Bal(l)ade entre EDP et probabilités

Paul-Éric Chaudru de Raynal

Habilitation à Diriger des Recherches - HDR

spécialité: Mathématiques appliquées

présentée et soutenue par

Paul-Eric CHAUDRU de RAYNAL
Laboratoire de Mathématiques Jean Leray UMR CNRS 6629

BAL(L)ADE ENTRE EDP ET PROBABILITÉS

---

le 30 mai 2024

après avis de

Franco FLANDOLI	Professor	Scuola Normale Superiore, Pisa
Kavita RAMANAN	Professor	Brown University
Nizar TOUZI	Professor	New York University

devant le jury composé de

Président :	François BOLLEY	Professeur des Universités	ENS Rennes
Examinateurs :	Philippe BRIAND	Professeur des Universités	Université Savoie Mont Blanc
	François DELARUE	Professeur des Universités	Université Côte d'Azur
	Franco FLANDOLI	Professor	Scuola Normale Superiore, Pisa
	Arnaud GUILLIN	Professeur des Universités	Université Clermont Auvergne
	Frédéric HERAU	Professeur des Universités	Nantes Université
	Kavita RAMANAN	Professor	Brown University
	Nizar TOUZI	Professor	New York University

Ballade *[English: Ballad] nom féminin (ancien provençal ballada, de ballar, danser). (i) Au Moyen Âge, poème lyrique d'origine chorégraphique, d'abord chanté, puis destiné seulement à la récitation. (ii) À partir du xive siècle, poème à forme fixe, composé de trois strophes suivies d'un envoi d'une demi-strophe. (iii) Poème narratif mis à la mode en Allemagne et en Angleterre à la fin du xviie siècle, relatant de façon pathétique une tradition historique ou légendaire. (iv) Pièce vocale ou instrumentale inspirée par une ballade littéraire (Chopin, Liszt, Brahms ont écrit des ballades pour piano).*

Balade *[English: Walk / Stroll] nom féminin (de balader) Familier. Promenade : Faire une balade dans les bois. Synonymes : excursion - marche - promenade - sortie - tour - virée*

💬 ***Note de l'auteur :*** *Nous sommes le 1er août 2022, Stéphane vient de répondre à l'éditeur afin de valider les épreuves de notre travail, en collaboration avec Igor, sur le caractère bien posé au sens fort de "la chaîne" avec seuils de régularité (semble-t-il) quasi minimaux sur les dérives. Ceci répond à la question que m'avait posée François, il y a 12 ans maintenant. Ma thèse est définitivement finie, il est temps de tourner la page et de commencer la rédaction de mon mémoire d'habilitation. Je créé un fichier HDR.tex et tape ces lignes. Advienne que pourra.*

Remerciements

Dix années se sont écoulées depuis la rédaction de mon manuscrit de thèse et ce mémoire vise à “conter” les grandes lignes de mes recherches depuis. Le métier de chercheur est souvent un métier de passion, dans une certaine mesure d’obsession, mais surtout de discussions et de partage. De fait, la démarcation entre vie professionnelle et privée est parfois ténue : les interactions que j’ai pu avoir durant cette période, professionnelles et/ou plus personnelles ont énormément compté. Je remercie donc toutes les personnes, qu’elles soient de la communauté scientifique ou d’ailleurs, avec qui j’ai pu échanger ces dix dernières années.

L’histoire commence avec François Delarue. François a accepté de m’initier aux subtilités de l’analyse stochastique il y a quatorze ans maintenant. Dès lors, il a fait preuve de bienveillance envers moi et s’est toujours enquis de l’orientation et de l’avancée de mes travaux. J’ai notamment pu, à plusieurs reprises, bénéficier de son extraordinaire culture scientifique. Qu’un chercheur aussi brillant et généreux ait non seulement accepté de guider mes premiers pas, mais aussi gardé un oeil sur l’avancée de mes recherches est une véritable chance. Merci pour tout.

Franco Flandoli has honored me by agreeing to review this manuscript. Franco’s work on regularization by noise has undeniably had a profound impact on my research. That Franco has accepted to evaluate this work in return means a lot to me. Thank you Franco.

Kavita Ramanan has also agreed to review this thesis. The fact that such a specialist in probability theory has agreed to review and comment on my work is an honor. It is her remarkable work on reflected processes and large deviations that motivated me to contact her. Thank you Kavita.

Nizar Touzi a lui aussi accepté de rapporter ces travaux. Les recherches de Nizar, à l’interaction de la théorie des processus, des EDP et de leurs applications à l’économie et la finance notamment, ont un impact majeur au sein de notre communauté. Pouvoir bénéficier de l’avis d’un tel spécialiste est évidemment une chance. Merci Nizar.

François Bolley a accepté de participer à ce jury. Depuis son arrivée à l’ENS de Rennes, François a toujours bien voulu prendre un peu de son temps pour que nous puissions discuter ou encadrer des étudiants. J’ai ainsi pu bénéficier de ses nombreuses connaissances et de sa générosité à de multiples reprises. J’espère pouvoir continuer ce partage dans le futur. Merci François.

Philippe Briand a indéniablement beaucoup compté pour mon intégration dans la communauté scientifique et l’épanouissement de mes recherches. Fraîchement diplômé de mon doctorat, Philippe m’a accueilli au LAMA à Chambéry et m’a très rapidement proposé une collaboration sur un sujet riche et passionnant. J’ai énormément appris à son contact, et sa générosité et sa bienveillance

m'ont permis, entre autres, de tenir six ans "à cheval" entre Rennes et Chambéry. Merci Philippe.

J'ai rencontré Arnaud Guillin grâce au premier travail que Philippe m'avait proposé. J'ai été frappé qu'un chercheur aussi reconnu et occupé prenne toujours le temps d'échanger avec l'ensemble des membres de la communauté, qu'ils soient jeunes chercheurs ou chercheurs chevronnés. Surtout de façon aussi accessible et sympathique. Je suis très heureux qu'un tel spécialiste des processus de McKean-Vlasov, entre autres, accepte d'évaluer mon travail. Merci Arnaud.

Finalement, Frédéric Hérau m'a fait l'honneur d'accepter de participer à ce jury d'habilitation. Frédéric est un chercheur renommé en analyse et ses appétences scientifiques l'amène à considérer, entre autres, des systèmes cinétiques et de type McKean-Vlasov, tous deux présents dans ce mémoire d'habilitation. Pouvoir bénéficier des commentaires d'un tel chercheur est d'une grande valeur. Merci Frédéric.

Ces travaux doivent beaucoup aux personnes qui ont accepté de collaborer avec moi. Ce fut toujours dans une ambiance de travail excellente et je les en remercie. Merci en particulier à Noufel, pour nos travaux sur les équations de McKean-Vlasov; à Ying, pour notre collaboration sur les processus avec contrainte faible et son accueil lors de ma délégation à Rennes; à Pierre, pour notre collaboration et nos échanges. Plus particulièrement encore, j'en profite pour remercier très chaleureusement Stéphane, avec qui j'ai pu partager, entre autres, "l'ascension de la chaîne". J'ai énormément appris à ses côtés, dans une atmosphère particulièrement agréable. Et ce sont des euphémismes. Merci pour tout Stéphane. J'espère pouvoir continuer à bénéficier, à l'avenir, de ta grande culture scientifique, et surtout de ton amitié.

Je profite de ces lignes pour remercier tous les collègues du LAMA, à Chambéry, où j'ai trouvé un environnement de travail propice à ma jeune carrière. Notamment : Céline, Laure, Marguerite, Jimmy, Georges, Sébastien et Didier. Je remercie aussi les collègues qui m'ont accueilli à l'IUT. Notamment : Pascale, Vincent, Julien, Christophe, encore Julien et Sylvain.

Je tiens aussi à remercier chaleureusement tous mes collègues du LMJL, à Nantes. Notamment Nicolas et Philippe, amis probabilistes et du RU, Frédéric et Anaïs, ex et présent co-bureau. Merci à Béatrice, pour ton aide pour la coordination entre autres. Le LMJL offre un cadre de travail particulièrement agréable et stimulant. C'est une véritable opportunité pour moi d'avoir pu y continuer mon parcours.

C'est aussi l'occasion de remercier les membres de l'IRMAR, notamment de l'équipe Probabilités. J'ai eu la chance de bénéficier par deux fois d'une délégation CNRS de six mois au sein de ce laboratoire, ainsi que d'un accueil récurrent à compter de ma première délégation lors de mes séances de travail rennaises. L'IRMAR m'a offert un cadre de travail extrêmement plaisant.

Je me permets d'évoquer encore trois collègues dont le soutien et la générosité ont indiscutablement contribué à ce manuscrit. Mihaï m'a donné l'opportunité d'encadrer ma première thèse avec lui. Pour ta gentillesse, ta confiance, ta prévenance et tous nos échanges, merci Mihaï. Thomas a choisi de nous faire confiance, à Mihaï et moi, pour superviser sa thèse de doctorat. Ce fut un très grand plaisir. J'espère que nos chemins continueront de se croiser. Merci Thomas. Enfin, Julian m'a invité deux fois à le rejoindre sur des projets ANR. Les échanges qui en ont résulté ont chaque fois été riches et stimulants. Merci Julian pour ta confiance et les projets à venir.

À nouveau, la frontière entre vie professionnelle et personnelle est ténue. Je ne saurais clore cette partie sans adresser mes plus sincères remerciements à toute ma famille. Merci à vous tous pour toutes les belles choses partagées, passées et à venir.

Merci à Salima, Inès et Neïl d'égayer chaque jour de ma vie, et de la rendre juste merveilleuse.

Merci Maman. Tu m'avais fait promettre de te communiquer ce manuscrit.. le voilà. Tu as planté la graine des mathématiques que voici, j'aurai tant aimé que tu puisses les voir.

Présentation du manuscrit

Le présent manuscrit vise à retracer les travaux effectués depuis la fin de la préparation de mon doctorat (2010-2013 - sous la direction de F. Delarue) à aujourd'hui. Il couvre ainsi la période d' ATER (2013-2014) effectué à L'université Nice Côte d'Azur, de Maître de Conférence à l'Université Savoie Mont Blanc (2014-2020) puis à Nantes Université (2020-).

Le choix du titre n'est pas si anodin: d'une part le manuscrit est écrit dans l'espoir de proposer au lecteur une *balade* de l'analyse classique à l'analyse stochastique, au gré des questions rencontrées et de certaines solutions proposées; d'autre part il aspire (dans une certaine mesure !) à être une *ballade*, les différents objets et concepts manipulés recélant (dans la limite de ce que mes compétences me permettent d'en comprendre et d'en exposer) une part de beauté. Surtout, la démarche permet d'égayer un peu la rédaction du mémoire. De fait, le manuscrit est écrit de façon *linéaire*. Des concepts, questions et raisonnements introduits en *amont* de la lecture seront régulièrement réutilisés en *aval*.

