}_0$ and independent from $\tilde{B}$. We know that, for each $b \in \mathfrak{B}_{T,d,\mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}$, there exists a unique strong solution $(\tilde{X}_t^b)_{t \in [0, T]}$ of the McKean-Vlasov SDE

$$\begin{cases} d\tilde{X}_t = b(t, \tilde{X}_t, \mathcal{L}_{\tilde{\mathbb{P}}}(\tilde{X}_t))dt + \sigma(t, \tilde{X}_t)d\tilde{B}_t, & t \in [0, T], \\ \tilde{X}_{|t=0} = \tilde{X}_0, \end{cases}$$

and that the law of $\tilde{X}$ under $\tilde{\mathbb{P}}$ is indeed $\bar{\mathbb{P}}_b$. Moreover, we denote by $(\tilde{X}_t^0)_{t \in [0, T]}$ the unique strong solution of the SDE

$$\begin{cases} d\tilde{X}_t = \sigma(t, \tilde{X}_t)d\tilde{B}_t, & t \in [0, T], \\ \tilde{X}_{|t=0} = \tilde{X}_0, \end{cases}$$

and its law under $\tilde{\mathbb{P}}$ by W (which is consistent with the notation given in the Introduction). Then, for every $b \in \mathfrak{B}_{T,d,\mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}$, as $\int_0^T |(c^{-1/2}b)(t, \tilde{X}_t^b, \mu_t^b)|^2 dt$ and $\int_0^T |(c^{-1/2}b)(t, \tilde{X}_t^0, \mu_t^b)|^2 dt$ are finite $\tilde{\mathbb{P}}$ -almost surely, thanks to Girsanov Theorem (see *e.g.* the result 7.6.4. in [131]), we have that $\bar{\mathbb{P}}_b$ and W are equivalent and,

$$\log \frac{d\bar{\mathbb{P}}_b}{dW}(X)|_{\mathcal{F}_T} = \int_0^T (c^{-1}b)(t, X_t, \mu_t^b)^\top dX_t - \frac{1}{2} \int_0^T |(c^{-1/2}b)(t, X_t, \mu_t^b)|^2 dt.$$

In the same way, for every $b_1, b_2 \in \mathfrak{B}_{T,d,\mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}$, $\bar{\mathbb{P}}_{b_1}$ and $\bar{\mathbb{P}}_{b_2}$ are equivalent and,

$$\begin{aligned} \log \frac{d\bar{\mathbb{P}}_{b_1}}{d\bar{\mathbb{P}}_{b_2}}(X)|_{\mathcal{F}_T} &= \int_0^T ((c^{-1}b_1)(t, X_t, \mu_t^{b_1}) - (c^{-1}b_2)(t, X_t, \mu_t^{b_2}))^\top dX_t \\ &\quad - \frac{1}{2} \int_0^T (|(c^{-1/2}b_1)(t, X_t, \mu_t^{b_1})|^2 - |(c^{-1/2}b_2)(t, X_t, \mu_t^{b_2})|^2) dt. \end{aligned}$$

Now, let $N \geq 1$ and $(c, \mu_0) \in \mathfrak{C}_{T,d,\mathfrak{s}+, \mathfrak{s}-, \mathfrak{c}_1} \times \mathfrak{M}_d$, and let $(\tilde{\Omega}^N, \tilde{\mathcal{F}}^N, (\tilde{\mathcal{F}}_t^N)_{t \in [0, T]}, \tilde{\mathbb{P}}^N)$ a (complete) filtered probability space endowed with N independant d -dimensional Brownian motions $\tilde{B}^i, 1 \leq i \leq N$. Let $(\tilde{X}_0^1, \dots, \tilde{X}_0^N) \sim \mu_0^{\otimes N}$, measurable with respect to $\tilde{\mathcal{F}}_0^N$ and independent from $\tilde{B}^N$. We know that, for each $b \in \mathfrak{B}_{T,d,\mathfrak{h}, \mathfrak{n}, \mathfrak{c}_2, \mathfrak{c}_3}$, there exists a unique strong solution $(\tilde{X}_t^{i,b})_{1 \leq i \leq N, t \in [0, T]}$ of the system of SDE's

$$\begin{cases} d\tilde{X}_t^i = b(t, \tilde{X}_t^i, \mu^N(\tilde{X}_t^1, \dots, \tilde{X}_t^N))dt + \sigma(t, \tilde{X}_t^i)d\tilde{B}_t^i, & 1 \leq i \leq N, t \in [0, T], \\ (\tilde{X}_t^1, \dots, \tilde{X}_t^N)|_{t=0} = (\tilde{X}_0^1, \dots, \tilde{X}_0^N), \end{cases}$$

and the law of $(\tilde{X}^{1,b}, \dots, \tilde{X}^{N,b})$ under $\tilde{\mathbb{P}}^N$ is indeed $\mathbb{P}_b^N$. Moreover, we denote by $(\tilde{X}_t^{i,0})_{1 \leq i \leq N, t \in [0, T]}$ the unique strong solution of the SDE

$$\begin{cases} d\tilde{X}_t^i = \sigma(t, \tilde{X}_t^i) d\tilde{B}_t^i, & t \in [0, T], \\ (\tilde{X}_t^1, \dots, \tilde{X}_t^N)|_{t=0} = (\tilde{X}_0^1, \dots, \tilde{X}_0^N), \end{cases}$$

and its law under $\tilde{\mathbb{P}}^N$ by $\mathbb{W}^N$ (which is consistent with the notation given in the Introduction). Then, for every $b \in \mathfrak{B}_{T,d,h,n,c_2,c_3}$, as the quantities

$$\sum_{i=1}^N \int_0^T |(c^{-1/2}b)(t, \tilde{X}_t^{i,b}, \mu^N(\tilde{X}_t^{1,b}, \dots, \tilde{X}_t^{N,b}))|^2 dt, \quad \sum_{i=1}^N \int_0^T |(c^{-1/2}b)(t, \tilde{X}_t^{i,0}, \mu^N(\tilde{X}_t^{1,0}, \dots, \tilde{X}_t^{N,0}))|^2 dt,$$

are finite $\tilde{\mathbb{P}}^N$ -almost surely, thanks to Girsanov Theorem (see *e.g.* the result 7.6.4. in [131]), we have that $\mathbb{P}_b^N$ and $\mathbb{W}^N$ are equivalent and,

$$\log \frac{d\mathbb{P}_b^N}{d\mathbb{W}^N}(\mathbb{X}^N)|_{\mathcal{F}_T} = \sum_{i=1}^N \int_0^T (c^{-1}b)(t, X_t^i, \mu_t^N)^\top dX_t^i - \frac{1}{2} \sum_{i=1}^N \int_0^T |(c^{-1/2}b)(t, X_t^i, \mu_t^N)|^2 dt.$$

In the same way, we know that, for each $b \in \mathfrak{B}_{T,d,h,n,c_2,c_3}$, there exists a unique strong solution $(\bar{X}_t^{i,b})_{1 \leq i \leq N, t \in [0, T]}$ of the system of SDE's

$$\begin{cases} d\bar{X}_t^i = b(t, \bar{X}_t^i, \mu_t^b) dt + \sigma(t, \bar{X}_t^i) d\tilde{B}_t^i, & 1 \leq i \leq N, t \in [0, T], \\ (\bar{X}_t^1, \dots, \bar{X}_t^N)|_{t=0} = (\tilde{X}_0^1, \dots, \tilde{X}_0^N), \end{cases}$$

and the law of $(\bar{X}^{1,b}, \dots, \bar{X}^{N,b})$ under $\tilde{\mathbb{P}}^N$ is indeed $\tilde{\mathbb{P}}_b^{\otimes N}$. Then, for every $b \in \mathfrak{B}_{T,d,h,n,c_2,c_3}$, as the quantities

$$\sum_{i=1}^N \int_0^T |(c^{-1/2}b)(t, \bar{X}_t^{i,b}, \mu_t^b)|^2 dt, \quad \sum_{i=1}^N \int_0^T |(c^{-1/2}b)(t, \tilde{X}_t^{i,0}, \mu_t^b)|^2 dt,$$

are finite $\tilde{\mathbb{P}}^N$ -almost surely, thanks to Girsanov Theorem, we have that $\tilde{\mathbb{P}}_b^{\otimes N}$ and $\mathbb{W}^N$ are equivalent and,

$$\log \frac{d\tilde{\mathbb{P}}_b^{\otimes N}}{d\mathbb{W}^N}(\mathbb{X}^N)|_{\mathcal{F}_T} = \sum_{i=1}^N \int_0^T (c^{-1}b)(t, X_t^i, \mu_t^b)^\top dX_t^i - \frac{1}{2} \sum_{i=1}^N \int_0^T |(c^{-1/2}b)(t, X_t^i, \mu_t^b)|^2 dt.$$

Finally, for every $b \in \mathfrak{B}_{T,d,h,n,c_2,c_3}$, $\mathbb{P}_b^N$ and $\tilde{\mathbb{P}}_b^{\otimes N}$ are equivalent, and

$$\begin{aligned} \log \frac{d\tilde{\mathbb{P}}_b^{\otimes N}}{d\mathbb{P}_b^N}(\mathbb{X}^N)|_{\mathcal{F}_T} &= \sum_{i=1}^N \int_0^T ((c^{-1}b)(t, X_t^i, \mu_t^b) - (c^{-1}b)(t, X_t^i, \mu_t^N))^\top dX_t^i \\ &\quad - \frac{1}{2} \sum_{i=1}^N \int_0^T (|(c^{-1/2}b)(t, X_t^i, \mu_t^b)|^2 - |(c^{-1/2}b)(t, X_t^i, \mu_t^N)|^2) dt. \end{aligned}$$

3 Proof of Proposition 32

We generalize the proof in [65]. By Pinsker's inequality (see Lemma 2.5. in [182]),

$$\|\mathbb{P}_b^N - \bar{\mathbb{P}}_b^{\otimes N}\|_{\text{TV}}^2 \leq \frac{1}{2} \mathbb{E}_{\bar{\mathbb{P}}_b^{\otimes N}} \left[\log \frac{d\bar{\mathbb{P}}_b^{\otimes N}}{d\mathbb{P}_b^N} \right],$$

so it is sufficient to prove that

$$\limsup_{N \rightarrow \infty} \sup_{b \in \mathfrak{B}_{T,d,b,n,c_2,c_3}} \mathbb{E}_{\bar{\mathbb{P}}_b^{\otimes N}} \left[\log \frac{d\bar{\mathbb{P}}_b^{\otimes N}}{d\mathbb{P}_b^N} \right] \lesssim 1.$$

By Girsanov theorem, we have

$$\begin{aligned} \mathbb{E}_{\bar{\mathbb{P}}_b^{\otimes N}} \left[\log \frac{d\bar{\mathbb{P}}_b^{\otimes N}}{d\mathbb{P}_b^N} \right] &= \frac{1}{2} \sum_{i=1}^N \int_0^T \mathbb{E}_{\bar{\mathbb{P}}_b^{\otimes N}} \left[|(c^{-1/2}b)(s, X_s^i, \mu_s^N) - (c^{-1/2}b)(s, X_s^i, \mu_s^b)|^2 \right] ds \\ &\lesssim \sum_{i=1}^N \int_0^T \mathbb{E}_{\bar{\mathbb{P}}_b^{\otimes N}} \left[|b(s, X_s^i, \mu_s^N) - b(s, X_s^i, \mu_s^b)|^2 \right] ds. \end{aligned}$$