Le manuscrit comprend ainsi 6 chapitres, un interlude et un bestiaire et se structure en deux parties : chapitres 1 et 2, interlude et chapitre 3, puis viennent les chapitres 4, 5 et 6. Les chapitres 1 et 4 ne contiennent aucun résultat personnel et ont vocation à introduire chacune des deux parties. Le bestiaire contient des définitions d'espaces fonctionnels utilisés dans le manuscrit.

La première partie est relative aux EDS *linéaires* (au sens de McKean-Vlasov). Le second chapitre, l'interlude et le troisième chapitre portent sur des travaux liés à la *régularisation par le bruit*, selon la terminologie de F. Flandoli dans ces notes cours à Saint Flour (2010) et introduite dans le chapitre 1. Ils exposent les résultats des travaux [4, 6, 7, 8] et [13] de la bibliographie personnelle, respectivement.

La seconde partie est liée aux EDS *non-linéaires* (au sens de McKean-Vlasov), telles qu'introduites dans le chapitre 4 . Le chapitre 5 porte sur les travaux [9, 12] et le chapitre 6 sur les travaux [5, 11].

Overview of the manuscript

The present manuscript aims to trace the work carried out from the end of my PhD preparation (2010-2013 - under the supervision of F. Delarue) up until today. It covers the period of my ATER (2013-2014) at the University of Nice Côte d'Azur, my position as Maître de Conférences at the University Savoie Mont Blanc (2014-2020), and my current position at Nantes University (2020-).

The title of this manuscript, "Bal(l)ade between Probability and PDE", was deliberately chosen to take the reader on a journey from classical analysis to stochastic analysis, following the questions explored and solutions proposed: this is the *balade*. Also, the manuscript aims to convey more than just technical exposition (to some extent!); the objects and concepts manipulated possess aesthetic qualities (to the best of my ability to comprehend and present them), which is the *ballade*. Especially, this approach adds an element of liveliness and interest to the writing of the manuscript. Therefore, the manuscript is written in a *linear* way. Concepts, questions, and reasoning introduced earlier in the text will be regularly reused downstream.

The manuscript consists of 6 chapters, an interlude and a Bestiary and is structured into two parts: Chapters 1 and 2, interlude, and Chapter 3, followed by Chapters 4, 5, and 6. Chapters 1 and 4 contain no personal results and aim to introduce each of the two parts. The bestiary contains definitions of functional spaces used in the manuscript.

The first part is related to *linear* SDEs (in the sense of McKean-Vlasov). The second chapter, the interlude, and the third chapter deal with work related to *regularization by noise*, according to the terminology of F. Flandoli in his lecture notes at Saint Flour (2010) and introduced in Chapter 1. They present the results of the work [4, 6, 7, 8] and [13] from the personal bibliography, respectively.

The second part is related to *non-linear* SDEs (in the sense of McKean-Vlasov), introduced in Chapter 4. Chapter 5 deals with the work [9, 12], and Chapter 6 deals with the work [5, 11].

Bibliographie personnelle - Personal bibliography

Nota. Les travaux présentés pour l'habilitation sont accompagnées d'un symbole 👉, les travaux en "transparents" sont ceux de la thèse de doctorat - The works presented for the thesis are accompanied by a symbol 👉, the works in "transparencies" are those of the PhD thesis.

[1] P.-É. Chaudru de Raynal. Strong existence and uniqueness for stochastic differential equation with Hölder drift and degenerate noise. Annales de l'Institut Henri Poincaré (B). Version pdf.

[2] P.-É. Chaudru de Raynal et C.A. Garcia Trillos. A cubature based algorithm to solve decoupled McKean-Vlasov Forward Backward Stochastic Differential Equations. Stochastic Processes and their Applications. Version pdf.

[3] 👉 P.-É. Chaudru de Raynal Strong well-posedness of McKean-Vlasov SDE with Hölder drift. Stochastic Processes and their Applications. Version pdf.

[4] 👉 P.-É. Chaudru de Raynal. Weak regularization by stochastic drift: result and counter-example. Discrete and Continuous Dynamical Systems (Series A). Version pdf.

[5] 👉 Ph. Briand, P.-É. Chaudru de Raynal, A. Guillin et C. Labart. Particles Systems and Numerical Schemes for Mean Reflected Stochastic Differential Equations. Annals of Applied Probability. Version pdf.

[6] 👉 P.-É. Chaudru de Raynal et S. Menozzi. Regularization effects of a noise propagating through a chain of differential equations: an almost sharp result.Transaction of the American Mathematical Society. Version pdf.

[7] 👉 P.-É. Chaudru de Raynal, I. Honoré et S. Menozzi. Sharp Schauder estimate for some degenerate Kolmogorov equation. Annali della scuola normale superior di Pisa. Version pdf.

[8] 👉 P.-É. Chaudru de Raynal, I. Honoré et S. Menozzi. Strong regularization by Brownian noise propagating through weak Hörmander structure. Probability Theory and Related Fields. Version pdf.

[9] 👉 P.-É. Chaudru de Raynal et N. Frikha. Well-posedness for some non-linear diffusion processes and related PDE on the Wasserstein space. Journal de Mathématiques pures et appliquées. Version pdf.

[10] 👉 P.-É. Chaudru de Raynal, S. Menozzi et E. Priola. Schauder estimates for drifted fractional operators in the supercritical case. Journal of Functional Analysis. Version pdf.

[11] 👉 Ph. Briand, P. Cardaliaguet, P.-É. Chaudru de Raynal et Y. Hu. Forward and Backward Stochastic Differential Equations with Normal Constraints in Law. Stochastic Processes and their Applications. Version pdf.

[12] 👉 P.-É. Chaudru de Raynal et N. Frikha. From the backward Kolmogorov equation on Wasserstein space to Propagation of Chaos for McKean-Vlasov SDEs. Journal de Mathématiques pures et appliquées. Version pdf.

[13] 👉 P.-É. Chaudru de Raynal et S. Menozzi. Multidimensional Stable driven SDEs with Besov drift. Electronic Journal of Probability. Version pdf.

[14] P.-É. Chaudru de Raynal, S. Menozzi et E. Priola. Weak well-posedness of multidimensional stable driven SDE in the critical case. Stochastics and Dynamics. Version pdf.

[15] M. H. Duong, P. Monmarché, M. Tomasevic and J. Tugaut. Reducing exit-times of diffusions with repulsive interactions. ESAIM: Probability and Statistics. Version pdf.

[16] P.-É. Chaudru de Raynal, S. Menozzi, A. Pesce et X. Zhang. Heat kernel and gradient estimates for kinetic SDEs with low regularity coefficients. Bulletin des Sciences Mathématiques. Version pdf.

[17] P.-É. Chaudru de Raynal, J.-F. Jabir et S. Menozzi. Multidimensional Stable driven McKean-Vlasov SDEs with distributional interaction kernel - a regularization by noise perspective. Stochastics and Partial Differential Equations: Analysis and Computations. Version pdf.

[18] P.-É. Chaudru de Raynal, J.-F. Jabir et S. Menozzi. Multidimensional Stable driven McKean-Vlasov SDEs with distributional interaction kernel - critical thresholds and related models. Soumis / Submitted Version pdf.

Chapter 1

Well-posedness of SDE

💬 *Throughout the manuscript, we define functional spaces on the fly, sometimes without explicitly specifying the ambient space when it seems understandable to us. This is done in order to enhance readability. At the end of the manuscript, the reader will find a brief bestiary of the spaces used, see 6.5.*

1.1 Regularization by noise

Since the 70's, many works have shown that Stochastic differential equations with coefficients that are less regular than Lipschitz can be well-posed in either a weak or strong sense. This phenomenon is known as regularization by noise. Indeed, ODEs are SDEs and it is therefore impossible to solve SDEs in full generality outside the Cauchy Lipschitz framework without taking advantage of the noise, see *e.g.* the celebrated work of Di Perna and Lions [DL89]. In this vein, we refer to the pioneering works of Zvonkin and Veretennikov: [Zvo74], [Ver80] for strong regularization, *i.e.* well-posedness holds in a strong sense and Stroock and Varadhan's [SV79] for weak regularization, *i.e.* well-posedness holds in weak sense, both outside the Lipschitz framework. Consider indeed the SDE with measurable coefficients $F, \sigma : \mathbf{R}_+ \times \mathbf{R}^d \to \mathbf{R}^d \times \mathbf{R}^{d\otimes d}$:

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

where B denotes a d-dimensional Brownian motion. If $\sigma = 0$ the SDE may be ill-posed as soon as F is less than Lipschitz. However, if σ is assumed to be uniformly non degenerate, the result of Zvonkin and Veretennikov tells us that the SDE is well-posed in a strong sense as soon as σ is Lipschitz in space and F only bounded and measurable. From Stroock and Varadhan's theory, we know that the Lipschitz condition on the coefficient σ may be relaxed to continuity for weak well-posedness to hold. In both cases, σ is assumed to be uniformly non degenerate, which justifies again that this is a regularization by noise phenomenon.

From both perspectives, the well-posedness of the SDE relies on obtaining a "good" theory for the (at least formally) associated PDE, *i.e.* driven by the generator of the Markov process that solves the SDE. The terminology of "good" depends on the type of well-posedness we are looking for. From the formal discussion below, we will see that a heuristic rule with vague contours seems to emerge for what "good" means: to obtain weak well-posedness, a sufficient condition is to control the gradient of the associated PDE in supremum norm while to obtain strong well-posedness, a sufficient condition is to control the second order derivatives of the PDE in a suitable Lebesgue

norm.

If we focus on the drift part of the equation, we can roughly describe the regularization phenomenon by considering the following framework: given a noise $(\mathcal{W}_t)_{t\geq 0}$ with fluctuations of order t^γ, $\gamma > 0$, what kind of minimal regularity $C_b^\beta(\mathbf{R}^d), \beta \leq 1$ must[1][2] be assumed on the coefficient F of the formal equation[3]

$$dY_t = F(t, Y_t)dt + \sigma d\mathcal{W}_t, \quad Y_0 = y \in \mathbf{R}^d, \tag{SDE}$$

with $\sigma > 0$, while preserving the well-posedness of the system in the weak or strong sense? To illustrate this, we will mainly consider the case $F : x \mapsto \mathrm{sign}(x)|x|^\beta$ in the following sections (Peano-like drift).

1.1.1 Weak regularization

The probabilistic viewpoint From a probabilistic viewpoint, an answer to the weak regularization phenomenon has been addressed in the work [DF14] by Delarue and Flandoli[4]: we have to compare the fluctuations of the noise $\mathcal{W}$ and of one of the extremal solutions of the deterministic version of the Peano equation associated with the SDE *i.e.*

$$dY_t = \mathrm{sign}(Y_t)|Y_t|^\beta dt, \quad Y_0 = 0.$$

To regularize the equation, the noise has to dominate the system in small time so that the solution leaves the singularity of the initial configuration. This competition can be made explicit. There must be a time $0 < t_\sigma < 1$ such that, below this instant, the noise dominates the system and pushes the solution far enough from the singularity, while above, the drift dominates the system and constrains the solution to fluctuate around one of the extremal solutions of the deterministic Peano equation (which was chosen by flipping a coin, in a Brownian case). A good way to see how the instant t_σ looks like is to compare the fluctuations of the extremal solutions $(\pm c_\beta t^{1/(1-\beta)})$ with the fluctuations of the noise "t^γ". This leads to the following equation for t_σ

$$\sigma t_\sigma^\gamma = t_\sigma^{\frac{1}{1-\beta}},$$

which gives $t_\sigma = \sigma^{(1-\beta)/(1-\gamma(1-\beta))}$ and leads to the condition:

$$\beta > 1 - \frac{1}{\gamma}. \tag{weak heuristic rule}$$

Thus, the more singular the drift is, the more "irregular" the noise has to be to restore well-posedness. We emphasize that, having this observation in mind, it is possible to build a rigorous class of counter-examples for weak well-posedness to hold below such a threshold (see Chapter 2). This is the reason why we say that this analysis is related to weak regularization.

[1] Here, $C_b^\beta(\mathbf{R}^d)$ denotes the space of bounded Hölder continuous function on $\mathbf{R}^d$, we refer to 6.5 for details on the functional spaces used in the manuscript.

[2] For a negative value of β, one may think of a generalized derivative (whence a distribution) of a $C_b^{\beta+1}$ map, and so on if $\beta < -1$.

[3] The terminology of "formal" relies on the fact that, in the case of a distributional drift, the meaning of the SDE has to be specified.

[4] We emphasize that the Authors consider the case of a Brownian noise, but we choose to formulate it more generally for pedagogical purposes.

The PDE viewpoint Let us restrict our discussion to a certain class of noises for which the associated SDE would have a Markovian solution[5]. For instance, α-stable noise with stability parameter α in $(0, 2]$, where the case $\alpha = 2$ corresponds to a Brownian motion. The PDE associated with the SDE can be written, in the framework of the Peano example, as follows:

$$\partial_t u(t,x) + \mathrm{sgn}(x)(|x|^\beta) Du(t,x) + \frac{1}{2}\Delta^{\frac{\alpha}{2}} u(t,x) = 0,$$

where $\Delta^{\alpha/2}$ stands for the usual fractional Laplacian. The previous discussion more or less suggests that, in order to preserve weak well-posedness, the transport term of the PDE should be a negligible perturbation of the fractional Laplacian in small time (*i.e.* the noise dominates). To see this, let us introduce the corresponding rescaled function $u_\lambda(t,x) := u(\lambda t, \lambda^{1/\alpha} x)$ for $\lambda > 0$ (reflecting the parabolic scale for t and x). We then get that u_λ satisfies the following equation:

$$\partial_t u_\lambda(t,x) + \lambda^{1-\frac{1}{\alpha}} \mathrm{sgn}(x)|\lambda^{\frac{1}{\alpha}} x|^\beta Du_\lambda(t,x) + \frac{1}{2}\Delta^{\frac{\alpha}{2}} u_\lambda(t,x) = 0,$$

so that the terms associated with the principal part of the partial differential operator in the above PDE, namely $\partial_t u_\lambda$ and $\Delta^{\alpha/2} u_\lambda$, are comparable. On the other hand, if $\beta > 1 - \alpha$, the scaled drift coefficient goes to zero with λ, and the principal part of the partial differential operator dominates; if $\beta = 1 - \alpha$, the scaled drift coefficient stays at a macro scale and the rescaled drift has the same order as the principal part of the partial differential operator (critical case); otherwise, the drift explodes when λ goes to zero. As in this case we have $\alpha = 1/\gamma$, we have thus recovered the weak heuristic rule from the probabilistic analysis done before.

1.1.2 Strong regularization

The case of strong uniqueness is much more involved. Let us first emphasize that all the results we found in the literature and that hold in any dimension[6], see *e.g.* [Zvo74, Ver80, KR05, Zha10, Dav07, FF11, Pri12a, BFGM19, CG16, WZ16, FFPV17, CZZ21, Pri18a, Ger22], suggest the following heuristic rule:

$$\beta \geq 1 - \frac{1}{2\gamma}. \qquad \text{(strong heuristic rule)}$$

Thus, there is a gap between strong and weak regularization. The main point now is that, to the best of our knowledge, there are no counter-examples that prove that this gap is indeed a price to pay for passing from weak to strong regularization. In order to understand where this gap comes from, we need to go a little further in the PDE based proof, known as Zvonkin Transform. This is the purpose of the next section. We believe that through this, the reader may be convinced that the above rule gives the optimal threshold for the method employed.

The Zvonkin Transform The main idea of the Zvonkin approach consists in rewriting the SDE (with $\sigma = 1$) as

$$X_t = u(t, X_t) + X_0 - u(0, X_0) - M_{0,t}(\alpha, u, X) + \mathcal{W}_t, \qquad \text{(Zvonkin transform)}$$

[5] Actually, only Markovian noises will be considered in this manuscript.

[6] The scalar case may indeed introduce some specific features, see *e.g.* [BC03, GO13, ABM20, 13]

with

$$M_{0,t}(\alpha,u,X) := \begin{cases} \int_0^t Du(s,X_s)\cdot d\mathcal{W}_s, & \text{if } \alpha = 2; \\ \int_0^t \int_{\mathbf{R}\setminus\{0\}} \tilde{N}(dr,dz)\{u(s,X_{s^-}+z)-u(s,X_{s^-})\}, & \text{if } \alpha<2, \end{cases}$$

with $\tilde{N}$ the compensated Poisson measure associated with $\mathcal{W}$ and where u is the solution of

$$\partial_t u(t,x) + \langle F(t,x), Du(t,x)\rangle + \frac{1}{2}\Delta^{\frac{\alpha}{2}} u(t,x) = F(t,x), \quad u(T,\cdot)=0.$$

The above stochastic dynamics is known as the Zvonkin transform. It follows from Itô's formula, provided that u is smooth enough. This transformation allows to eliminate the bad drift in the equation and to replace it with the solution of a parabolic PDE that benefits from the smoothing effect associated with the generator of the noise. Indeed, the mild solution of the above formal PDE writes

$$u(t,x) = \int_t^T ds P^\alpha_{s-t}[\{F(s,\cdot)+F(s,\cdot)\cdot Du(s,\cdot)\}](x), \ x\in\mathbf{R}^d,$$

where $(P^\alpha_t)_t$ denotes the usual stable semi-group. It is well known that this semi-group maps C^β_b to $C^{\beta+\alpha}_b$ with a critical singularity (*w.c.s.*)[7]. This is also known as Schauder estimates, see *e.g.* [Fri64, KP10] for $\alpha=2$ or [10, MP14] in the pure jump case. Therefore, we could expect at best that $u(t,\cdot)\in C^{\alpha+\beta}_b$.

Note now that the Zvonkin transform is (almost) a classical SDE, without any clear non-degeneracy assumption on the noise. Hence, its strong well-posedness relies on the usual stability argument of Grönwall type. Therefore, we need the stochastic integral $M(\alpha,u,X)$ to be Lipschitz continuous in the spatial variable. When $\alpha<2$, it follows from the interpolation type Lemma 4.1 in [Pri12a] that a sufficient condition is that for any t in $[0,T]$, $u(t,\cdot)$ must belong to $C^{1+\eta}_b$, $\eta>\alpha/2$. When $\alpha=2$, it suffices to have that for any t in $[0,T]$, $u(t,\cdot)$ belongs to C^2_b.

Therefore, we obtain that β must satisfy $\alpha+\beta>1+\alpha/2 \Leftrightarrow \beta>1-\alpha/2$. In other words, denoting the self similarity index of the noise by γ again, we get that strong well-posedness holds for

$$\beta > 1-\frac{1}{2\gamma}.$$

Thus, we can conclude that the threshold for strong well-posedness coming from strong heuristic rule is optimal for the Zvonkin approach, although it appears in non-Markovian cases where, obviously, PDE does not come into play.

[7] 💬 *The reader may wonder why, given the form of the above mild representation, a critical singularity is satisfying. This comes from the fact that, for non-integer β, it is possible to control the $C^{\beta+\alpha}_b$ of the integrated semi-group uniformly in t, i.e. there exists $C>0$ (not depending on t) such that for any f in C^β_b, $|\int_t^T P^\alpha_{s-t} f ds|_{C^{\beta+\alpha}_b} < C|f|_{C^\beta_b}$. From now on, the sentence w.c.s. will refer to this fact.*

Chapter 2

Regularization by degenerate noise

In this chapter, we summarize the results obtained in the works [4],[6] and [7],[8]. The common feature of all this works is the model considered: we investigated the smoothing properties of a noise propagating through a chain of n differential equations or oscillators. When $n = 2$, the model can be written as a classical kinetic model. In other words, we were concerned with the smoothing effect of a degenerate noise. This was studied at three levels: weak regularization was derived in [4],[6], Schauder estimates for the corresponding PDE were derived in [7] and strong regularization in [8].

To connect this chapter with Chapter 1, the reader may consider these works as an attempt (and a successful one !) to make rigorous the discussion on the thresholds and the associated smoothing effect related to the strong and weak regularization phenomenon. As previously noted, in the classical Brownian and sub-critical stable case (*i.e.* when the parameter α of the Markovian stable noise belongs to $(1, 2]$), the weak heuristic rule (to derive weak well-posedness) may lead to consider distributional drift, which implies additional issues to be resolved (we refer to Chapter 3 for further investigations in this direction). In this chapter, we somehow disrupt the noise by making it propagate through a weak Hörmander structure. As a consequence, the noise added into the system degenerates and its associated scaling exponent "γ" is bigger than 1: the noise is more regular. In view of the rules previously obtained, this lead to positive thresholds for the minimal regularity required on the drift, so that we can work with functions. This is more flexible to use tools from stochastic calculus.

2.1 Peano-like system perturbed by a noise propagating through a weak Hörmander structure

In this section, we present the archetypal model considered in the previously mentioned works from our personal bibliography. We have simplified the exposition as much as possible, while retaining the essential properties of the model. Furthermore, we have chosen this model to fit as close as possible with the discussion presented in Chapter 1.