We use Lemma 111 and Theorem 2 of [74]. In the case (i) of Assumption 16, we have

$$\begin{aligned} \mathbb{E}_{\bar{\mathbb{P}}_b^{\otimes N}} \left[\log \frac{d\bar{\mathbb{P}}_b^{\otimes N}}{d\mathbb{P}_b^N} \right] &\lesssim N \int_0^T \mathbb{E}_{\bar{\mathbb{P}}_b^{\otimes N}} [\mathcal{W}_1(\mu_s^N, \mu_s^b)^2] ds \\ &\lesssim N \int_0^T \int_0^\infty \bar{\mathbb{P}}_b^{\otimes N}(\mathcal{W}_1(\mu_s^N, \mu_s^b) \geq z^{1/2}) dz ds \\ &\lesssim N \int_0^\infty e^{-cNz} dz \\ &\lesssim 1. \end{aligned}$$

In the case (iii) of Assumption 16, we have by Lemma 112, for N large enough

$$\begin{aligned} \mathbb{E}_{\bar{\mathbb{P}}_b^{\otimes N}} \left[\log \frac{d\bar{\mathbb{P}}_b^{\otimes N}}{d\mathbb{P}_b^N} \right] &\lesssim \sum_{i=1}^N \int_0^T \mathbb{E}_{\bar{\mathbb{P}}_b^{\otimes N}} \left[\sum_{\ell=1}^{k_b^*-1} \frac{1}{\ell!} \left| \int_{(\mathbb{R}^d)^\ell} \delta_\mu^\ell b(s, X_s^i, y^\ell, \mu_s^b) (\mu_s^N - \mu_s^b)^{\otimes \ell} (dy^\ell) \right|^2 \right] ds \\ &\lesssim N \sum_{\ell=1}^{k_b^*-1} \frac{1}{\ell!} \int_0^T \frac{K_\ell}{N - (k_b^* - 1)} |\delta_\mu^\ell b(s, \cdot)|_{\text{Lip}}^2 ds \lesssim 1. \end{aligned}$$

Finally, in the case (ii) of Assumption 16,

$$\begin{aligned} \mathbb{E}_{\bar{\mathbb{P}}_b^{\otimes N}} \left[\log \frac{d\bar{\mathbb{P}}_b^{\otimes N}}{d\mathbb{P}_b^N} \right] &\lesssim \sum_{i=1}^N \int_0^T \mathbb{E}_{\bar{\mathbb{P}}_b^{\otimes N}} \left[\sum_{\ell=1}^{k_b^*-1} \frac{1}{\ell!} \left| \int_{(\mathbb{R}^d)^\ell} \delta_\mu^\ell b(s, X_s^i, y^\ell, \mu_s^b) (\mu_s^N - \mu_s^b)^{\otimes \ell} (dy^\ell) \right|^2 \right] ds \\ &\quad + N \int_0^T \mathbb{E}_{\bar{\mathbb{P}}_b^{\otimes N}} [\mathcal{W}_1(\mu_s^N, \mu_s^b)^2 \wedge \mathcal{W}_1(\mu_s^N, \mu_s^b)^{2k_b^*}] ds. \end{aligned}$$

For the first part in the right term we mimic the proof of the case (iii). For the second part, we only show it for the subcase $d \geq 3$ and $k_b^* \geq d/2$ (the case $d = 2$ gives the same result, but is more technical because of the log term in this case of Theorem 2 in [74]) :

$$\mathbb{E}_{\bar{\mathbb{P}}_b^{\otimes N}} [\mathcal{W}_1(\mu_s^N, \mu_s^b)^2 \wedge \mathcal{W}_1(\mu_s^N, \mu_s^b)^{2k_b^*}] = \int_0^\infty \bar{\mathbb{P}}_b^{\otimes N} (\mathcal{W}_1(\mu_s^N, \mu_s^b)^2 \wedge \mathcal{W}_1(\mu_s^N, \mu_s^b)^{2k_b^*} \geq z) dz$$

$$\lesssim \int_0^1 \bar{\mathbb{P}}_b^{\otimes N}(\mathcal{W}_1(\mu_s^N, \mu_s^b) \geq z^{1/2k_b^*}) dz + \int_1^\infty \bar{\mathbb{P}}_b^{\otimes N}(\mathcal{W}_1(\mu_s^N, \mu_s^b) \geq z^{1/2}) dz.$$

On the one hand

$$\int_0^1 \bar{\mathbb{P}}_b^{\otimes N}(\mathcal{W}_1(\mu_s^N, \mu_s^b) \geq z^{1/2k_b^*}) dz \lesssim \int_0^1 e^{-\mathfrak{C}N z^{d/2k_b^*}} dz \lesssim \frac{1}{N},$$

and on the other hand

$$\int_1^\infty \bar{\mathbb{P}}_b^{\otimes N}(\mathcal{W}_1(\mu_s^N, \mu_s^b) \geq z^{1/2}) dz \lesssim \int_1^\infty e^{-\mathfrak{C}N z} dz \lesssim \frac{1}{N}.$$

So finally

$$\mathbb{E}_{\bar{\mathbb{P}}_b^{\otimes N}} \left[\log \frac{d\bar{\mathbb{P}}_b^{\otimes N}}{d\mathbb{P}_b^{K,N}} \right] \lesssim 1.$$

To establish the propagation of chaos result, we proceed like in [142]. Let $1 \leq K \leq N$, we have, for every $t \in [0, T]$, and $\mathcal{B}((\mathbb{R}^d)^K)$ the Borel σ -algebra of $(\mathbb{R}^d)^K$,

$$\begin{aligned} \|\mu_t^{K,N} - \mu_t^{b \otimes K}\|_{\text{TV}} &= \sup_{B \in \mathcal{B}((\mathbb{R}^d)^K)} |\mu_t^{K,N}(B) - \mu_t^{b \otimes K}(B)| \\ &= \sup_{B \in \mathcal{B}((\mathbb{R}^d)^K)} |\mathbb{P}_b^{K,N}((X_t^1, \dots, X_t^K) \in B) - \bar{\mathbb{P}}_b^{\otimes K}((X_t^1, \dots, X_t^K) \in B)| \\ &\leq \|\mathbb{P}_b^{K,N} - \bar{\mathbb{P}}_b^{\otimes K}\|_{\text{TV}}, \end{aligned}$$

and, again by Pinsker's inequality,

$$\|\mathbb{P}_b^{K,N} - \bar{\mathbb{P}}_b^{\otimes K}\|_{\text{TV}} \leq \frac{1}{\sqrt{2}} \mathbb{E}_{\bar{\mathbb{P}}_b^{\otimes K}} \left[\log \frac{d\bar{\mathbb{P}}_b^{\otimes K}}{d\mathbb{P}_b^{K,N}} \right]^{1/2}.$$

As $\mathbb{E}_{\bar{\mathbb{P}}_b^{\otimes K}} \left[\log \frac{d\bar{\mathbb{P}}_b^{\otimes K}}{d\mathbb{P}_b^{K,N}} \right]$ is the Kullback-Leibler divergence between $\bar{\mathbb{P}}_b^{\otimes K}$ and $\mathbb{P}_b^{K,N}$, we also have (see *e.g.* [53, 11]) that

$$\mathbb{E}_{\bar{\mathbb{P}}_b^{\otimes K}} \left[\log \frac{d\bar{\mathbb{P}}_b^{\otimes K}}{d\mathbb{P}_b^{K,N}} \right] \leq \frac{K}{N} \mathbb{E}_{\bar{\mathbb{P}}_b^{\otimes N}} \left[\log \frac{d\bar{\mathbb{P}}_b^{\otimes N}}{d\mathbb{P}_b^{K,N}} \right],$$

so finally,

$$\|\mu_t^{K,N} - \mu_t^{b \otimes K}\|_{\text{TV}} \lesssim \sqrt{\frac{K}{N}}$$

4 Parametric estimation of the drift : discrete observations

4.1 Consistency of the AMLE : Proof of Theorem 59

As Θ is well compact, we will follow Theorem 3.2.8 in [54]. We have first to show that $\vartheta \in \Theta \mapsto U^{N,M}(\vartheta; y)$ is continuous in every $y = (y_j^i)_{1 \leq i \leq N, 0 \leq j \leq M} \in (\mathbb{R}^d)^{N \times (M+1)}$, and equal to

$$-\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M \left((c^{-1}b_0)(\vartheta; t_{j-1}^M, y_{j-1}^i, \mu^N(y_{j-1})) - (c^{-1}b_0)(\vartheta_0; t_{j-1}^M, y_{j-1}^i, \mu^N(y_{j-1})) \right)^\top (y_j^i - y_{j-1}^i)$$

$$+ \frac{1}{2} \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M \left(\left| (c^{-1/2} b_0)(\vartheta; t_{j-1}^M, y_{j-1}^i, \mu^N(y_{j-1})) \right|^2 - \left| (c^{-1/2} b_0)(\vartheta_0; t_{j-1}^M, y_{j-1}^i, \mu^N(y_{j-1})) \right|^2 \right) \Delta_M,$$

the continuity of $\vartheta \in \Theta \mapsto U^{N,M}(\vartheta; y)$ follows directly from the continuity of $\vartheta \in \Theta \mapsto b_0(\vartheta; \cdot)$. Then, we have to show that our contrast process $U^{N,M}(\vartheta; \mathbb{X}^{N,M})$ verifies a contrast property (cf [54], [79]), that is to say there exists some function $K : \Theta \times \Theta \rightarrow \mathbb{R}^+$ such that for every $\vartheta_0 \in \Theta$, $\vartheta \mapsto K(\vartheta, \vartheta_0)$ has a strict minimum for $\vartheta = \vartheta_0$, and $U^{N,M}(\vartheta; \mathbb{X}^{N,M})$ tends in probability when $N, M \rightarrow \infty$ to $K(\vartheta, \vartheta_0)$.

Firstly, as under $\mathbb{P}_{\vartheta_0}^N$,

$$X_{t_j^M}^i - X_{t_{j-1}^M}^i = \int_{t_{j-1}^M}^{t_j^M} b_0(\vartheta_0; s, X_s^i, \mu_s^N) ds + \int_{t_{j-1}^M}^{t_j^M} \sigma(s, X_s^i) dB_s^i,$$

we have that $U^{N,M}(\vartheta; \mathbb{X}^{N,M})$ is also equal, under $\mathbb{P}_{\vartheta_0}^N$, to

$$\begin{aligned} & - \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M \left((c^{-1} b_0)(\vartheta; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N) - (c^{-1} b_0)(\vartheta_0; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N) \right)^\top \int_{t_{j-1}^M}^{t_j^M} \sigma(s, X_s^i) dB_s^i \\ & - \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M \left((c^{-1} b_0)(\vartheta; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N) - (c^{-1} b_0)(\vartheta_0; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N) \right)^\top \int_{t_{j-1}^M}^{t_j^M} b_0(\vartheta_0; s, X_s^i, \mu_s^N) ds \\ & + \frac{1}{2N} \sum_{i=1}^N \sum_{j=1}^M \left(\left| (c^{-1/2} b_0)(\vartheta; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N) \right|^2 - \left| (c^{-1/2} b_0)(\vartheta_0; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N) \right|^2 \right) \Delta_M, \end{aligned}$$

and simple calculations show that this quantity is also equal to $A_1^{N,M} + A_2^{N,M} + A_3^{N,M}$ where

$$\begin{aligned} A_1^{N,M} &:= - \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M \left((c^{-1} b_0)(\vartheta; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N) - (c^{-1} b_0)(\vartheta_0; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N) \right)^\top \\ & \quad \times \int_{t_{j-1}^M}^{t_j^M} \sigma(s, X_s^i) dB_s^i, \\ A_2^{N,M} &:= - \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M \left((c^{-1} b_0)(\vartheta; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N) - (c^{-1} b_0)(\vartheta_0; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N) \right)^\top \\ & \quad \int_{t_{j-1}^M}^{t_j^M} (b_0(\vartheta_0; s, X_s^i, \mu_s^N) - b_0(\vartheta_0; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N)) ds, \\ A_3^{N,M} &:= \frac{1}{2N} \sum_{i=1}^N \sum_{j=1}^M \left| (c^{-1/2} b_0)(\vartheta; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N) - (c^{-1/2} b_0)(\vartheta_0; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N) \right|^2 \Delta_M. \end{aligned}$$

Let us first show that $A_1^{N,M}$ and $A_2^{N,M}$ tend in probability to 0 under $\mathbb{P}_{\vartheta_0}^N$ as $N, M \rightarrow \infty$. Let m a positive integer. We rewrite $A_1^{N,M}$ as

$$- \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M f_i(t_{j-1}^M) \int_{t_{j-1}^M}^{t_j^M} g_i(s) dB_s^i,$$

where

$$\begin{aligned} f_i(s) &:= ((c^{-1}b_0)(\vartheta; s, X_s^i, \mu_s^N) - (c^{-1}b_0)(\vartheta_0; s, X_s^i, \mu_s^N))^\top, \\ g_i(s) &:= \sigma(s, X_s^i). \end{aligned}$$

We have

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} [|A_1^{N,M}|^m] &\lesssim \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} \left[\left| \frac{1}{N^2} \sum_{i=1}^N \int_0^T |g_i(s) \sum_{j=1}^M f_i(t_{j-1}^M) \mathbf{1}_{s \in [t_{j-1}^M, t_j^M)}|^2 ds \right|^m \right]^{1/2} \\ &\lesssim \frac{1}{N^{(m+1)/2}} \left(\sum_{i=1}^N \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} \left[\int_0^T |g_i(s) \sum_{j=1}^M f_i(t_{j-1}^M) \mathbf{1}_{s \in [t_{j-1}^M, t_j^M)}|^{2m} ds \right] \right)^{1/2} \\ &\lesssim \frac{|\vartheta - \vartheta_0|^m}{N^{m/2}} \sigma_+^m \left(1 + \sup_{t \in [0, T]} \max_{r=2r_1m, 2r_2m} \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} [\mathbf{m}_r(\mu_t^N)] \right)^{1/2} \\ &\lesssim \frac{|\vartheta - \vartheta_0|^m}{N^{m/2}} \end{aligned}$$

which tends to 0 as $N, M \rightarrow \infty$. Now we turn to $A_2^{N,M}$, we rewrite it as

$$-\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M f_i(t_{j-1}^M) \int_{t_{j-1}^M}^{t_j^M} h_i(t_{j-1}^M, s) ds,$$

where

$$h_i(t_{j-1}^M, s) := b_0(\vartheta_0; s, X_s^i, \mu_s^N) - b_0(\vartheta_0; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N).$$

We have

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} [|A_2^{N,M}|^m] &\lesssim \frac{1}{N} \sum_{i=1}^N \int_0^T \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} \left[\left| \sum_{j=1}^M f_i(t_{j-1}^M) h_i(t_{j-1}^M, s) \mathbf{1}_{s \in [t_{j-1}^M, t_j^M)} \right|^m \right] ds \\ &\lesssim \frac{1}{N} \sum_{i=1}^N \int_0^T \sum_{j=1}^M \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} [|f_i(t_{j-1}^M)|^m |h_i(t_{j-1}^M, s)|^m] \mathbf{1}_{s \in [t_{j-1}^M, t_j^M)} ds \\ &\lesssim \frac{|\vartheta - \vartheta_0|^m}{M^{m\gamma}} \end{aligned}$$

which tends to 0 as $N, M \rightarrow \infty$. Finally, let us show that the term $A_3^{N,M}$ tends in probability under $\mathbb{P}_{\vartheta_0}^N$ as $N, M \rightarrow \infty$ to our contrast function

$$K(\vartheta, \vartheta_0) := \frac{1}{2} \int_0^T \langle |(c^{-1/2}b_0)(\vartheta; t, \cdot, \mu_t^{\vartheta_0}) - (c^{-1/2}b_0)(\vartheta_0; t, \cdot, \mu_t^{\vartheta_0})|^2, \mu_t^{\vartheta_0} \rangle dt.$$

We use the notation

$$f(\vartheta, \vartheta_0; s, x, \nu) := (c^{-1/2}b_0)(\vartheta; s, x, \nu) - (c^{-1/2}b_0)(\vartheta_0; s, x, \nu).$$

We then have that $\mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} [|A_3^{N,M} - K(\vartheta, \vartheta_0)|]$ is equal to

$$\mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} \left[\left| \frac{1}{2} \sum_{j=1}^M \int_{t_{j-1}^M}^{t_j^M} \langle |f(\vartheta, \vartheta_0; t_{j-1}^M, \cdot, \mu_{t_{j-1}^M}^N)|^2, \mu_{t_{j-1}^M}^N \rangle ds - \frac{1}{2} \sum_{j=1}^M \int_{t_{j-1}^M}^{t_j^M} \langle |f(\vartheta, \vartheta_0; s, \cdot, \mu_s^{\vartheta_0})|^2, \mu_s^{\vartheta_0} \rangle ds \right| \right]$$

$$\begin{aligned}
 &\lesssim \frac{1}{2} \sum_{j=1}^M \int_{t_{j-1}^M}^{t_j^M} \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} [|f(\vartheta, \vartheta_0; s, X_s^i, \mu_s^N) - f(\vartheta, \vartheta_0; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N)|^2] ds \\
 &\quad + \frac{1}{2} \sum_{j=1}^M \int_{t_{j-1}^M}^{t_j^M} \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} [\langle |f(\vartheta, \vartheta_0; s, \cdot, \mu_s^N)|^2, \mu_s^N \rangle - \langle |f(\vartheta, \vartheta_0; s, \cdot, \mu_s^{\vartheta_0})|^2, \mu_s^{\vartheta_0} \rangle] ds \\
 &\lesssim \frac{1}{2} \sum_{j=1}^M \int_{t_{j-1}^M}^{t_j^M} \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} [|s - t_{j-1}^M|^{2\gamma} + |X_s^i - X_{t_{j-1}^M}^i|^2 + \mathcal{W}_1(\mu_s^N, \mu_{t_{j-1}^M}^N)^2] ds \\
 &\quad + \frac{1}{2} \sum_{j=1}^M \int_{t_{j-1}^M}^{t_j^M} \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} [\langle |f(\vartheta, \vartheta_0; s, \cdot, \mu_s^N)|^2, \mu_s^N \rangle - \langle |f(\vartheta, \vartheta_0; s, \cdot, \mu_s^{\vartheta_0})|^2, \mu_s^{\vartheta_0} \rangle] ds .
 \end{aligned}$$

As for the first part, we use the fact that for every $t, t' \in [0, T]$,

$$\mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} [\mathcal{W}_1(\mu_t^N, \mu_{t'}^N)] \leq \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} [|X_t^N - X_{t'}^N|] \lesssim |t - t'|^{1/2} ,$$

which gives that

$$\frac{1}{2} \sum_{j=1}^M \int_{t_{j-1}^M}^{t_j^M} \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} [|s - t_{j-1}^M|^{2\gamma} + |X_s^i - X_{t_{j-1}^M}^i|^2 + \mathcal{W}_1(\mu_s^N, \mu_{t_{j-1}^M}^N)^2] ds \lesssim \frac{1}{M^{2\gamma}} + \frac{1}{M} .$$

As for the second part, we apply Lemma 1.21 with the function $\phi(\vartheta; s, x, \nu) := |f(\vartheta, \vartheta_0; s, x, \nu)|^2$, which verifies

$$|\phi(\vartheta; s, x', \nu') - \phi(\vartheta; s, x, \nu)| \lesssim (|x' - x| + \mathcal{W}_1(\nu', \nu))(1 + |x|^{r_1} + |x'|^{r_1} + \mathbf{m}_{r_2}(\nu) + \mathbf{m}_{r_2}(\nu')) .$$

K is a non-negative function and $K(\vartheta, \vartheta_0)$ equals 0 if and only if

$$\bar{\mathbb{P}}_{\vartheta_0}(\forall t \in [0, T], b_0(\vartheta; t, X_t^1, \mu_t^{\vartheta_0}) = b_0(\vartheta_0; t, X_t^1, \mu_t^{\vartheta_0})) = 0 .$$

We can also rewrite $K(\vartheta, \vartheta_0)$ in the following form :

$$\begin{aligned}
 K(\vartheta, \vartheta_0) &= \frac{1}{2} \sum_{j=1}^d \int_0^T \int_{\mathbb{R}^d} \left| \int_0^1 \nabla_{\vartheta}(c^{-1/2} b_0)^j(\vartheta_0 + \lambda(\vartheta - \vartheta_0); t, x, \mu_t^{\vartheta_0})^\top (\vartheta - \vartheta_0) d\lambda \right|^2 \mu_t^{\vartheta_0}(dx) dt , \\
 &= \frac{1}{2} \langle \mathbb{I}(\vartheta_0, \vartheta)(\vartheta - \vartheta_0), \vartheta - \vartheta_0 \rangle
 \end{aligned}$$

where we have used the notation

$$\mathbb{I}(\vartheta_0, \vartheta) := \left(\sum_{j=1}^d \int_0^T \int_{\mathbb{R}^d} \nabla_{\vartheta}(c^{-1/2} b_0)^j([\vartheta_0, \vartheta]; t, x, \mu_t^{\vartheta}) \nabla_{\vartheta_0}(c^{-1/2} b_0)^j([\vartheta_0, \vartheta]; t, x, \mu_t^{\vartheta_0})^\top \mu_t^{\vartheta_0}(dx) dt \right) ,$$

and as the statistical experiment is supposed strongly non-degenerate in the sense of Definition 49, we know that $\mathbb{I}(\vartheta_0, \vartheta)$ is a definite positive matrix for every $\vartheta \in \Theta$ (such that $[\vartheta_0, \vartheta] \subset \Theta$). So $K(\vartheta, \vartheta_0)$ equals 0 if and only if $\vartheta = \vartheta_0$.

Finally we have to show that there exist some decreasing to 0 sequences $(\eta_k)_k, (\epsilon_k)_k$ such that for every k ,

$$\mathbb{P}_{\vartheta_0}^N \left(\sup_{|\vartheta - \vartheta'| \leq \eta_k} |U^{N,M}(\vartheta'; \mathbb{X}^{N,M}) - U^{N,M}(\vartheta; \mathbb{X}^{N,M})| > \epsilon_k \right) \xrightarrow{k \rightarrow \infty} 0.$$

In order to do so, as

$$\begin{aligned} & \mathbb{P}_{\vartheta_0}^N \left(\sup_{|\vartheta - \vartheta'| \leq \eta_k} |U^{N,M}(\vartheta'; \mathbb{X}^{N,M}) - U^{N,M}(\vartheta; \mathbb{X}^{N,M})| > \epsilon_k \right) \\ & \leq \frac{1}{\epsilon_k} \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} \left[\sup_{|\vartheta - \vartheta'| \leq \eta_k} |U^{N,M}(\vartheta'; \mathbb{X}^{N,M}) - U^{N,M}(\vartheta; \mathbb{X}^{N,M})| \right], \end{aligned}$$

we plan to use Kolmogorov's continuity theorem (see the Appendix of [98]). Let $\vartheta, \vartheta' \in \Theta$, and m a positive integer. We can use the previous results by replacing ϑ_0 by ϑ' in the definitions of $A_k^{N,M}$, $k = 1, 2, 3$. It remains to deal with the term $\mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} [|A_3^{N,M}|^m]$, and we easily have

$$\mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} [|A_3^{N,M}|^m] \lesssim \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} [|f_i(t_{j-1}^M)|^m] \lesssim |\vartheta' - \vartheta|^{2m}.$$

So, as Θ is compact,

$$\mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} [|U^{N,M}(\vartheta'; \mathbb{X}^{N,M}) - U^{N,M}(\vartheta; \mathbb{X}^{N,M})|^m] \lesssim \frac{|\vartheta' - \vartheta|^m}{N^{m/2}} + \frac{|\vartheta' - \vartheta|^m}{M^{m\gamma}} + |\vartheta' - \vartheta|^{2m} \lesssim |\vartheta' - \vartheta|^m,$$

and taking $m > d$, we have by Kolmogorov's theorem

$$\mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} \left[\sup_{|\vartheta - \vartheta'| \leq \eta_k} |U^{N,M}(\vartheta'; \mathbb{X}^{N,M}) - U^{N,M}(\vartheta; \mathbb{X}^{N,M})| \right] \lesssim \eta_k^{(m-d)/m}.$$

It remains to take, for instance, $m = 2d$, $\eta_k = k^{-1}$ and $\epsilon_k = \eta_k^{1/4}$ to conclude. It ends the proof of Theorem 4.1.