💬 *In the following, we invite the reader to perform the calculations in the case $n = 2$ to avoid confusion between all indices and perhaps gain a better understanding of what is happening, with the kinetic case in mind.*

The noise Let us first introduce the archetypal noise propagating through the weak Hörmander structure. For $n \geq 2$, it writes,

$$d\mathcal{W}_t = A\mathcal{W}_t dt + E dB_t, \text{ with } A = \begin{pmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ 1 & 0 & \cdots & 0 & \vdots \\ 0 & 1 & 0 & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}, \quad X_0 = \mathbf{0}, \tag{2.1.1}$$

where $E = (1, 0, \dots, 0)^* = (1, \mathbf{0}_{1,n-1})^*$. In other words, $\mathcal{W}$ is simply a Brownian motion and its iterated time integral (we emphasize that this could have been a d-dimensional Brownian motion and its associated iterated time integral). This simple model is a natural generalization of the Kolmogorov's example [Kol34] ($n = 2$). In this work, Kolmogorov showed that this process is Gaussian and admits a smooth density, so this noise is said to be hypoelliptic while being non-elliptic, see [Hör67]. The main point to notice now is that each component of this noise lives at its own time scale: the i^{th} component $\mathcal{W}^i$ lives at a time scale or order $t^{i-1/2}$, $1 \leq i \leq n$. This can be clearly seen when considering the associated density q which satisfies, for $x, y \in \mathbf{R}^n$, $t < s \in (\mathbf{R}_+)^2$,

$$q(t, s, x, y) \asymp g_c(t, s, x, y) := \frac{1}{(s-t)^{\frac{n^2 d}{2}}} \exp\left(-c^{-1}\frac{1}{2}(s-t)|T_{s-t}^{-1}(e^{tA}x - y)|^2\right),$$

with $c = 1$ and where T_r is the (diagonal) intrinsic scale matrix of the system *i.e.* $(T_r)_{i,i} = r^i$, $1 \leq i \leq n$. We point out that the Gaussian density in the above right hand side also suggests that for any $2 \leq i \leq n$, the i^{th} component feels the transport of the initial conditions of the previous levels[1] $1 \leq j < i$. This is reminiscent of the model: these transports, which are at scale of t^j, $1 \leq j < i \leq n$, are not negligible in small time, as the typical time scale of the i^{th} component is $t^{i-1/2}$.

The Peano-like system We now pick a collection of indices $(\beta_i^j)_{2 \leq i \leq j \leq n}$ each of them in $(0, 1)$ and we consider the following Peano-like system:

$$dX_t = P(X_t)dt, \quad P = (P_1, \dots, P_n)^*, \quad P_i = \sum_{j=i}^{n} P_i^j, \quad P_i^j = p_{i,j}\text{sign}(\mathbf{x}_j)|\mathbf{x}_j|^{\beta_i^j}.$$

(Peano like system)

This system is precisely designed to be perturbed by the above noise. The perturbed version reads as:

$$dX_t = (AX_t + P(X_t))\, dt + E dB_t. \tag{perturbed Peano like system}$$

[1] This is well-seen from the specific sub-diagonal structure of A and fact that the matrice is nilpotent of degree n.

2.2 Failure of uniqueness for a perturbed Peano-like system

For $i \in [\![2, n]\!]$ and $j \in [\![i, n]\!]$, we consider the Peano like system with $p_{k,\ell} = 0$ for any $(k, \ell) \neq (i, j)$. Namely,

$$\begin{cases} \dot{x}_t^1 = \dot{x}_t^2 = \ldots = \dot{x}_t^{i-1} = 0 \\ \dot{x}_t^i = \mathrm{sign}(x_t^j)|x_t^j|^{\beta_i^j} \\ \dot{x}_t^{i+1} = x_t^i, \ldots, \dot{x}_t^j = x_t^{j-1} \\ \dot{x}_t^{j+1} = \dot{x}_t^{j+2} = \ldots = \dot{x}_t^n = 0 \end{cases} \qquad t \geq 0,$$

and $x_0^l = 0$ for all l in $[\![1, j]\!]$. Each entry $(x_t^k)_{t\geq 0, k\in[\![1,n]\!]}$ of the above dynamics is scalar. It is well seen that the global well-posedness of this system relies on the well-posedness of the i^{th} equation, whose extremal solutions write $\pm c_{\beta_i^j}^{ij} t^{((j-i)\beta_i^j+1)/(1-\beta_i^j)}$, for some positive $c_{\beta_i^j}^{ij}$. At this stage, it is worth noting that this system also have some interesting features: not only do the extremal solutions depend on the level ("i") of the chain, but they also depend on their position ("j") in the equation.

Still reasoning by analogy with Chapter 1, we begin by comparing the fluctuations of the noise and the extremal solutions. To this end, we add a small parameter $\sigma > 0$ in the front of the Brownian motion in the above system (*i.e.* with σE instead of E). Let us now fix $2 \leq i \leq j \leq n$ and consider thus the component SDE

$$\begin{aligned} Z_t^{i,j} &= x_0 + \sigma \mathcal{W}_t^i \qquad \text{(component SDE)} \\ &+ \int_0^t \mathrm{sign}\left(\int_0^s \ldots \int_0^{s_{i+2}} Z_{s_{i+1}}^{i,j} ds_{i+1} \ldots ds_j \right) \left| \int_0^s \ldots \int_0^{s_{i+2}} Z_{s_{i+1}}^{i,j} ds_{i+1} \ldots ds_j \right|^{\beta_i^j} ds \end{aligned}$$

where $\mathcal{W}^i$ is the $(i-1)^{\text{th}}$ iterated integral in time of the Brownian motion. As this noise (impulsed in the i^{th} level of the chain) lives at scale $i - 1/2$, we are thus looking for a small time $t_\sigma^{i,j}$ solving

$$\sigma t^{\frac{2i-1}{2}} = \sigma t^{\frac{(j-i)\beta_i^j+1}{1-\beta_i^j}},$$

which gives as a necessary condition for $t_\sigma^{i,j}$ to be defined:

$$\beta_i^j > \frac{2i-3}{2j-1} \qquad (n = 2 \Rightarrow \beta_2^2 > 1/3).$$

It turns out that such a comparison can also yield counter-examples. Let us recall that, in their work, Delarue and Flandoli showed that for the noise to restore well-posedness, there must exist a time $0 < t_\sigma^{i,j} < 1$ such that below this time, the solution behaves like the noise, and above it, it fluctuates around one of the extremal solutions of the Peano system. Let us assume that this instant does not exist *i.e.* that $\beta_i^j < [2i-3/(2j-1)]$. It is thus natural to wonder wether as soon as the solution starts just above the singularity, it will remain around the positive extremal solution of the Peano system with high probability, as the intensity of the noise is not "strong enough" to

balance the fluctuations of the extremal solution. This can be made rigorous. Introduce for a given parameter $\lambda \in (0,1)$, for any continuous path Y from $\mathbf{R}_+$ to $\mathbf{R}$, the following stoping time

$$\tau(Y) = \inf\{t \geq 0 \ : \ Y_t \leq (1-\lambda) c^{ij}_{\beta_i^j} t^{((j-i)\beta_i^j+1)/(1-\beta_i^j)}\}.$$

we prove in [4] ($n=2$) and [6] the following lemma.

Lemma 2.2.1. *Let $Z^{i,j}$ be the weak solution of the component SDE starting from some $x_0 > 0$ and suppose that $\beta_i^j < [2i-3/(2j-1)]$. Then, there exists a positive ρ, depending on λ, β_i^j, i and $\mathbf{E}|\mathcal{W}_1^i|$ only, such that*

$$\mathbf{P}_{x_0}(\tau(Z^{i,j}) \geq \rho) \geq 3/4.$$

Letting the initial condition x_0 tend to 0, we can deduce that the weak solution $(Z^{i,j}, \mathcal{W}^i)$ of the component SDE starting from 0 at time 0 remains around the positive extremal with high probability. But the symmetry of the system implies that this is also the case for $(-Z^{i,j}-, \mathcal{W}^i)$, and if uniqueness in law holds, this means that

$$\mathbf{P}_0(\tau(Z^{i,j}) \geq \rho) \geq 3/4 \text{ and } \mathbf{P}_0(\tau(-Z^{i,j}) \geq \rho) \geq 3/4,$$

which is obviously impossible. We deduce the following result taken from [4] ($n=2$) and [6].

Theorem 2.2.2 (Failure of weak uniqueness)**.** *For any*

$$\beta_i^j < \frac{2i-3}{2j-1},$$

uniqueness in law fails for the component SDE with $x_0 = 0$.

To resume, for a diagonal system, *i.e.* $p_{i,j} = 0$, $i \neq j$, we obtain in the case $n=2$ (kinetic model) that uniqueness fails as soon as $\beta_2^2 < 1/3 = 1 - 1/(3/2)$ where $t^{3/2}$ represents the typical order of fluctuations of $\int_0^t B_s ds$. More generally, this proves that uniqueness fails as soon as $\beta_i^i < 1 - 2/(2i-1) = 1 - 1/(i-1/2)$ where $i - 1/2$ stands for the typical order of fluctuations of the $(i-1)^{\text{th}}$ iterated in time integral of a Brownian motion: this is precisely the weak heuristic rule in Chapter 1.

2.3 Smoothing properties of the PDE associated to the Peano like system: tools

The results on weak and strong regularization may rely heavily, in the Markovian setting, on the smoothing properties of the associated PDE. Here, we summarize the main (and common) ingredients that will be useful for further investigations in these directions. We first write the operator and PDE associated to the perturbed Peano example. For any positive time horizon $T > 0$, we define the Cauchy problem $\mathscr{C}(T, f, g, \mathcal{L})$ as follows:

$$\begin{aligned}
&\partial_t u(t,x_1,\dots,x_n) + \mathcal{L}u(t,x_1,\dots,x_n) := \partial_t u(t,x_1,\dots,x_n) + \frac{1}{2}\Delta_{x_1} u(t,x_1,\dots,x_n) \\
&\quad + \langle Ax, D_x u(t,x_1,\dots,x_n)\rangle + \sum_{i=2}^n \sum_{j=i}^n \text{sign}(x_j)(|x_j|^{\beta_i^j}) \cdot D_{x_i} u(t,x_1,\dots,x_n) \\
&= -f(t,x_1,\dots,x_n), \quad (t,x_1,\dots,x_n) \in [0,T) \times \mathbf{R}^n \\
&u(T,x_1,\dots,x_n) = g(x_1,\dots,x_n), \ (x_1,\dots,x_n) \in \mathbf{R}^n,
\end{aligned}$$

where the maps f, g will be specified later on.

Our strategy to study the smoothing properties of $\mathcal{L}$ on the data (P, f, g) rests upon the parametrix approach see *e.g.* the work of McKean and Singer [MS67] or the book of Friedman [Fri64]. Roughly speaking, this approach is a perturbative argument consisting in expanding the operator $\mathcal{L}$ around a good proxy, usually denoted by $\tilde{\mathcal{L}}$. In our setting, the term good proxy relates to the fact that the operator $\tilde{\mathcal{L}}$ is the generator of the "closest" Gaussian approximation of the process formally associated with the generator $\mathcal{L}$. The associated fundamental solutions, denoted by p and $\tilde{p}$, play a key role (assuming they exist, which is the case when the coefficients of the generator and the data (f, g) are smooth by Hörmander theorem [Hör67]).

The parametrix kernel When employing the parametrix approach, we are led to control the approximation error of the operator $\mathcal{L}$ by the proxy one $\tilde{\mathcal{L}}$. This error is encoded in what we call the parametrix kernel and usually takes the form, on a given time interval $[t, T]$,

$$\left| \int_t^T [(\tilde{\mathcal{L}} - \mathcal{L})\psi_1(s, y)]\psi_2(s, y) dy ds \right|$$

for some suitable maps ψ_1, ψ_2 for which the above formulation makes sense, and which mainly depend on the way we chose to approximate $\mathcal{L}$. Let us only clarify at this stage that one of these maps usually takes the form of the density of the proxy (*i.e.* of the fundamental solution associated with it). In all the cases we consider, it is necessary to prove that this error is negligible in small time so that the generator $\tilde{\mathcal{L}}$ is indeed a good approximation of $\mathcal{L}$. As one of the maps involved in the above error takes the form of the density of the proxy, we understand that the perturbation must be of the same order as the typical fluctuations of this process.

The proxy and associated freezing curve In our case, the "good proxy" is the generator $\tilde{\mathcal{L}}$ of the Gaussian (Ornstein-Uhlenbeck) process $\tilde{X}$ whose dynamics writes

$$d\tilde{X}_t = P(\theta_t)dt + A\tilde{X}_t dt + E dB_t, \quad \tilde{X}_t = e^{tA}\tilde{X}_0 + \int_0^t e^{-(s-t)A} P(\theta_s) ds + \int_0^t e^{-(s-t)A} E dB_s$$

for some freezing curve $\mathbf{R}_+ \ni t \mapsto \theta_t \in \mathbf{R}^n$ and where for the second equality in the above, we explicitly solved the equation. This process admits a Gaussian density $\tilde{p}$ which is also the fundamental solution of $\tilde{\mathcal{L}}$.

To go further, we need to choose a freezing curve. In the previous paragraph, we emphasized that the perturbation must be of the typical order of the process. Let us consider, for instance, the case of a non degenerate Gaussian noise (*i.e.* $n = 1$ in our setting). In such a case, the mean of the Gaussian process is the map $m_{t,s} := \mathrm{id}$, for all $t < s \in \mathbf{R}_+^2$ and if we want to approximate $\mathcal{L}$ on a given interval $[t, s]$, the idea consists in either freezing the coefficients on the initial condition x at time t (forward parametrix) or on the terminal condition y at time s (backward parametrix). This reproduces the typical fluctuations of the process well. Each of the two approaches has its benefits and drawbacks. The first one is very robust and allows to handle a large class of problems but requires more *a priori* knowledge on the operator $\mathcal{L}$; the second is more subtle but appears to be more delicate to work with and less flexible, especially because, at the end of the day, we somehow loose the fact that we work with a density.

In the general case $n \geq 2$, the story becomes rather different since, as we saw on the density q of the noise, the i^{th} component feels the transport of the j^{th} previous one, $j \leq i-1$. Thus, to reproduce the typical fluctuations of the process, we need to freeze along the transport term of the system *i.e.* along the curve given as the solution of

$$\dot{\theta}_{t,s}(y) = F(t, \theta_{t,s}(y)), \ \theta_{s,s}(y) = y$$

which we call the backward flow or

$$\dot{\theta}_{t,s}(x) = F(t, \theta_{t,s}(x)), \ \theta_{t,t}(x) = x$$

which we call the forward flow. Note that in our setting, the existence of a solution to such equations follows from Peano's theorem, while without mollifying procedure, the uniqueness is not guaranteed. However, this does not prevent us from choosing, once and for all, a given solution and to working with it.

As a consequence, we distinguish two kinds of degenerate Gaussian type kernel $g_{c,m}^{\text{backward}}$ and $g_{c,m}^{\text{forward}}$ which reproduce the behavior of the O.U. process. We indeed know from the work of Delarue and Menozzi [DM10] that the following estimates hold in either the forward or backward cases: for any $l \in [\![0, n]\!]$, $r \in \{0, 1\}$:

$$|D_{\mathbf{x}_l} D_{\mathbf{x}_1}^k \tilde{p}(t, s, x, y)| \ \leq \ \frac{\bar{c}}{(s-t)^{(l-\frac{1}{2})+\frac{k}{2}}} g_c(t, s, x, y),$$

where

$$g_c(t, s, x, y) =: g_c^{\text{backward}}(t, s, x, y) = \frac{c}{(s-t)^{\frac{n^2 d}{2}}} \exp\left(-c^{-1}(s-t)|T_{s-t}^{-1}(x - \theta_{t,s}(y))|^2\right)$$

in the backward case and

$$g_c(t, s, x, y) =: g_c^{\text{forward}}(t, s, x, y) = \frac{c}{(s-t)^{\frac{n^2 d}{2}}} \exp\left(-c^{-1}(s-t)|T_{s-t}^{-1}(\theta_{t,s}(x) - y)|^2\right)$$

otherwise. The above estimates reflect the different scales at which each component of the system lives. In particular, compared to the typical behavior of non degenerate Gaussian noise, it must be noted that even a first order differentiation w.r.t. a degenerate variable may produce a non-integrable time singularity.

2.4 Weak regularization for the perturbed Peano like system

In the light of the previous section, it is unnecessary to look for solutions of the perturbed Peano like system in the regime $\beta_i^j < [2i-3/(2j-1)]$. We thus focus from now on on the complementary set, up to the critical case *i.e.* we assume that

$$\beta_i^j > \frac{2i-3}{2j-1}, \ 2 \leq i \leq j \leq n.$$

Under such conditions, the continuity assumed on the coefficients allows to use a compactness argument to obtain existence of a solution. The tricky part concerns uniqueness in law. The crucial point here is that, in this framework, few informations are known on the law or the generator $\mathcal{L}$ of the process. On the other hand, we obtained numerous results on the solution of a frozen (along a suitable transport term) version of the Peano like system in the previous paragraph. We thus aim to take advantage of this knowledge, viewing the generator $\mathcal{L}$ as a perturbation of the frozen one. This is where parametrix comes into play. The approach implemented in [6] is in the spirit of the original approach of Stroock and Varadhan to prove uniqueness for the Martingale problem [SV79]. The objective, to obtain uniqueness, is to prove that for a sufficiently large class of functions $f : [0,T] \times \mathbf{R}^n \to \mathbf{R}$ it holds that for any two martingale solutions $\mathbf{P}_i$, $i = 1, 2$ starting from x at time t, it holds that

$$\mathbf{E}_{\mathbf{P}_1(t,x)}\left[\int_t^T f(s, X_s)ds\right] = \mathbf{E}_{\mathbf{P}_2(t,x)}\left[\int_t^T f(s, X_s^{t,x})ds\right].$$

To do so, the idea consists in approaching the above quantities by the one associated to the proxy. Let $f \in C_0^\infty([0,T) \times \mathbf{R}^n, \mathbf{R})$ be given. Denoting by $\tilde{\mathbf{P}}$ the associated law, we may define

$$\tilde{G}f(t,x) = \mathbf{E}_{\tilde{\mathbf{P}}(t,x)}\left[\int_t^T f(s, X_s)ds\right] = \int_t^T \int_{\mathbf{R}^n} f(s,y)\tilde{p}(t,s,x,y)dyds.$$

Assuming that this is a smooth map from $[0,T) \times \mathbf{R}^n \to \mathbf{R}$, this quantity solves the Cauchy problem

$$\partial_t \tilde{G}f + \tilde{M}_t f = -f,\ (t,x) \in [0,T) \times \mathbf{R}^n,$$

where, for all $(t,x) \in [0,T) \times \mathbf{R}^n$,

$$\tilde{M}_t f := \int_t^T \int_{\mathbf{R}^n} \tilde{\mathcal{L}}\tilde{p}(t,s,x,y)f(s,y)dyds,$$

(the choice of such an operator will be justified latter on) and we may expand the Green kernel $\tilde{G}$ applied to some smooth map f along any given martingale solution through Itô's formula to obtain

$$\begin{aligned} 0 = \mathbf{E}\left[\tilde{G}f(T, X_T^{t,x})\right] &= \tilde{G}f(t,x) + \mathbf{E}\left[\int_t^T (\partial_s + \mathcal{L})\tilde{G}f(s, X_s^{t,x})ds\right] \\ &= \tilde{G}f(t,x) + \mathbf{E}\left[\int_t^T (\mathcal{L}\tilde{G} - \tilde{M}_s)f(s, X_s^{t,x})ds\right] - \mathbf{E}\left[\int_t^T f(s, X_s^{t,x})ds\right], \end{aligned}$$

using as well the Cauchy problem satisfied by $\tilde{G}f$. From here, the strategy consists in first controlling the first two terms in the right hand side above in terms of the $L^q([0,T], L^p(\mathbf{R}^n))$-norm of f to obtain the a.e. existence of a density for the $\mathbf{P}_i(t,x)$, $i = 1, 2$ lying in the dual Lebesgue space. This follows from a pointwise control of the parametrix kernel, which is given by $(\mathcal{L}\tilde{G} - \tilde{M}_s)f$ and by the Gaussian estimate on the density $\tilde{p}$. Once this is proven, it remains to estimate suitably the $L^q([0,T], L^p(\mathbf{R}^n))$ norm of the parametrix kernel and deduce that the operator $[I - (\mathcal{L}\tilde{G} - \tilde{M}_s)]$ is invertible on this space. We thus choose f in the above to be the inverse of this operator applied to some function $\phi \in L^q([0,T], L^p(\mathbf{R}^n))$ and obtain, for $i = 1, 2$ that

$$\mathbf{E}_{\mathbf{P}_i}\left[\int_t^T \phi(s, X_s^{t,x})ds\right] = \tilde{G}\Big[[I - (\mathcal{L}\tilde{G} - \tilde{M}_s)]^{-1}\Big]\phi(t,x),$$

from which uniqueness follows. Let us noticed that, in order to conclude, it is crucial to obtain a good control on the parametrix kernel. We briefly detail how this control works. Let $f \in C_0^\infty([0,T)\times\mathbf{R}^n,\mathbf{R})$, we have, for any $(t,x)\in[0,T)\times\mathbf{R}^n$, using estimates on the derivative of the proxy:

$$\begin{aligned}
&\left|(\mathcal{L}\tilde{G}-\tilde{M}_s)f(t,x)\right| \\
=\;&\left|\int_t^T\int_{\mathbf{R}^n} f(s,y)(\mathcal{L}-\tilde{\mathcal{L}})\tilde{p}(t,s,x,y)dyds\right| \\
\leq\;&\int_t^T\int_{\mathbf{R}^n}|f(s,y)|\sum_{i=2}^n\sum_{j=i}^n\left|\left(\mathrm{sign}(x_j)|x_j|^{\beta_i^j})-\mathrm{sign}(\theta_s)|\theta_s|^{\beta_i^j}\right)D_{x_i}\tilde{p}(t,s,x,y)\right|dyds \\
\leq\;&\int_t^T\int_{\mathbf{R}^n}|f(s,y)|\sum_{i=2}^n\sum_{j=i}^n\left|x_j-\theta_s^j\right|^{\beta_i^j}\frac{\bar{c}}{(s-t)^{(i-\frac{1}{2})}}g_{c,m}(t,s,x,y)dyds.
\end{aligned}$$