4.2 Asymptotic normality of the AMLE : proof of Theorem 60

We will follow Part 3.3.4. of [54]. First, for every y , $\vartheta \mapsto U^{N,M}(\vartheta, y)$ is two-times continuously differentiable on a neighbourhood of ϑ_0 by assumption on b . Now, let us show that

$$\sqrt{N} \nabla_{\vartheta} U^{N,M}(\vartheta_0, \mathbb{X}^{N,M}) \rightarrow \mathcal{N}_p(0, \mathbb{I}(\vartheta_0))$$

in law under $\mathbb{P}_{\vartheta_0}^N$ as $N, M \rightarrow \infty$. Firstly

$$\begin{aligned} \partial_{\vartheta_k} U^{N,M}(\vartheta, y) &= -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M \partial_{\vartheta_k} (c^{-1} b_0)(\vartheta; t_{j-1}^M, y_{j-1}^i, \mu^N(y_{j-1}))^\top (y_j^i - y_{j-1}^i) \\ &\quad + \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M \partial_{\vartheta_k} (c^{-1/2} b_0)(\vartheta; t_{j-1}^M, y_{j-1}^i, \mu^N(y_{j-1}))^\top (c^{-1/2} b_0)(\vartheta; t_{j-1}^M, y_{j-1}^i, \mu^N(y_{j-1})) \Delta_M. \end{aligned}$$

So, under $\mathbb{P}_{\vartheta_0}^N$, we have $\partial_{\vartheta_k} U^{N,M}(\vartheta_0, \mathbb{X}^{N,M}) = A_{4,k}^{N,M} + A_{5,k}^{N,M} + A_{6,k}^{N,M}$, where

$$\begin{aligned} A_{4,k}^{N,M} &:= -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M \partial_{\vartheta_k}(c^{-1}b_0)(\vartheta_0, t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N)^\top \\ &\quad \times \int_{t_{j-1}^M}^{t_j^M} (b_0(\vartheta_0; s, X_s^i, \mu_s^N) - b_0(\vartheta_0; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N)) ds, \\ A_{5,k}^{N,M} &:= -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M \int_{t_{j-1}^M}^{t_j^M} (\partial_{\vartheta_k}(c^{-1}b_0)(\vartheta_0; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N) - \partial_{\vartheta_k}(c^{-1}b_0)(\vartheta_0; s, X_s^i, \mu_s^N))^\top \\ &\quad \times \sigma(s, X_s^i) dB_s^i, \\ A_{6,k}^{N,M} &:= -\frac{1}{N} \sum_{i=1}^N \int_0^T \partial_{\vartheta_k}(c^{-1/2}b_0)(\vartheta_0; s, X_s^i, \mu_s^N)^\top dB_s^i. \end{aligned}$$

We treat the term $A_{4,k}^{N,M}$ with the same method as for the term $A_2^{N,M}$ in the preceding part, and we obtain

$$\mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} [|A_{4,k}^{N,M}|] \lesssim \frac{1}{M^\gamma}.$$

So, if $N^{1/2}M^{-\gamma} \rightarrow 0$ as $N, M \rightarrow \infty$, we obtain that $N^{1/2}A_4^{N,M} \rightarrow 0$ in $\mathbb{P}_{\vartheta_0}^N$ -probability as $N, M \rightarrow \infty$.

We treat the term $A_{5,k}^{N,M}$ with the same method as for the term $A_1^{N,M}$, and we obtain

$$\mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} [|A_{5,k}^{N,M}|^2] \lesssim \frac{1}{N} \left(\frac{1}{M^{2\gamma'}} + \frac{1}{M} \right),$$

and so $N^{1/2}A_5^{N,M} \rightarrow 0$ in $\mathbb{P}_{\vartheta_0}^N$ -probability as $N, M \rightarrow \infty$.

Finally, for the term $A_6^{N,M} = (A_{6,k}^{N,M})_{1 \leq k \leq p}^\top$, we have

$$A_6^{N,M} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^d \int_0^T \nabla_{\vartheta}(c^{-1/2}b)^j(\vartheta_0; s, X_s^i, \mu_s^N) d(B_s^i)^j,$$

and we already know that

$$\sqrt{N}A_6^{N,M} \rightarrow \mathcal{N}_p(0, \mathbb{I}(\vartheta_0))$$

in law under $\mathbb{P}_{\vartheta_0}^N$ as $N, M \rightarrow \infty$.

Now we will show that

$$D_{\vartheta}^2 U^{N,M}(\vartheta_0, \mathbb{X}^{N,M}) \rightarrow \mathbb{I}(\vartheta_0)$$

in $\mathbb{P}_{\vartheta_0}^N$ -probability as $N, M \rightarrow \infty$. Let $1 \leq k, l \leq p$, we have that $\partial_{\vartheta_l \vartheta_k}^2 U^{N,M}(\vartheta, y)$ is equal to

$$\begin{aligned} & -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M \partial_{\vartheta_l \vartheta_k}^2(c^{-1}b_0)(\vartheta; t_{j-1}^M, y_{j-1}^i, \mu_{j-1}^N(y))^\top (y_j^i - y_{j-1}^i) \\ & + \frac{\Delta_M}{N} \sum_{i=1}^N \sum_{j=1}^M \left(\partial_{\vartheta_l \vartheta_k}^2(c^{-1}b_0)(\vartheta; t_{j-1}^M, y_{j-1}^i, \mu_{j-1}^N(y))^\top b_0(\vartheta; t_{j-1}^M, y_{j-1}^i, \mu_{j-1}^N(y)) \right. \end{aligned}$$

$$+ \partial_{\vartheta_k}(c^{-1}b_0)(\vartheta; t_{j-1}^M, y_{j-1}^i, \mu_{j-1}^N(y))^\top \partial_{\vartheta_l} b_0(\vartheta; t_{j-1}^M, y_{j-1}^i, \mu_{j-1}^N(y)) \Big) .$$

So, under $\mathbb{P}_{\vartheta_0}^N$, $\partial_{\vartheta_l \vartheta_k}^2 U^{N,M}(\vartheta, \mathbb{X}^{N,M}) = A_7^{N,M} + A_8^{N,M} + A_9^{N,M} + A_{10}^{N,M}$, where

$$\begin{aligned} A_7^{N,M} &:= -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M \partial_{\vartheta_l \vartheta_k}^2 (c^{-1}b_0)(\vartheta; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N)^\top \\ &\quad \times \int_{t_{j-1}^M}^{t_j^M} (b_0(\vartheta_0; s, X_s^i, \mu_s^N) - b_0(\vartheta_0; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N)) ds , \\ A_8^{N,M} &:= -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M \partial_{\vartheta_l \vartheta_k}^2 (c^{-1}b_0)(\vartheta; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N)^\top \int_{t_{j-1}^M}^{t_j^M} \sigma(s, X_s^i) dB_s^i , \\ A_9^{N,M} &:= \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M \int_{t_{j-1}^M}^{t_j^M} \left(\partial_{\vartheta_k}(c^{-1/2}b_0)(\vartheta_0; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N)^\top \partial_{\vartheta_l}(c^{-1/2}b_0)(\vartheta_0; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N) \right. \\ &\quad \left. - \partial_{\vartheta_k}(c^{-1/2}b_0)(\vartheta_0; s, X_s^i, \mu_s^N)^\top \partial_{\vartheta_l}(c^{-1/2}b_0)(\vartheta_0; s, X_s^i, \mu_s^N) \right) ds , \\ A_{10}^{N,M} &:= \frac{1}{N} \sum_{i=1}^N \int_0^T \partial_{\vartheta_k}(c^{-1/2}b_0)(\vartheta_0; s, X_s^i, \mu_s^N)^\top \partial_{\vartheta_l}(c^{-1/2}b_0)(\vartheta_0; s, X_s^i, \mu_s^N) ds . \end{aligned}$$

In the same way as before, we show that, $A_7^{N,M}, A_8^{N,M}, A_9^{N,M}$ tend to 0 in $\mathbb{P}_{\vartheta_0}^N$ -probability as $N, M \rightarrow \infty$, and that

$$A_{10}^{N,M} \rightarrow \int_0^T \int_{\mathbb{R}^d} \partial_{\vartheta_k}(c^{-1/2}b_0)(\vartheta_0; s, x, \mu_s^{\vartheta_0})^\top \partial_{\vartheta_l}(c^{-1/2}b_0)(\vartheta_0; s, x, \mu_s^{\vartheta_0}) \mu_s^{\vartheta_0}(dx) ds = [\mathbb{I}(\vartheta_0)]_{kl}$$

in $\mathbb{P}_{\vartheta_0}^N$ -probability as $N, M \rightarrow \infty$.

Finally, we have to show that, for every $r > 0$ such that $\bar{B}(\vartheta_0, r) \subset \Theta$, we have, for some $m > d$,

$$\mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} \left[\sup_{|u| \leq r} |\partial_{\vartheta_l \vartheta_k}^2 U^{N,M}(\vartheta_0 + u, \mathbb{X}^{N,M}) - \partial_{\vartheta_l \vartheta_k}^2 U^{N,M}(\vartheta_0, \mathbb{X}^{N,M})| \right] \lesssim r^{1-d/m} .$$

In order to do so, we again make use of the Kolmogorov's theorem : let $\vartheta, \vartheta' \in \Theta$, and a positive integer m , we have, under $\mathbb{P}_{\vartheta_0}^N$,

$$\partial_{\vartheta_l \vartheta_k}^2 U^{N,M}(\vartheta', \mathbb{X}^{N,M}) - \partial_{\vartheta_l \vartheta_k}^2 U^{N,M}(\vartheta, \mathbb{X}^{N,M}) = A_{11}^{N,M} + A_{12}^{N,M} + A_{13}^{N,M} ,$$

where

$$\begin{aligned} A_{11}^{N,M} &:= -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M (\partial_{\vartheta_l \vartheta_k}^2 (c^{-1}b_0)(\vartheta'; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N) - \partial_{\vartheta_l \vartheta_k}^2 (c^{-1}b_0)(\vartheta; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N))^\top \\ &\quad \times \int_{t_{j-1}^M}^{t_j^M} (b_0(\vartheta_0; s, X_s^i, \mu_s^N) - b_0(\vartheta_0; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N)) ds , \end{aligned}$$

$$\begin{aligned}
 A_{11}^{N,M} &:= -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M (\partial_{\vartheta_i \vartheta_k}^2 (c^{-1} b_0)(\vartheta'; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N) - \partial_{\vartheta_i \vartheta_k}^2 (c^{-1} b_0)(\vartheta; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N))^\top \\
 &\quad \times \int_{t_{j-1}^M}^{t_j^M} \sigma(s, X_s^i) dB_s^i, \\
 A_{12}^{N,M} &:= \frac{\Delta_M}{N} \sum_{i=1}^N \sum_{j=1}^M (\partial_{\vartheta_k} (c^{-1/2} b_0)(\vartheta'; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N)^\top \partial_{\vartheta_i} (c^{-1/2} b_0)(\vartheta; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N) \\
 &\quad - \partial_{\vartheta_k} (c^{-1/2} b_0)(\vartheta; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N)^\top \partial_{\vartheta_i} (c^{-1/2} b_0)(\vartheta; t_{j-1}^M, X_{t_{j-1}^M}^i, \mu_{t_{j-1}^M}^N)) .
 \end{aligned}$$