To conclude, the procedure would now consist in choosing the curve θ along with we freeze the coefficients in $\mathcal{L}$ in order to fit the exponential decay in the Gaussian density g[2]: this is what was hinted in the previous paragraph by saying that the perturbation must be of the typical order of the frozen process. Here, this gives us that θ should be chosen as the backward flow with terminal condition y at time s, which was defined in the previous section. This justifies, *a posteriori*, the choice of the operator $\tilde{M}$ above which allows to integrate w.r.t. the terminal conditions. In this case, starting from the last inequality obtained in the previous equation, we obtain that (by deteriorating the variance of the Gaussian kernel)

$$\begin{aligned}
&\left|(\mathcal{L}\tilde{G}-\tilde{M}_s)f(t,x)\right| \\
\leq\;&C\sum_{i=2}^n\sum_{j=i}^n\int_t^T\frac{ds}{(s-t)^{i-\frac{1}{2}-\beta_i^j(j-\frac{1}{2})}}\int_{\mathbf{R}^n}g_{c,m}^{\mathrm{backward}}(t,s,x,y)|f(s,y)|dy,
\end{aligned}$$

for some $C>0$. The condition on the family of indexes (β_i^j) is precisely designed to give that for all $i\in[\![2,n]\!], j\in[\![i,n]\!]$, $1-\{(i-1)+1/2-\beta_i^j(j-1+1/2)>0$ so that all the above time singularities are integrable. To conclude this computation, we eventually need to prove that $g_{c,m}^{\mathrm{backward}}$ is comparable to a density. Indeed, the dependence on the covariance matrix w.r.t. the spatial integration variable in $\tilde{p}$ (the space integration is done over the terminal condition which appears in the freezing curve) breaks the density property of $g_{c,m}^{\mathrm{backward}}$. If the coefficients in the equation solved by the curve θ were Lipschitz, θ would be a Lipschitz homeomorphism, and in such case, one could use the fact that

$$\begin{aligned}
C^{-1}(s-t)|\mathbf{T}_{s-t}^{-1}(\theta_{s,t}(x)-y)|^2 &\leq (s-t)|\mathbf{T}_{s-t}^{-1}(x-\theta_{t,s}(y))|^2 \\
&\leq C(s-t)|\mathbf{T}_{s-t}^{-1}(\theta_{s,t}(x)-y)|^2,
\end{aligned}$$

for some $C>0$. The point is that, in our setting, the solution of the transport equation is not smooth and is even non unique. We thus managed to prove that the above estimate holds for the chosen flow up to some additional constant in the upper and lower bounds, and where the latter may involved any other Peano solution of the transport equation. This allowed us to deduce the result.

[2]By using the inequality: $\forall\eta>0,\ \forall q>0,\ \exists\bar{C}>0$ s.t. $\forall\sigma>0,\ \sigma^q e^{-\eta\sigma}\leq\bar{C}$.

Let us emphasize that the choice of the freezing curve as described above also leads to other pathological phenomena. In such a case, the Cauchy problem associated with $\tilde{G}f$ and $\tilde{M}_t f$ stated above is ill-posed, as the lost of the density property prevents us from benefiting from the smoothness of f around the singularity in t. This can be circumvent through suitable truncation procedure, which we will not elaborate on here.

2.5 Schauder estimates

Once weak well-posedness has been proved, we can investigate the associated Cauchy problem $\mathscr{C}(T, f, g, \mathcal{L})$. It is indeed known that well-posedness of the martingale problem gives rise to the existence and uniqueness of a mild solution (*i.e.* as a semi-group) for the PDE. Having this mild solution at hand, it is reasonable to investigate what the smoothing effect of the semi-group is (especially in the perspective of building a strong solution). Namely, denoting by $(P_{t,s})_{0\leq t<s\leq T}$ the family of semi-groups associated with the Markovian (weak) solution of the perturbed Peano like system, we aim to establish, in a rather optimal way, the set of functions to which $P_{t,s}\varphi$ belongs for any $0 \leq t < s \leq T$, when φ lies is in a suitable (here anisotropic) Hölder space. This translates into PDE terminology as Schauder estimates fo the Cauchy problem $\mathscr{C}(T, f, g, \mathcal{L})$ where f and g belong to anisotropic Hölder spaces. This is the purpose of the work [7] which we sketch here.

The approach still relies on the parametrix method. We begin by considering a sequence of solutions $(u_m)_m$ of the Cauchy problem where the parameter m represents the smoothing parameter of the coefficients of the equation. Under this mollified setting, we know from Hörmander's theorem that the PDE is hyppoelliptic, so the sequence of maps $(u_m)_m$ is a sequence of smooth functions. As the Cauchy problem associated with both $\mathcal{L}^m$ and $\tilde{\mathcal{L}}$ are well posed for any $m > 0$, we can write

$$(\partial_t + \mathcal{L}^m)u_m = -f \Leftrightarrow (\partial_t + \tilde{\mathcal{L}})u_m = -f + (\tilde{\mathcal{L}} - \mathcal{L}^m)u_m.$$

Since $\tilde{p}$ is the fundamental solution associated with $\tilde{\mathcal{L}}$, we obtain that

$$\begin{aligned} u_m(t,x) = \int_{\mathbf{R}^n} g(y)\tilde{p}(t,T,x,y)dy \quad &+ \quad \int_t^T \int_{\mathbf{R}^n} f(s,y)\tilde{p}(t,s,x,y)dyds \\ &+ \quad \int_t^T \int_{\mathbf{R}^n} (\mathcal{L}^m - \tilde{\mathcal{L}})u_m(s,y)\tilde{p}(t,s,x,y)dyds. \end{aligned}$$

Having this representation, the strategy now consists in investigating for any j in $[\![1, n]\!]$, for any $(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n) \in \mathbf{R}^{n-1}$ the sensitivity of the map

$$x_j \mapsto u_m(t, x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_n), \quad t \in [0, T]$$

w.r.t. the sensitivity of f and g, and the parametrix kernel, uniformly in $m > 0$. In the previous part, we proved that under the condition assumed on the family of Hölder coefficients β (allowing weak well-posedness to hold), the parametrix kernel is negligible in small time, or equivalently, the drift (or transport term) is a negligible perturbation of the principal part of the operator (so that for any $m > 0$, $\tilde{\mathcal{L}} - \mathcal{L}^m \approx 0$), so we can forget the third term in the r.h.s. of the previous equation. Therefore, we mainly focus on the second term in the above right-hand side, the estimation of the first one following from similar arguments.

From the computations we have done in the previous section, we understand that it is hopeless to estimate derivatives in the degenerate direction (*i.e.* w.r.t. x_j, $j \geq 2$) without imposing any threshold on the Hölder regularity on the data f. We therefore investigate the Hölder norm of the solution instead. Let $x, x' \in \mathbf{R}^n$ such that $x_1 = x'_1$ (we do not focus on the non degenerate direction, which is more classical). We aim to estimate $u_m(t,x) - u_m(t,x')$, for t in $[0,T]$. However, the strategy differs depending on the regime in which the estimate is performed: if the spatial perturbation is small w.r.t. the order of the typical fluctuation of the process, which is the so-called diagonal regime, we compare the density; otherwise we are in the off-diagonal regime and we usually clumsily estimate the norm of the difference as the sum of the norms of each term.

To distinguish between the two regimes, we introduce the splitting time t_0 where the change of regime occurs. It can be easily derived from the various scalings associated with the proxy and takes the form:

$$t_0 := t + c_0 \left\{ \sum_{\ell=2}^{n} |x_\ell - x'_\ell|^{\frac{1}{2\ell-1}} \right\}^2 =: t + c_0 \mathbf{d}^2(x,x'), \quad c_0 > 0.$$

In other words, the regime occurs at different scales depending on the variable we consider. This naturally defines a homogeneous quasi-distance denoted above by $\mathbf{d}$. The change of regime occurs at two levels: at a global level where we naturally compare the global time length interval $T-t$ with the spatial perturbation, and at a local level inside the time integration when the global regime is diagonal (otherwise it remains off-diagonal once and for all). We now sketch the computations in the (global and local) diagonal regime to better understand what kind of smoothing effect can be expected. We have (keeping in mind that we illustrate the computations for $g \equiv 0$ and $\tilde{\mathcal{L}} - \mathcal{L} \approx 0$),

$$\begin{aligned} u(t,x) - u(t,x') &\approx \int_{t_0 \wedge T}^{T} \int_{\mathbf{R}^n} f(s,y) \{ \tilde{p}(t,s,x,y) - \tilde{p}(t,s,x',y) \} dyds \\ &= \int_{t_0 \wedge T}^{T} \int_{\mathbf{R}^n} f(s,y) \int_0^1 D_x \tilde{p}(t,s,x' + \lambda(x'-x), y) d\lambda \cdot (x - x') dyds, \end{aligned}$$

thanks to Taylor expansion. The point is to notice that we now need to exploit the regularity of f to quantify the smoothing effect of the generator. This is done through cancellation argument by rewriting the above term as:

$$\int_{t_0 \wedge T}^{T} \int_{\mathbf{R}^n} \{ f(s,y) - f(s,\theta_s) \} \int_0^1 D_x \tilde{p}(t,s,x' + \lambda(x'-x), y) d\lambda \cdot (x - x') dyds.$$

The remainder term is equal to o if $\tilde{p}$ is a density. This, together with the hope of using the Gaussian exponential decay to smoothen the singularity resulting from the above differentiation (in order to obtain a non-exploding bound in small time) suggests choosing the forward freezing curve with starting conditions (τ, ξ) as the freezing curve, and then set the parameters (τ, ξ) to (t,x) after differentiating and expanding the densities. Together with the Gaussian bound on the density of the proxy, we deduce that the above term can be estimated by:

$$\begin{aligned} & C \max_{1 \leq j \leq n} |f|_{L^\infty([f_j]_\infty^{\gamma_j})} \int_{t_0 \wedge T}^{T} ds \sum_{i=2}^{n} |(x-x')_i| (s-t)^{(i-\frac{1}{2})(\gamma_j - 1)} \\ \leq \; & C \max_{1 \leq j \leq n} |f|_{L^\infty([f_j]_\infty^{\gamma_j})} \sum_{i=2}^{n} \left\{ \sum_{j=2}^{n} |(x-x')_j|^{\frac{1}{2j-1}} \right\}^{2+\gamma_i(2i-1)}, \end{aligned}$$

where $|f|_{L^\infty([f_j]_\infty^{\gamma_j})}$ denotes the supremum in time of the Hölder-modulus of the map f w.r.t. its j^{th} variable, assumed to be uniform w.r.t. the other space variables. This leads us to choose $\gamma_j = \gamma/(2j-1)$ for any j in $[\![2,n]\!]$ to obtain that

$$\left| \int_{t_0 \wedge T}^{T} \int_{\mathbf{R}^n} f(s,y)\{\tilde{p}(t,s,x,y) - \tilde{p}(t,s,x',y')\} dyds \right| \leq C\mathbf{d}^{2+\gamma}(x,x'), \quad \forall m > 0.$$

This *a posteriori* justifies the choice of data (f,g) in anisotropic Hölder spaces associated with the quasi-distance $\mathbf{d}$ defined above, which reflects the particular scales of the system.