And we obtain

$$\mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} \left[\left| \partial_{\vartheta_i \vartheta_k}^2 U^{N,M}(\vartheta', \mathbb{X}^{N,M}) - \partial_{\vartheta_i \vartheta_k}^2 U^{N,M}(\vartheta, \mathbb{X}^{N,M}) \right|^m \right] \lesssim |\vartheta' - \vartheta|^m ,$$

and so, by Kolmogorov's theorem, for $m > d$,

$$\mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} \left[\sup_{|u| \leq r} \left| \partial_{\vartheta_i \vartheta_k}^2 U^{N,M}(\vartheta_0 + u, \mathbb{X}^{N,M}) - \partial_{\vartheta_i \vartheta_k}^2 U^{N,M}(\vartheta_0, \mathbb{X}^{N,M}) \right| \right] \lesssim r^{1-d/m} .$$

Let us show briefly how to conclude the proof. We have

$$\begin{aligned}
 0 &= \sqrt{N} \nabla_{\vartheta} U^{N,M}(\tilde{\vartheta}^{N,M}, \mathbb{X}^{N,M}) = \sqrt{N} \nabla_{\vartheta} U^{N,M}(\vartheta_0, \mathbb{X}^{N,M}) \\
 &\quad + \int_0^1 D_{\tilde{\vartheta}}^2 U^{N,M}(\vartheta_0 + \lambda(\tilde{\vartheta}^{N,M} - \vartheta_0), \mathbb{X}^{N,M}) d\lambda \times \sqrt{N}(\tilde{\vartheta}^{N,M} - \vartheta_0) .
 \end{aligned}$$

So, as $\nabla_{\vartheta} U^{N,M}(\vartheta_0, \mathbb{X}^{N,M}) = 0$, if, for large enough N, M , $\int_0^1 D_{\tilde{\vartheta}}^2 U^{N,M}(\vartheta_0 + \lambda(\tilde{\vartheta}^{N,M} - \vartheta_0), \mathbb{X}^{N,M}) d\lambda$ is invertible, we have

$$\sqrt{N}(\tilde{\vartheta}^{N,M} - \vartheta_0) = - \left(\int_0^1 D_{\tilde{\vartheta}}^2 U^{N,M}(\vartheta_0 + \lambda(\tilde{\vartheta}^{N,M} - \vartheta_0), \mathbb{X}^{N,M}) d\lambda \right)^{-1} \times \sqrt{N} \nabla_{\vartheta} U^{N,M}(\vartheta_0, \mathbb{X}^{N,M}) .$$

We have already shown that

$$\sqrt{N} \nabla_{\vartheta} U^{N,M}(\vartheta_0, \mathbb{X}^{N,M}) \rightarrow \mathcal{N}_p(0, \mathbb{I}(\vartheta_0))$$

in law under $\mathbb{P}_{\vartheta_0}^N$ as $N, M \rightarrow \infty$, it then remains to show that

$$\int_0^1 D_{\tilde{\vartheta}}^2 U^{N,M}(\vartheta_0 + \lambda(\tilde{\vartheta}^{N,M} - \vartheta_0), \mathbb{X}^{N,M}) d\lambda \rightarrow \mathbb{I}(\vartheta_0)$$

in $\mathbb{P}_{\vartheta_0}^N$ -probability as $N, M \rightarrow \infty$. As $\mathbb{I}(\vartheta_0)$ is invertible, we will then have for large enough N, M , $\int_0^1 D_{\tilde{\vartheta}}^2 U^{N,M}(\vartheta_0 + \lambda(\tilde{\vartheta}^{N,M} - \vartheta_0), \mathbb{X}^{N,M}) d\lambda$ invertible $\mathbb{P}_{\vartheta_0}^N$ -almost surely.

We already know that

$$D_{\tilde{\vartheta}}^2 U^{N,M}(\vartheta_0, \mathbb{X}^{N,M}) \rightarrow \mathbb{I}(\vartheta_0)$$

in $\mathbb{P}_{\vartheta_0}^N$ -probability as $N, M \rightarrow \infty$, and it just remains to prove that

$$\int_0^1 V^{N,M}(\vartheta_0, \tilde{\vartheta}^{N,M}, \lambda) d\lambda \rightarrow 0$$

in $\mathbb{P}_{\vartheta_0}^N$ -probability as $N, M \rightarrow \infty$, where we have defined

$$V^{N,M}(\vartheta_0, \vartheta, \lambda) := D_{\vartheta}^2 U^{N,M}(\vartheta_0 + \lambda(\vartheta - \vartheta_0), \mathbb{X}^{N,M}) - D_{\vartheta}^2 U^{N,M}(\vartheta_0, \mathbb{X}^{N,M}).$$

We have, for every $r > 0$,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} \left[\left| \int_0^1 V^{N,M}(\vartheta_0, \tilde{\vartheta}^{N,M}, \lambda) d\lambda \right| \right] &\leq \int_0^1 \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} [|V^{N,M}(\vartheta_0, \tilde{\vartheta}^{N,M}, \lambda)| \mathbf{1}_{|\tilde{\vartheta}^{N,M} - \vartheta_0| \leq r}] d\lambda \\ &\quad + \int_0^1 \mathbb{E}_{\mathbb{P}_{\vartheta_0}^N} [|V^{N,M}(\vartheta_0, \tilde{\vartheta}^{N,M}, \lambda)| \mathbf{1}_{|\tilde{\vartheta}^{N,M} - \vartheta_0| > r}] d\lambda \\ &\lesssim r^{1-d/m} + \mathbb{P}_{\vartheta_0}^N(|\tilde{\vartheta}^{N,M} - \vartheta_0| > r). \end{aligned}$$

So for every $\epsilon > 0$, taking $r^{1-d/m} = \epsilon/2$, as $|\tilde{\vartheta}^{N,M} - \vartheta_0| \rightarrow 0$ in $\mathbb{P}_{\vartheta_0}^N$ -probability as $N, M \rightarrow \infty$, for large enough N, M , $\mathbb{P}_{\vartheta_0}^N(|\tilde{\vartheta}^{N,M} - \vartheta_0| > r) \leq \epsilon/2$, which ends the proof of Theorem 60.

BIBLIOGRAPHIE

- [1] Kweku Abraham and Richard Nickl. On statistical Calderón problems. Mathematical Statistics and Learning, 2(2) :165–216, 2020.
- [2] Yacine Aït-Sahalia and Jean Jacod. High-frequency financial econometrics. In High-Frequency Financial Econometrics. Princeton University Press, 2014.
- [3] Chiara Amorino, Akram Heidari, Vytautė Pilipauskaitė, and Mark Podolskij. Parameter estimation of discretely observed interacting particle systems. arXiv preprint arXiv :2208.11965, 2022.
- [4] Luciana Angiuli and Luca Lorenzi. Compactness and invariance properties of evolution operators associated with Kolmogorov operators with unbounded coefficients. Journal of mathematical analysis and applications, 379(1) :125–149, 2011.
- [5] Donald G. Aronson and Piotr Besala. Parabolic equations with unbounded coefficients. Journal of Differential Equations, 3(1) :1–14, 1967.
- [6] Javier Baladron, Diego Fasoli, Olivier Faugeras, and Jonathan Touboul. Mean-field description and propagation of chaos in networks of Hodgkin–Huxley and FitzHugh–Nagumo neurons. Journal of Mathematical Neurosciences, pages 2 :Art 10,50, 2012.
- [7] Martin T. Barlow and Marc Yor. Semimartingale inequalities via the Garsia–Rodemich–Rumsey lemma, and applications to local times. J. Functional Analysis, 49(2) :198–229, 1982.
- [8] Denis Belomestny, Vytautė Pilipauskaitė, and Mark Podolskij. Semiparametric estimation of McKean–Vlasov SDEs. arXiv preprint arXiv :2107.00539, 2021.
- [9] Denis Belomestny and John Schoenmakers. Projected particle methods for solving McKean–Vlasov stochastic differential equations. SIAM Journal on Numerical Analysis, 56(6) :3169–3195, 2018.
- [10] Gérard Ben Arous and Marc Brunaud. Methode de Laplace : etude variationnelle des fluctuations de diffusions de type champ moyen. Communications in Statistics-Simulation and Computation, 31(1-4) :79–144, 1990.
- [11] Gérard Ben Arous and Ofer Zeitouni. Increasing propagation of chaos for mean field models. In Annales de l’Institut Henri Poincaré (B) Probability and Statistics, volume 35, pages 85–102. Elsevier, 1999.

- [12] Said Benachour, Bernard Roynette, Denis Talay, and Pierre Vallois. Nonlinear self-stabilizing processes—i existence, invariant probability, propagation of chaos. Stochastic processes and their applications, 75(2) :173–201, 1998.
- [13] Oumaima Bencheikh and Benjamin Jourdain. Bias behaviour and antithetic sampling in mean-field particle approximations of SDEs nonlinear in the sense of McKean. ESAIM : Proceedings and Surveys, 65 :219–235, 2019.
- [14] Rudolf Beran. The role of Hájek’s convolution theorem in statistical theory. Kybernetika, 31(3) :221–237, 1995.
- [15] Lucien Birgé and Pascal Massart. Minimum contrast estimators on sieves : exponential bounds and rates of convergence. Bernoulli, 4(3) :329–375, 1998.
- [16] Jaya P.N. Bishwal. Estimation in interacting diffusions : Continuous and discrete sampling. Applied Mathematics, 2(9) :1154–1158, 2011.
- [17] Vladimir I. Bogachev, Nicolai V. Krylov, Michael Röckner, and Stanislav V. Shaposhnikov. Fokker–Planck–Kolmogorov Equations, volume 207. American Mathematical Society, 2022.
- [18] Vladimir I. Bogachev, Stanislav V. Shaposhnikov, and Alexander Y. Veretennikov. Differentiability of solutions of stationary Fokker–Planck–Kolmogorov equations with respect to a parameter. Discrete & Continuous Dynamical Systems, 36(7) :3519, 2016.
- [19] François Bolley, José A. Canizo, and José A. Carrillo. Stochastic mean-field limit : non-Lipschitz forces and swarming. Mathematical Models and Methods in Applied Sciences, 21(11) :2179–2210, 2011.
- [20] François Bolley, Ivan Gentil, and Arnaud Guillin. Uniform convergence to equilibrium for granular media. Archive for Rational Mechanics and Analysis, 208(2) :429–445, 2013.
- [21] François Bolley, Arnaud Guillin, and Cédric Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. Probability Theory and Related Fields, 137(3-4) :541–593, 2007.
- [22] Ludwig Boltzmann. Weitere studien über das wärmeleichgewicht unter gasmolekülen. In Kinetische Theorie II, pages 115–225. Springer, 1970.
- [23] Mattia Bongini, Massimo Fornasier, Markus Hansen, and Mauro Maggioni. Inferring interaction rules from observations of evolutive systems I : The variational approach. Mathematical Models and Methods in Applied Sciences, 27(05) :909–951, 2017.
- [24] Mireille Bossy. Some stochastic particle methods for nonlinear parabolic PDEs. In ESAIM : proceedings, volume 15, pages 18–57. EDP Sciences, 2005.
- [25] Mireille Bossy and Denis Talay. A stochastic particle method for some one-dimensional nonlinear pde. Mathematics and computers in simulation, 38(1-3) :43–50, 1995.
- [26] Mireille Bossy and Denis Talay. A stochastic particle method for the McKean–Vlasov and the Burgers equation. Mathematics of computation, 66(217) :157–192, 1997.
- [27] Alexandre Boumezoued, Marc Hoffmann, and Paulien Jeunesse. Nonparametric adaptive inference of birth and death models in a large population limit. Mathematical Statistics and Learning, 3(1) :1–69, 2021.
- [28] David R. Brillinger. Learning a potential function from a trajectory. In Selected works of David Brillinger, pages 361–364. Springer, 2012.

- [29] David R. Brillinger, Haiganoush K. Preisler, Alan A. Ager, and J.G. Kie. The use of potential functions in modelling animal movement. In Selected Works of David Brillinger, pages 385–409. Springer, 2012.
- [30] David R. Brillinger, H.K. Preisler, and M.J. Wisdom. Modelling particles moving in a potential field with pairwise interactions and an application. Brazilian Journal of Probability and Statistics, 25(3) :421–436, 2011.
- [31] Valerii V. Buldygin and Yuriy V. Kozačenko. Sub-Gaussian random variables. Ukrainian Mathematical Journal, 32(6) :483–489, 1980.
- [32] Martin Burger, Vincezo Capasso, and Daniela Morale. On an aggregation model with long and short range interactions. Nonlinear Analysis and Real World Applications, 8(3) :939–958, 2007.
- [33] Claudio Canuto, Fabio Fagnani, and Paolo Tilli. An Eulerian approach to the analysis of Krause’s consensus models. SIAM Journal on Control and Optimization, 50(1) :243–265, 2012.
- [34] Pierre Cardaliaguet. Notes on mean field games. Technical report.
- [35] Pierre Cardaliaguet, François Delarue, Jean-Michel Lasry, and Pierre-Louis Lions. The master equation and the convergence problem in mean field games, volume 201 of Annals of Mathematics Studies. Princeton University Press, Princeton, NJ, 2019.
- [36] Pierre Cardaliaguet and Charles Lehalle. Mean Field Game of Controls and An Application To Trade Crowding. Mathematics and Financial Economics, 12(3) :335–363, 2019.
- [37] René Carmona, François Delarue, et al. Probabilistic Theory of Mean Field Games with Applications I-II. Springer, 2018.
- [38] Patrick Cattiaux, Arnaud Guillin, and Florent Malrieu. Probabilistic approach for granular media equations in the non-uniformly convex case. Probability theory and related fields, 140(1) :19–40, 2008.
- [39] Louis-Pierre Chaintron and Antoine Diez. Propagation of chaos : a review of models, methods and applications. I. Models and methods. arXiv preprint arXiv :2203.00446, 2022.
- [40] Louis-Pierre Chaintron and Antoine Diez. Propagation of chaos : a review of models, methods and applications. II. Applications. arXiv preprint arXiv :2106.14812, 2022.
- [41] Jean-François Chassagneux, Lukasz Szpruch, and Alvin Tse. Weak quantitative propagation of chaos via differential calculus on the space of measures. The Annals of Applied Probability, 32(3) :1929–1969, 2022.
- [42] Jean-François Chassagneux, Lukasz Szpruch, and Alvin Tse. Weak quantitative propagation of chaos via differential calculus on the space of measures. The Annals of Applied Probability, 32(3) :1929–1969, 2022.
- [43] Bernard Chazelle, Quansen Jiu, Qianxiao Li, and Chu Wang. Well-posedness of the limiting equation of a noisy consensus model in opinion dynamics. Journal of Differential Equations, 263(1) :365 – 397, 2017.
- [44] Xiaohui Chen. Maximum likelihood estimation of potential energy in interacting particle systems from single-trajectory data. Electronic Communications in Probability, 26 :1–13, 2021.
- [45] Julien Chevallier. Approximation par champ-moyen : le couplage à la Sznitman pour les nuls. 2017.

- [46] Michele Coghi, Jean-Dominique Deuschel, Peter K. Friz, and Mario Maurelli. Pathwise McKean–Vlasov theory. 2018.
- [47] Fabienne Comte. Estimation non-paramétrique. Spartacus-Idh, 2017.
- [48] Fabienne Comte and Valentine Genon-Catalot. Nonparametric drift estimation for iid paths of stochastic differential equations. The Annals of Statistics, 48(6) :3336–3365, 2020.
- [49] Fabienne Comte and Nicolas Marie. Nonparametric Drift Estimation from Diffusions with Correlated Brownian Motions. 2022.
- [50] Harald Cramér. Mathematical methods of Statistics. Department of Mathematical SU, 1946.
- [51] Dan Crisan and Eamon McMurray. Smoothing properties of McKean–Vlasov SDEs. Probability Theory and Related Fields, 171(1) :97–148, 2018.
- [52] Dan Crisan and Eamon McMurray. Cubature on Wiener space for McKean–Vlasov SDEs with smooth scalar interaction. The Annals of Applied Probability, 29(1) :130–177, 2019.
- [53] Imre Csiszár. Sanov property, generalized I–projection and a conditional limit theorem. The Annals of Probability, pages 768–793, 1984.
- [54] Didier Dacunha-Castelle and Marie Duflo. Probability and Statistics : Volume II, volume 2. Springer Science & Business Media, 2012.
- [55] Donald A. Dawson. Critical dynamics and fluctuations for a mean–field model of cooperative behavior. Journal of Statistical Physics, 31(1) :29–85, 1983.
- [56] Donald A. Dawson and Jürgen Gärtner. Large deviations from the McKean–Vlasov limit for weakly interacting diffusions. Stochastics : An International Journal of Probability and Stochastic Processes, 20(4) :247–308, 1987.
- [57] Donald A. Dawson and Jürgen Gärtner. Large deviations, free energy functional and quasi–potential for a mean field model of interacting diffusions, volume 78. American Mathematical Soc., 1989.
- [58] P-E Chaudru de Raynal, J-F Jabir, and S Menozzi. Multidimensional Stable driven McKean–Vlasov SDEs with distributional interaction kernel : a regularization by noise perspective. arXiv preprint arXiv :2205.11866, 2022.
- [59] Paul-Eric Chaudru de Raynal and Noufel Frikha. From the backward Kolmogorov PDE on the Wasserstein space to propagation of chaos for McKean–Vlasov SDEs. Journal de Mathématiques Pures et Appliquées, 156 :1–124, 2021.
- [60] Paul-Eric Chaudru de Raynal and Noufel Frikha. From the backward Kolmogorov PDE on the Wasserstein space to propagation of chaos for McKean–Vlasov SDEs. Journal de Mathématiques Pures et Appliquées, 156 :1–124, 2021.
- [61] Paul-Eric Chaudru de Raynal and Noufel Frikha. Well–posedness for some non–linear SDEs and related PDE on the Wasserstein space. Journal de Mathématiques Pures et Appliquées, 159 :1–167, 2022.
- [62] PE Chaudru de Raynal. Strong well posedness of McKean–Vlasov stochastic differential equations with hölder drift. Stochastic Processes and their Applications, 130(1) :79–107, 2020.
- [63] Thomas Deck and Susanne Kruse. Parabolic differential equations with unbounded coefficients—a generalization of the parametrix method. Acta Applicandae Mathematica, 74(1) :71–91, 2002.

- [64] Laetitia Della Maestra and Marc Hoffmann. The LAN property for McKean–Vlasov models in a mean–field regime. arXiv preprint arXiv :2205.05932, 2022.
- [65] Laetitia Della Maestra and Marc Hoffmann. Nonparametric estimation for interacting particle systems : McKean–Vlasov models. Probability Theory and Related Fields, 182(1) :551–613, 2022.
- [66] Christophe Denis, Charlotte Dion-Blanc, and Miguel Martinez. A ridge estimator of the drift from discrete repeated observations of the solution of a stochastic differential equation. Bernoulli, 27(4) :2675–2713, 2021.
- [67] Marie Doumic, Marc Hoffmann, Nathalie Krell, and Lydia Robert. Statistical estimation of a growth–fragmentation model observed on a genealogical tree. Bernoulli, 21(3) :1760–1799, 2015.
- [68] Marie Doumic, Marc Hoffmann, Patricia Reynaud-Bouret, and Vincent Rivoirard. Nonparametric estimation of the division rate of a size–structured population. SIAM Journal on Numerical Analysis, 50(2) :925–950, 2012.
- [69] Paul Ehrenfest and Tatiana Ehrenfest. The conceptual foundations of the statistical approach in mechanics. Courier Corporation, 1990.
- [70] Lawrence C. Evans. Partial differential equations, volume 19. American Mathematical Soc., 2010.
- [71] Begoña Fernandez and Sylvie Méléard. A Hilbertian approach for fluctuations on the McKean–Vlasov model. Stochastic Process. Appl., 71(1) :33–53, 1997.
- [72] William Finnoff. Law of large numbers for a general system of stochastic differential equations with global interaction. Stochastic Processes and their Applications, 46(1) :153–182, 1993.
- [73] Jean-Pierre Fouque and Li-Hsien Sun. Systemic risk illustrated. Handbook on Systemic Risk, Eds J.P. Fouque and J. Langsam, 2013.
- [74] Nicolas Fournier and Arnaud Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. Probab. Theory Related Fields, 162(3-4) :707–738, 2015.
- [75] Avner Friedman. Partial differential equations of parabolic type. Courier Dover Publications, 2008.
- [76] Tadahisa Funaki. A certain class of diffusion processes associated with nonlinear parabolic equations. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 67(3) :331–348, 1984.
- [77] Angela Ganz Bustos. Approximations des distributions d’équilibre de certains systèmes stochastiques avec interactions McKean–Vlasov. PhD thesis, Nice, 2008.
- [78] Jürgen Gärtner. On the McKean–Vlasov limit for interacting diffusions. Math. Nachr., 137 :197–248, 1988.
- [79] Valentine Genon-Catalot and Jean Jacod. On the estimation of the diffusion coefficient for multi–dimensional diffusion processes. In Annales de l’IHP Probabilités et Statistiques, volume 29, pages 119–151, 1993.
- [80] Valentine Genon-Catalot and Jean Jacod. Estimation of the diffusion coefficient for diffusion processes : random sampling. Scand. J. Statist., 21(3) :193–221, 1994.
- [81] Valentine Genon-Catalot and Catherine Laredo. Parametric inference for small variance and long time horizon McKean–Vlasov diffusion models. Electronic Journal of Statistics, 15(2) :5811–5854, 2021.