Let us briefly explain the main consequences of the above calculations when dealing with the off-diagonal regime. To conclude the above computations, we *a posteriori* chose the same initial condition $\xi = x$ for the forward transport of $u_m(t,x)$ and $u_m(t,x')$. However, it is clear that such a choice for the latter is anything but natural, since the associated proxy starts from x'. In the off-diagonal regime, we do not compare the density, but rather estimate the norm of each term in the difference. Hence, this requires the forward transport associated with each term to be initialized at point x and x', respectively. As a consequence, we need to investigate the sensitivity of the parametrix to a change of initial condition in the freezing curve. This can be achieved through a careful analysis of the mean and covariance of the proxy.

Finally, we discuss the control of the parametrix kernel. The implicit representation also implies many issues to solve. The first one may be the fact that the parametrix kernel involves differentiation of the solution u_m with respect to the degenerate variable, although we saw that the derivatives of u_m cannot be estimated uniformly with respect to the mollification parameter, at least for "general" data f, g. To overcome this problem, we rebalance the derivative onto the density, thanks to an integration by parts argument, and use a cancellation argument to take advantage of the *a priori* Hölder regularity of u_m. We thus obtain our estimates in terms of... our estimates. We first conclude in small times through a circular argument made possible by scaling the PDE and suitably equilibrating the constant appearing in the splitting time t_0 (where the change of regime occurs).

To conclude, we obtained, for any γ in $(0,1)$, denoting by $C^\gamma_{b,\mathbf{d}}$ the anisotropic Hölder space roughly introduced above, that for any $t < T \in \mathbf{R}_+$, the family of semi-group $(P_{t,s})_{t \leq s < T}$ maps $C^\gamma_{b,\mathbf{d}}$ to $C^{2+\gamma}_{b,\mathbf{d}}$. This means that, for any $j \in [\![1,n]\!]$, the map $x_j \mapsto P_{t,s}f(\cdot,x_j,\cdot)$ is in $C_b^{(2+\gamma)/2j-1}$: the smoothing effect translates into a gain of regularity of order $2/(2j-1)$. In other words, the smoothing effect decreases as one moves away from the source of noise. This particular feature is reminiscent of the weak Hörmander-like structure.

2.6 Strong regularization

As emphasized in Chapter 1, our strategy for proving strong well-posedness rests upon the Zvonkin transform. In the considered setting, it writes:

$$X_t = u(t,X_t) + x_0 - u(0,x_0) - \int_0^t Du(s,X_s)EdB_s + EB_t,$$

where $u = (u_1, \dots, u_n)^*$ and each u_i satisfies

$$\partial_t u_i(t,x) + \langle \underbrace{Ax + P(x)}_{=F(x)=(F_1,\dots,F_n)(x)}, Du_i(t,x)\rangle + \frac{1}{2}\Delta_{x_1} u_i(t,x) = \mathbf{F}_i(x).$$

We then need to obtain a Lipschitz continuous in space integrand in the stochastic integral associated with the Zvonkin transform. Due to the particular structure of the embedding matrix E (recall that $E = (1, \mathbf{0}_{1,n-1})^*$), this means that the gradient $D_1 u$ (w.r.t. the non degenerate variable) must be Lipschitz in all spatial directions. This requires the solution u to have the same regularity in all the spatial directions. As a consequence, the regularity exponent of the drift does not depend on the level "i" of the chain anymore, and this is why we now work with the family $\beta = (\beta_i^j)_{i,j}$ satisfying

$$\forall i,k \in [\![1,n]\!], \quad \beta_i^j = \beta_k^j =: \beta^j, \quad \forall j \in [\![1,n]\!].$$

As highlighted in the precedent section through Schauder estimates, we expect, for any i in $[\![1,n]\!]$, any j in $[\![1,n]\!]$, for fixed $(x_{1:j-1}, x_{j+1:n})$ in $\mathbf{R}^{n-1}$ and t in $[0,T]$, the map

$$(u_i)_j(t,\cdot) : \mathbf{R} \ni z_j \mapsto u_i(t, \mathbf{x}_{1:j-1}, z_j, \mathbf{x}_{j+1:n}) \in \mathbf{R} \tag{2.6.1}$$

to be in $C^{2/(2j-1)+\beta^j}$ for any β^j in $\big(0, 1/[2j-1]\big)$, uniformly w.r.t. t and $(\mathbf{x}_{1:j-1}, \mathbf{x}_{j+1:n})$. Especially, we can deduce that, as $(u_i)_j(t,\cdot)$ belongs to $C^{2/(2j-1)+\beta^j}$, $(D_{x_1} u_i)_j(t,\cdot)$ belongs to $C_b^{1/(2j-1)+\beta^j}$, *i.e.* due to homogeneity reasons, one differentiation w.r.t. the non degenerate variable induces a loss of Hölder regularity of order $1/(2j-1)$ w.r.t. the j^{th} degenerate variable. Since the previous Schauder estimates only hold for a bounded source, we need to extend them to the unbounded case. In this case, we cannot expect the solutions $(u_i)_{i\in[\![1,n]\!]}$ to be bounded anymore but rather to have linear growth (and consequently bounded gradients). Therefore, we specifically state the parabolic bootstrap in terms of usual Hölder spaces on the gradients. We prove that for any β^j in $((2j-2)/(2j-1), 1)$, the map $(D_{x_1} u_i)_j(t,\cdot)$ uniformly belongs to $C_b^{1/(2j-1)+\beta_j-\varepsilon}$ for any $0 < \varepsilon << 1$ (the infinitesimal loss of regularity is due to the fact that we did not need a "true" Schauder estimate to conclude the proof, which allowed us to bypass the scaling and the balancing procedure described previously). To obtain Lipschitz control in all the spatial directions of the gradient $D_{x_1} u$, we require $1/(2j-1) + \beta^j > 1 \Leftrightarrow \beta^j > (2j-2)/(2j-1)$. Not that the previous thresholds write $\beta^j > (2j-2)/(2j-1) = 1 - 1/(2j-1) = 1 - 1/[2 \times (j-1/2)]$ where $(j-1/2)$ denotes the typical fluctuations of the $(j-1)^{\text{th}}$ iterated time integral of a Brownian motion. Thus, we end up with the thresholds in strong heuristic rule, which we have shown to be optimal with respect to the methodology.

2.7 General model and results, comments

For some $n \geq 1$, we consider the following perturbed chain of differential equations

$$\begin{aligned}
dX_t^1 &= F_1(t, X_t^1, \dots, X_t^n)dt + \sigma(t, X_t^1, \dots, X_t^n)dB_t, \\
dX_t^2 &= F_2(t, X_t^1, \dots, X_t^n)dt, \\
dX_t^3 &= F_3(t, X_t^2, \dots, X_t^n)dt, \\
&\vdots \\
dX_t^n &= F_n(t, X_t^{n-1}, X_t^n)dt,
\end{aligned} \qquad t \geq 0$$

where $(B_t)_{t\geq 0}$ stands for a d-dimensional Brownian motion and for all $i \in [\![1,n]\!]$, $t \geq 0$ the component X_t^i is $\mathbf{R}^d$-valued as well (i.e. $X_t \in \mathbf{R}^{nd}$). The diffusion coefficient σ is assumed to be (at least) uniformly non degenerate and bounded. At the PDE level, the system writes, for any horizon time $T > 0$,

$$\begin{aligned}
&\partial_t u(t,x_1,\dots,x_n) + \langle F_1(t,x_1,\dots,x_n), D_{x_1}u(t,x_1,\dots,x_n)\rangle + \frac{1}{2}\mathrm{Tr}(D^2_{x_1}u(t,x_1,\dots,x_n)a(t,x_1,\dots,x_n))\rangle \\
&\quad + \sum_{i=2}^{n}\langle F_i(t,x_{i-1},\dots,x_n), D_{x_i}u(t,x_1,\dots,x_n) = -f(t,x_1,\dots,x_n), \quad (t,x_1,\dots,x_n) \in [0,T)\times \mathbf{R}^{nd} \\
&u(T,x_1,\dots,x_n) = g(x_1,\dots,x_n),\ (x_1,\dots,x_n) \in \mathbf{R}^{nd},
\end{aligned} \tag{2.7.1}$$

with $a = \sigma\sigma^*$, for given data $g : \mathbf{R}^{nd} \to \mathbf{R}$ and $f : [0,T]\times \mathbf{R}^{nd} \to \mathbf{R}$.

Let us discuss the weak Hörmander structure in this general case. It takes the following form: the noise propagates in the i^{th} component through the $(i-1)^{\text{th}}$, $2 \leq i \leq n$, thanks to the non-degeneracy of the gradients $\left(D_{x_{i-1}}F_i(t,\cdot)\right)_{i\in[\![2,n]\!]}$ (components which transmit the noise). In other words, up to a mollification of the diffusion coefficient, the drift is needed to span the space through Lie Bracketing. Moreover, the specific drift structure we consider here requires only one additional Lie bracket at each level of the chain to generate the corresponding directions, up to some regularization of the diffusion coefficient again. Eventually, we do not require the drift of each component to be smooth w.r.t. its other entries, but only to be Hölder continuous. This is the main reason why this framework falls under the regularization by noise theory.

Mathematically, this translates into the following assumptions:

(ML) The coefficients F and σ are measurable in time and $F(t,0,\dots,0)$ is bounded uniformly in time.

(UE) The diffusion matrix $a := \sigma\sigma^*$ is uniformly elliptic and bounded, uniformly in time, *i.e.* there exists $\Lambda \geq 1$ s.t. for all $t \geq 0$, $(\mathbf{x},\zeta) \in \mathbf{R}^{nd}\times\mathbf{R}^d$:

$$\Lambda^{-1}|\zeta|^2 \leq \langle a(t,\mathbf{x})\zeta,\zeta\rangle \leq \Lambda|\zeta|^2.$$

($\mathbf{H}_\eta$) For all $i \in [\![2,n]\!]$, there exists a closed convex subset $\mathcal{E}_{i-1} \subset GL_d(\mathbf{R})$ (set of invertible $d\times d$ matrices) s.t., for all $t \geq 0$ and $(x_{i-1},\dots,x_n) \in \mathbf{R}^{(n-i+2)d}$, $D_{x_{i-1}}F_i(t,x_{i-1},\dots,x_n) \in \mathcal{E}_{i-1}$. For example, $\mathcal{E}_{i-1}$ may be a closed ball included in $GL_d(\mathbf{R})$ the latter being an open set. Moreover, $D_{\mathbf{x}_{i-1}}F_i$ is η-Hölder continuous w.r.t. x_{i-1} uniformly in $x_{i:n}$ and time. Importantly, we assume as well that $D_{x_{i-1}}F_i$ is also bounded (which is automatically the case if the $\mathcal{E}_{i-1}$ are balls).