- [82] Valentine Genon-Catalot and Catherine Laredo. Probabilistic properties and parametric inference of small variance nonlinear self-stabilizing stochastic differential equations. Stochastic Processes and their Applications, 142 :513–548, 2021.
- [83] Valentine Genon-Catalot and Dominique Picard. Eléments de statistique asymptotique. Springer, 1993.
- [84] Kay Giesecke, Gustavo Schwenkler, and Justin A. Sirignano. Inference for large financial systems. Mathematical Finance, 30(1) :3–46, 2020.
- [85] Emmanuel Gobet. Monte–Carlo methods and stochastic processes : from linear to non–linear. Chapman and Hall/CRC, 2016.
- [86] Emmanuel Gobet and Stefano Pagliarani. Analytical approximations of non–linear SDEs of McKean–Vlasov type. Journal of Mathematical Analysis and Applications, 466(1) :71–106, 2018.
- [87] Alexander Goldenshluger and Oleg Lepski. Universal pointwise selection rule in multivariate function estimation. Bernoulli, 14(4) :1150–1190, 2008.
- [88] Alexander Goldenshluger and Oleg Lepski. Bandwidth selection in kernel density estimation : oracle inequalities and adaptive minimax optimality. Ann. Statist., 39(3) :1608–1632, 2011.
- [89] Alexander Goldenshluger and Oleg Lepski. On adaptive minimax density estimation on R^d . Probab. Theory Related Fields, 159(3-4) :479–543, 2014.
- [90] Jaroslav Hájek. Local asymptotic minimax and admissibility in estimation. In Proceedings of the sixth Berkeley symposium on Mathematical Statistics and Probability, volume 1, pages 175–194, 1972.
- [91] Yi Han. Smoothness of the density for McKean–Vlasov SDEs with measurable kernel. arXiv preprint arXiv :2208.02771, 2022.
- [92] Samuel Herrmann, Peter Imkeller, and Dierk Peithmann. Large deviations and a kramers, Å type law for self-stabilizing diffusions. The Annals of Applied Probability, 18(4) :1379–1423, 2008.
- [93] Masuyuki Hitsuda and Itaru Mitoma. Tightness problem and stochastic evolution equation arising from fluctuation phenomena for interacting diffusions. Journal of Multivariate Analysis, 19(2) :311–328, 1986.
- [94] Van Ha Hoang, Thanh Mai Pham Ngoc, Vincent Rivoirard, and Viet Chi Tran. Nonparametric estimation of the fragmentation kernel based on a partial differential equation stationary distribution approximation. Scandinavian Journal of Statistics, 49(1) :4–43, 2022.
- [95] Marc Hoffmann and Adélaïde Olivier. Nonparametric estimation of the division rate of an age dependent branching process. Stochastic Process. Appl., 126(5) :1433–1471, 2016.
- [96] Minyi Huang, Roland P. Malhamé, and Peter E. Caines. Large population stochastic dynamic games : closed–loop McKean–Vlasov systems and the Nash certainty equivalence principle. Communications in Information & Systems, 6(3) :221–252, 2006.
- [97] Martin Hutzenhaler, Arnulf Jentzen, and Peter E. Kloeden. Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients. The Annals of Applied Probability, 22(4) :1611–1641, 2012.
- [98] Ildar A. Ibragimov and Rafail Z. Hasminskii. Statistical estimation : asymptotic theory, volume 16. Springer Science & Business Media, 2013.

- [99] Pierre-Emmanuel Jabin and Zhenfu Wang. Mean field limit for stochastic particle systems. In Active Particles, Volume 1, pages 379–402. Springer, 2017.
- [100] Jean-Francois Jabir. Rate of propagation of chaos for diffusive stochastic particle systems via Girsanov transformation. arXiv preprint arXiv :1907.09096, 2019.
- [101] Jean Jacod and Philip Protter. Discretization of processes, volume 67. Springer Science & Business Media, 2011.
- [102] Jean Jacod and Albert N. Shiryaev. Limit theorems for stochastic processes, volume 288 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, second edition, 2003.
- [103] James H. Jeans. On the theory of star-streaming and the structure of the universe. Monthly Notices of the Royal Astronomical Society, 76 :70–84, 1915.
- [104] James H. Jeans. Problems of cosmogony and stellar dynamics. University press, 1919.
- [105] Paulien Jeunesse. Non Parametric Estimation of General Population Process Parameters. PhD thesis, Université Paris sciences et lettres, 2019.
- [106] Jan Johannes. Deconvolution with unknown error distribution. Ann. Statist., 37(5A) :2301–2323, 2009.
- [107] Benjamin Jourdain and Sylvie Méléard. Propagation of chaos and fluctuations for a moderate model with smooth initial data. In Annales de l’Institut Henri Poincaré (B) Probability and Statistics, volume 34, pages 727–766. Elsevier, 1998.
- [108] Benjamin Jourdain and Alvin Tse. Central limit theorem over non-linear functionals of empirical measures with applications to the mean-field fluctuation of interacting diffusions. Electronic Journal of Probability, 26 :1–34, 2021.
- [109] Mark Kac. Foundations of kinetic theory. In Proceedings of The third Berkeley symposium on mathematical statistics and probability, volume 3, pages 171–197, 1956.
- [110] Ioannis Karatzas and Steven Shreve. Brownian motion and stochastic calculus, volume 113. Springer Science & Business Media, 2012.
- [111] Raphael A. Kasonga. Maximum likelihood theory for large interacting systems. SIAM Journal on Applied Mathematics, 50(3) :865–875, 1990.
- [112] Arturo Kohatsu-Higa and Shigeyoshi Ogawa. Weak rate of convergence for an Euler scheme of nonlinear SDE, Ås. 1997.
- [113] Nikolai V. Krylov. Lectures on elliptic and parabolic equations in Hölder spaces. Number 12. American Mathematical Soc., 1996.
- [114] Shigeo Kusuoka and Yozo Tamura. Gibbs measures for mean field potentials. J. Fac. Sci. Univ. Tokyo Sect. IA Math, 31(1) :223–245, 1984.
- [115] Daniel Lacker. Mean field games and interacting particle systems. Preprint, 2018.
- [116] Daniel Lacker. On a strong form of propagation of chaos for McKean-Vlasov equations. Electron. Commun. Probab., 23 :Paper No. 45, 11, 2018.
- [117] Claire Lacour, Pascal Massart, and Vincent Rivoirard. Estimator selection : a new method with applications to kernel density estimation. Sankhya A, 79(2) :298–335, 2017.
- [118] O.A. Ladyzhenskaya, V.A. Solonnikov, and N.N. UralCeva. Linear and quasilinear equations of parabolic type, Translations of Mathematical Monographs. 23 (1968). American Mathematical Society, Providence RI.

- [119] Quanjun Lang and Fei Lu. Learning interaction kernels in mean-field equations of 1st-order systems of interacting particles. arXiv preprint arXiv :2010.15694, 2020.
- [120] Quanjun Lang and Fei Lu. Identifiability of interaction kernels in mean-field equations of interacting particles. arXiv preprint arXiv :2106.05565, 2021.
- [121] Jean-Michel Lasry and Pierre-Louis Lions. Jeux à champ moyen. I–le cas stationnaire. Comptes Rendus Mathématique, 343(9) :619–625, 2006.
- [122] Jean-Michel Lasry and Pierre-Louis Lions. Jeux à champ moyen. II–Horizon fini et contrôle optimal. Comptes Rendus Mathématique, 343(10) :679–684, 2006.
- [123] Lucien Le Cam. Asymptotic methods in statistical decision theory. Springer Series in Statistics. Springer–Verlag, New York, 1986.
- [124] Christian Léonard. Sur la limite en loi et les fluctuations de certains modèles dynamiques d’interaction. PhD thesis, 1984.
- [125] Christian Léonard. Une loi des grands nombres pour des systèmes de diffusions avec interaction et à coefficients non bornés. In Annales de l’IHP Probabilités et Statistiques, volume 22, pages 237–262, 1986.
- [126] Christian Léonard. Large deviations and law of large numbers for a mean field type interacting particle systems. Stochastic Processes and their Applications, 25 :215–235, 1987.
- [127] Oleg V. Lepski. A problem of adaptive estimation in Gaussian white noise. Teor. Veroyatnost. i Primenen., 35(3) :459–470, 1990.
- [128] Eugenio Elia Levi. Sulle equazioni lineari totalmente ellittiche alle derivate parziali. Rendiconti del Circolo Matematico di Palermo (1884-1940), 24(1) :275–317, 1907.
- [129] Zhongyang Li and Fei Lu. On the coercivity condition in the learning of interacting particle systems. arXiv preprint arXiv :2011.10480, 2020.
- [130] Zhongyang Li, Fei Lu, Mauro Maggioni, Sui Tang, and Cheng Zhang. On the identifiability of interaction functions in systems of interacting particles. Stochastic Processes and their Applications, 132 :135–163, 2021.
- [131] Robert S. Liptser and Albert N. Shiryaev. Statistics of random processes : General theory, volume 394. Springer, 1977.
- [132] Meiqi Liu and Huijie Qiao. Parameter estimation of path-dependent McKean–Vlasov stochastic differential equations. Acta Mathematica Scientia, 42(3) :876–886, 2022.
- [133] Yating Liu. Optimal Quantization : Limit Theorem, Clustering and Simulation of the McKean–Vlasov Equation. PhD thesis, Sorbonne université, 2019.
- [134] Eva Löcherbach. LAN and LAMN for systems of interacting diffusions with branching and immigration. In Annales de l’IHP Probabilités et Statistiques, volume 38, pages 59–90, 2002.
- [135] Mark G. Low. Nonexistence of an adaptive estimator for the value of an unknown probability density. Ann. Statist., 20(1) :598–602, 1992.
- [136] Fei Lu, Mauro Maggioni, and Sui Tang. Learning interaction kernels in heterogeneous systems of agents from multiple trajectories. J. Mach. Learn. Res., 22 :32–1, 2021.
- [137] Fei Lu, Mauro Maggioni, and Sui Tang. Learning interaction kernels in stochastic systems of interacting particles from multiple trajectories. Foundations of Computational Mathematics, 22(4) :1013–1067, 2022.

- [138] Fei Lu, Ming Zhong, Sui Tang, and Mauro Maggioni. Nonparametric inference of interaction laws in systems of agents from trajectory data. Proceedings of the National Academy of Sciences, 116(29) :14424–14433, 2019.
- [139] Alessandra Lunardi. Analytic semigroups and optimal regularity in parabolic problems. Springer Science & Business Media, 2012.
- [140] Mauro Maggioni, Jason Miller, Hongda Qiu, and Ming Zhong. Learning Interaction Kernels for Agent Systems on Riemannian Manifolds. arXiv preprint arXiv :2102.00327, 2021.
- [141] Mylène Maïda, Tien Dat Nguyen, Thanh Mai Pham Ngoc, Vincent Rivoirard, and Viet Chi Tran. Statistical deconvolution of the free fokker-planck equation at fixed time. Bernoulli, 28(2) :771–802, 2022.
- [142] Florent Malrieu. Logarithmic Sobolev inequalities for some nonlinear PDE’s. Stochastic Process. Appl., 95(1) :109–132, 2001.
- [143] Oxana Manita and Stanislav Shaposhnikov. Nonlinear parabolic equations for measures. St. Petersburg Mathematical Journal, 25(1) :43–62, 2014.
- [144] Nicolas Marie and Amélie Rosier. Nadaraya–Watson Estimator for IID Paths of Diffusion Processes. Scandinavian Journal of Statistics, 2021.
- [145] Pascal Massart. Concentration inequalities and model selection, volume 6. Springer, 2007.
- [146] Henry P. McKean. A class of Markov processes associated with nonlinear parabolic equations. Proceedings of the National Academy of Sciences of the United States of America, 56(6) :1907, 1966.
- [147] Henry P. McKean. Propagation of chaos for a class of non-linear parabolic equations. Stochastic Differential Equations (Lecture Series in Differential Equations, Session 7, Catholic Univ., 1967), pages 41–57, 1967.
- [148] Sylvie Méléard. Asymptotic behaviour of some interacting particle systems : McKean–Vlasov and Boltzmann models. In Probabilistic models for nonlinear partial differential equations, pages 42–95. Springer, 1996.
- [149] Sylvie Méléard and Sylvie Roelly-Coppoletta. A propagation of chaos result for a system of particles with moderate interaction. Stochastic Processes and their Applications, 26 :317–332, 1987.
- [150] Sylvie Méléard and Sylvie Roelly-Coppoletta. Systèmes de particules et mesures–martingales : un théorème de propagation du chaos. Séminaire de probabilités de Strasbourg, 22 :438–448, 1988.
- [151] Jason Miller, Sui Tang, Ming Zhong, and Mauro Maggioni. Learning Theory for Inferring Interaction Kernels in Second-Order Interacting Agent Systems. arXiv preprint arXiv :2010.03729, 2020.
- [152] Stéphane Mischler. Sur le programme de Kac concernant les limites de champ moyen. Séminaire Équations aux dérivées partielles (Polytechnique) dit aussi” Séminaire Goulaouic-Schwartz”, pages 1–19, 2010.
- [153] Itaru Mitoma. Generalized Ornstein–Uhlenbeck process having a characteristic operator with polynomial coefficients. Probability theory and related fields, 76(4) :533–555, 1987.
- [154] Alexander Mogilner and Leah Edelstein-Keshet. A non-local model for a swarm. Journal of Mathematical Biology, 38(6) :534–570, 1999.

- [155] François Monard, Richard Nickl, and Gabriel P. Paternain. Consistent Inversion of Noisy Non-Abelian X-Ray Transforms. Communications on Pure and Applied Mathematics, 74(5) :1045–1099, 2021.
- [156] Pierre Luc Morien. Autour des équations aux dérivées partielles stochastiques : systèmes en interaction et problème non-linéaire, régularité et approximation des densités. PhD thesis, Paris 6, 1996.
- [157] Elizbar A. Nadaraya. On a regression estimate. Teor. Verojatnost. i Primenen., 9 :157–159, 1964.
- [158] Masao Nagasawa and Hiroshi Tanaka. On the propagation of chaos for diffusion processes with drift coefficients not of average form. Tokyo journal of mathematics, 10(2) :403–418, 1987.
- [159] Richard Nickl. On Bayesian inference for some statistical inverse problems with partial differential equations. Bernoulli News, 24(2) :5–9, 2017.
- [160] Richard Nickl. Bernstein–von Mises theorems for statistical inverse problems I : Schrödinger equation. Journal of the European Mathematical Society, 22(8) :2697–2750, 2020.
- [161] Karl Oelschläger. A martingale approach to the law of large numbers for weakly interacting stochastic processes. The Annals of Probability, pages 458–479, 1984.
- [162] Karl Oelschläger. A law of large numbers for moderately interacting diffusion processes. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 69(2) :279–322, 1985.
- [163] Emanuel Parzen. On estimation of a probability density function and mode. The Annals of Mathematical Statistics, 33(3) :1065–1076, 1962.
- [164] Panpan Ren and Jiang-Lun Wu. Least squares estimator for path-dependent McKean–Vlasov SDEs via discrete-time observations. Acta Mathematica Scientia, 39(3) :691–716, 2019.
- [165] Daniel Revuz and Marc Yor. Continuous martingales and Brownian motion, volume 293 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, third edition, 1999.
- [166] Murray Rosenblatt. Remarks on some nonparametric estimates of a density function. The Annals of Mathematical Statistics, pages 832–837, 1956.
- [167] Thomas J. Rothenberg. Identification in parametric models. Econometrica : Journal of the Econometric Society, pages 577–591, 1971.
- [168] Michael Scheutzow. Uniqueness and non-uniqueness of solutions of Vlasov–McKean equations. Journal of the Australian Mathematical Society, 43(2) :246–256, 1987.
- [169] Louis Sharrock, Nikolas Kantas, Panos Parpas, and Grigorios A. Pavliotis. Parameter Estimation for the McKean–Vlasov Stochastic Differential Equation. arXiv preprint arXiv :2106.13751, 2021.
- [170] Tokuzo Shiga and Hiroshi Tanaka. Central limit theorem for a system of Markovian particles with mean field interactions. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 69(3) :439–459, 1985.
- [171] Daniel W. Stroock and S.R. Srinivasa Varadhan. Multidimensional diffusion processes, volume 233. Springer Science & Business Media, 1997.
- [172] Alain-Sol Sznitman. Nonlinear reflecting diffusion process, and the propagation of chaos and fluctuations associated. J. Funct. Anal., 56(3) :311–336, 1984.

- [173] Alain-Sol Sznitman. Topics in propagation of chaos. In Ecole d'été de Probabilités de Saint-Flour XIX–1989, pages 165–251. Springer, 1991.
- [174] Lukasz Szpruch, Shuren Tan, and Alvin Tse. Iterative multilevel particle approximation for McKean–Vlasov SDEs. The Annals of Applied Probability, 29(4) :2230–2265, 2019.
- [175] Lukasz Szpruch and Alvin Tse. Antithetic multilevel sampling method for nonlinear functionals of measure. The Annals of Applied Probability, 31(3) :1100–1139, 2021.
- [176] Denis Talay and Olivier Vaillant. A stochastic particle method with random weights for the computation of statistical solutions of McKean–Vlasov equations. The Annals of Applied Probability, 13(1) :140–180, 2003.
- [177] Yozo Tamura. Free energy and the convergence of distributions of diffusion processes of McKean type. Journal of the Faculty of Science, the University of Tokyo. Sect. 1 A, Mathematics, 34(2) :443–484, 1987.
- [178] Hiroshi Tanaka. Limit theorems for certain diffusion processes with interaction. In North–Holland Mathematical Library, volume 32, pages 469–488. Elsevier, 1984.
- [179] Hiroshi Tanaka and Masuyuki Hitsuda. Central limit theorem for a simple diffusion model of interacting particles. Hiroshima Mathematical Journal, 11(2) :415–423, 1981.
- [180] Alvin Tse. Higher order regularity of nonlinear Fokker–Planck PDEs with respect to the measure component. Journal de Mathématiques Pures et Appliquées, 150 :134–180, 2021.
- [181] Edison Tse. Information matrix and local identifiability of parameters. In Joint Automatic Control Conference, number 11, pages 611–619, 1973.
- [182] Alexandre B. Tsybakov. Introduction to Nonparametric Estimation. In Springer Series in Statistics, 2009.
- [183] Suzanne Varet, Claire Lacour, Pascal Massart, and Vincent Rivoirard. Numerical performance of Penalized Comparison to Overfitting for multivariate kernel density estimation. arXiv preprint arXiv :1902.01075, 2019.
- [184] Cédric Villani. Limite de champ moyen. Cours de DEA, 2002 :49, 2001.
- [185] Cédric Villani. Optimal transport : old and new, volume 338. Springer, 2009.
- [186] Anatoly A. Vlasov. On the kinetic theory of an assembly of particles with collective interaction. Russ. Phys. J., 9 :25–40, 1945.
- [187] Anatoly A. Vlasov. The vibrational properties of an electron gas. Soviet Physics Uspekhi, 10(6) :721, 1968.
- [188] Pierre Weiss. L'hypothèse du champ moléculaire et la propriété ferromagnétique. J. Phys. Theor. Appl., 6(1) :661–690, 1907.
- [189] Jianghui Wen, Xiangjun Wang, Shuhua Mao, and Xinpeng Xiao. Maximum likelihood estimation of McKean–Vlasov stochastic differential equation and its application. Applied Mathematics and Computation, 274 :237–246, 2016.
- [190] Ming Zhong, Jason Miller, and Mauro Maggioni. Data-driven discovery of emergent behaviors in collective dynamics. Physica D : Nonlinear Phenomena, 411 :132542, 2020.

RÉSUMÉ

Cette thèse se situe à l'interface de la Statistique, paramétrique et non-paramétrique, des processus stochastiques, et de l'analyse des équations aux dérivées partielles paraboliques. Le fil conducteur des travaux présentés ici a été d'obtenir des informations sur le flot de mesures de probabilités, solution d'une équation de Fokker-Planck-Kolmogorov non-linéaire, dite équation de McKean-Vlasov, et sur le coefficient de drift de cette équation, à l'aide de l'observation continue de la solution du système linéaire de diffusions en interaction l'approchant en limite de champ moyen, *i.e.* quand le nombre de particules de ce système croît vers l'infini.

Dans une première partie, nous présentons les résultats obtenus quand le coefficient de drift dépend d'un paramètre inconnu. Nous suivons le programme d'estimation de Ibragimov et Hasminskii : nous montrons que la propriété de normalité locale asymptotique est vérifiée par ce modèle et nous en déduisons des propriétés fines de convergence et de normalité asymptotique de l'estimateur du maximum de vraisemblance de ce paramètre. Nous proposons un nouveau critère de vérification de la non-dégénérescence et de l'identifiabilité de ce problème statistique.

Dans une deuxième partie, nous présentons les résultats obtenus quand le coefficient de drift est inconnu, mais appartient à une certaine classe de fonctions caractérisées par leur régularité : les fonctions localement hôlderiennes. Cette partie repose sur une nouvelle inégalité de concentration de type Bernstein, et sur des techniques de statistique non-paramétrique : nous construisons des estimateurs à noyaux adaptatifs de la densité solution de l'équation de McKean-Vlasov et de son drift à l'aide de la méthode de Goldenshluger-Lepski, et nous montrons que ces estimateurs sont optimaux au sens minimax pour un risque ponctuel. Enfin, nous construisons un estimateur de la force d'interaction dans le cas d'un drift de type Vlasov à l'aide d'une méthode de déconvolution.

MOTS CLÉS

Statistique paramétrique/non-paramétrique, Statistique des processus aléatoires, modèle de McKean-Vlasov, Systèmes de particules en interaction, Limite de Champ moyen, Propagation du chaos, Inégalités de concentration, Equation aux dérivées partielles paraboliques non-linéaires, Equation de Fokker-Planck-Kolmogorov non-linéaire, Espaces de Hölder, Estimation minimax adaptative, Normalité locale asymptotique, Programme d'Ibragimov-Hasminski, Estimateur du maximum de vraisemblance, Inégalités oracles, Estimateurs à noyaux, Sélection de fenêtres, Méthode de Goldenshluger-Lepski, Déconvolution.

ABSTRACT

This thesis lies at the crossroads of non-parametric and parametric Statistics of processes, and parabolic partial differential equations analysis. The aim of the present work was to obtain informations about the flow of probability measures solution of a nonlinear Fokker-Planck-Kolmogorov equation, known as McKean-Vlasov equation, and about the drift coefficient of this equation, thanks to the continuous observation of the solution of a linear system of interacting diffusions as an approximation in a mean-field limit, *i.e.* as the number of particles in this system grows to infinity.

In a first part, we present our results when the drift coefficient depends on an unknown parameter. We follow the estimation program of Ibragimov and Hasminskii : we show that our model verifies the local asymptotic normality property, and we deduce from it precise properties of convergence and asymptotic normality of the maximum likelihood estimator of the parameter. We give a new criterium for verifying that this statistical problem is non-degenerate and identifiable.

In a second part, we present our results when the drift coefficient is unknown, but lies in a certain class of functions characterized by their regularity : the local Hölder functions. This part relies on a new Bernstein type concentration inequality, and on nonparametric statistics tools : we build adaptive kernel estimators of the density solution of the McKean-Vlasov equation and of its drift, and we show that these estimators are optimal in a minimax sense for a pointwise risk. Finally, we construct an estimator of the interaction force in the case of a Vlasov type drift by means of a deconvolution method.

KEYWORDS

Nonparametric statistics, Parametric statistics, Statistics of stochastic processes, McKean-Vlasov model, Interacting particle systems, Mean-field limit, Propagation of chaos, Concentration inequalities, Nonlinear Parabolic Partial differential equation, Nonlinear Fokker-Planck-Kolmogorov equation, Adaptive minimax estimation, Local Asymptotic Normality, Ibragimov-Hasminski's program, Maximum likelihood estimator, Oracle inequality, Kernel estimation, Bandwidth selection, Goldenshluger-Lepski's method, Deconvolution.