($\mathbf{T}_\beta$) There exists a family of coefficients $\beta = (\beta_i^j)_{1\leq i\leq j\leq n}$ each of them lying in $(0,1)$ such that for any $i \in [\![1,n]\!]$, $j \in [\![1,n]\!]$ the map $x_j \mapsto F_i(\cdot,x_j,\cdot)$ is uniformly β_i^j-Hölder continuous.

The results obtained in the works [6, 7, 8] generalize the ones previously sketched. Specifically, the following theorems were obtained in those works.

Theorem 2.7.1 ([6]). *Under the above assumptions, if the family of coefficients β satisfies*

$$\forall i \leq j \in [\![1,n]\!]^2, \quad \beta_i^j > (2i-3)/(2j-1),$$

and if moreover the coefficient σ is η-Hölder continuous, then, weak well-posedness hold for the system. The result remains true if F_1 is only bounded or in $L^q(\mathbf{R}_+, L^p(\mathbf{R}^{nd}))$ with $[n^2d/p]+[2/q] < 1$, $\quad p \geq 2, q > 2$.

To the best of our knowledge, one of the first results on the well-posedness of the system without Lipschitz assumption on the diffusion coefficient goes back to Menozzi [Men11] who extended the result to a spatially continuous diffusion coefficient in [Men18]. A particular configuration of the system was also considered in [Zha18]. In all cases, the results do not provide further information except for the diffusion coefficient. The results (and counter-examples) obtained in [6] have been successfully extend to the case of a perturbation by an α-stable process in [MM21]. All these results enter the weak heuristic rule of Chapter 1. Regarding the counter-example, it was built based on the study in [DF14]. It *a psoteriori* appears that similar arguments were used by Tanaka, Tsuchiya and Watanabe in [TTW74] to prove that uniqueness fails for non degenerate α-stable driven SDE when $\beta < 1 - \alpha$, which again emphasizes the heuristic.

Concerning the smoothing effect of the semi-group (Schauder estimates) the result we proved is the following.

Theorem 2.7.2 ([7]). *Let $\gamma \in (0,1)$. Assume the conditions of the previous theorem to hold and assume moreover that $a = \sigma\sigma^* \in L^\infty([0,T], C^\gamma_{b,\mathbf{d}}(\mathbf{R}^{nd}, \mathbf{R}^{d\otimes d}))$ where $C^\delta_{b,\mathbf{d}}(\mathbf{R}^{nd}, \mathbf{R})$, $\delta > 0$, denotes the anisotropic Hölder space associated with the homogeneous distance $\mathbf{d}$. Let f, g belong respectively to $C^{2+\gamma}_{b,\mathbf{d}}(\mathbf{R}^{nd}, \mathbf{R})$ and $L^\infty([0,T], C^\gamma_{b,\mathbf{d}}(\mathbf{R}^{nd}, \mathbf{R}))$. Then, the Cauchy problem $\mathscr{C}(T, f, g, \mathcal{L})$ admits a unique mild and weak solution u in $C^{2+\gamma}_{b,\mathbf{d}}(\mathbf{R}^{nd}, \mathbf{R})$ which verifies*

$$\|u\|_{L^\infty(C^{2+\gamma}_{b,\mathbf{d}})} \leq C(\|g\|_{C^{2+\gamma}_{b,\mathbf{d}}} + \|f\|_{L^\infty(C^\gamma_{b,\mathbf{d}})}). \tag{2.7.2}$$

for some constant $C := C(T) > 0$.

As far as we know, the first Schauder estimates for this model goes back to the seminal work of Lunardi [Lun97]. The estimates were obtained for homogeneous linear drift and diffusion coefficient satisfying some structural assumptions. Later, Lorenzi extended these results to the kinetic framework ($n = 2$) in [Lor05], by removing the structural assumptions and allowing for unbounded diffusion. Then, Priola considered the same model with an additional non-linear homogeneous drift coefficient for the non degenerate variable in [Pri09]. This last result has also been proved to hold for inhomogeneous drift (for the non degenerate variable) in the kinetic case by Imbert and Mouhot in [IM21]. Hence, in the current framework of degenerate Kolmogorov equations, focusing on the drift, the Schauder estimates was only proved, to the best of our knowledge, for either linear drifts or Hölder perturbations on the non degenerate variable of a linear drift. The result has also been considered for non-local operator (associated with α-stable noise) in [ZH20] in the kinetic case and [Mar20] in the general setting.

We eventually derived the following result on strong well-posedness.

Theorem 2.7.3 ([8]). *Under the above assumptions, if moreover the family of coefficients β satisfies*

$$\forall j \in [\![1, n]\!], \quad \beta^j_i > (2j-2)/(2j-1), \quad \forall i \in [\![1, n]\!]$$

and the coefficient σ is Lipschitz continuous in space uniformly in time, then, strong well-posedness hold for the system of perturbed chain of differential equations.

For strong regularization, one of the first result has been obtained in the kinetic case and was originally presented in the personal work [1]. This work was later revisited and extended in [WZ16, FFPV17] and eventually in [Pri21] from a path-by-path perspective. In the recent work

[Ger22], Gerencser studied the diagonal Peano perturbed system with fractional Brownian motion. Once again, all the results confirme the strong heuristic rule obtained in Chapter 1 for strong regularization.

2.8 (Some) perspectives

Before considering some perspectives w.r.t. the previous results, let us first mention that the somehow natural question on heat kernel estimate on the density of the process, under the assumptions of weak uniqueness, has been investigated in the work [16], for $n = 2$ (kinetic system) in in collaboration with S. Menozzi, A. Pesce and X. Zhang.

- **How to reach critical thresholds ?** To start with, let us comment on the assumptions on the coefficients as well as their optimality. We first notice that the assumptions on the η-Hölder regularity of the gradient in the non degenerate direction of the drift of the degenerate variable clearly appears to be artificial. The presence of a small regularity parameter only avoids to use harmonic analysis tools to remove it, as in this case, we would face to the integration of a critical singularity. Also, all the critical case for the weak and strong thresholds remain open as they also imply a critical singularity.

- **Is there a price to pay to pass from weak to strong uniqueness ?** A main question raised by the above results lies in the gap between the minimal regularity asked on the drifted to obtain weak and strong well-posedness. As mentioned in Chapter 1, all the results we found in the literature present the same gap, although the approach used mainly differs from the one we exposed here. It seems to be a quite challenging question. Firstly, because it is not clear at all that this gap is indeed a price to pay to pass from weak to strong well-posedness. As an example, in the scalar case ($d = 1$ and n is arbitrary), it can be easily proved that the system has a unique strong solution under the assumptions of weak well-posedness. Secondly, as far as we know, the only counter example to strong well-posedness, while weak well-posedness holds, based on the drift, is due to Tsirelson (see *e.g.* [Che05]) and relies on non markovian structure.

Interlude
Quantification of the smoothing effect of a drifted super-critical noise

In this interlude, we briefly describe the result obtained in [10] which are related to the smoothing effect of the semi-group associated with an α-stable noise in the so-called super-critical regime, *i.e.* for α in $(0,1)$. We investigate this effect through Schauder estimates, deriving such estimates for drifted fractional operators in the super-critical case. Furthermore, the drift is allowed to be unbounded.

This result is also linked to the weak heuristic rule in Chapter 1. As explained at the beginning of the previous chapter, to work with functions (*i.e.* with a coefficient $\beta > 0$) the weak heuristic rule forces to choose noises with fluctuations (or self-similarity index) of order $\gamma > 1$. For α-stable noises, this condition writes $1/\alpha > 1 \Leftrightarrow \alpha < 1$: this is the super-critical regime. As such, this framework presents some similarities with the previous works on degenerate Brownian noise exposed in Chapter 2. However, the noise is here non degenerate and possesses its own drawbacks, especially since the associated density does not integrate more moments than $\tilde{\alpha} < \alpha$.

The system

For a fixed time horizon $T > 0$, we investigate the smoothing effect of a drifted α-stable noise whose generator writes, for any smooth function $\varphi : \mathbf{R}^d \to \mathbf{R}$:

$$\forall t \in [0,T],\ \mathcal{L}_t^\alpha \varphi := \Delta^{\frac{\alpha}{2}} \varphi(x) + F(t,x) \cdot D\varphi(x),$$

where the drift $F : [0,T] \times \mathbf{R}^d \to \mathbf{R}^d$ is locally β-Hölder continuous and where $\Delta^{\alpha/2}$ denotes the fractional Laplacian.

As previously mentioned, we quantify the smoothing effect of the noise through Schauder estimates. Specifically, we aim to establish global Schauder estimates for the Cauchy problem $\mathscr{C}(T, f, g, \mathcal{L}^\alpha)$:

$$\begin{aligned} \partial_t u(t,x) + \Delta^{\frac{\alpha}{2}} u(t,x) + F(t,x) \cdot D_x u(t,x) &= -f(t,x), \quad \text{on } [0,T) \times \mathbf{R}^d, \\ u(T,x) &= g(x), \quad \text{on } \mathbf{R}^d, \end{aligned} \tag{2.8.1}$$

where the source $f : [0,T] \times \mathbf{R}^d \to \mathbf{R}$ and terminal condition $g : \mathbf{R}^d \to \mathbf{R}$ are assumed to belong to some suitable Hölder spaces and to be bounded.

We focus on the super-critical case, *i.e.* $\alpha \in (0,1)$, even though the estimates could be extended to the simpler case $\alpha \in [1,2)$. The difficulty lies in the fact that in Fourier space, the principal part of the operator $\mathcal{L}^\alpha$, namely $\Delta^{\alpha/2}$, is of order α and does not dominate the drift term, which

is roughly speaking of order one. From the PDE heuristic for the weak heuristic rule exposed in Chapter 1 (see the "PDE viewpoint" therein), we understand that the Hölder regularity "β" of the drift must compensate for the low smoothing effect of the operator $\Delta^{\alpha/2}$. This implies that β is such that $\beta+\alpha>1$. We also refer to the class of counter-examples in the previous chapter to have a probabilistic viewpoint on this constraint.

Strategy

To obtain the Schauder estimates for the Cauchy problem $\mathscr{C}(T,f,g,\mathcal{L}^\alpha)$, we implement the procedure based on parametrix expansion exposed in Chapter 2. Following this latter, the choice of the proxy to approximate $\mathcal{L}^\alpha$ is crucial, and following the degenerate case, the proxy must involve a (forward) frozen flow associated with the system (a flow of the ODE $dy_t = F(t,y_t)dt$). Indeed, as the fluctuations of the associated driftless process typically behave like $t^{1/\alpha}$, the transport of the initial condition, which is of order t, cannot be neglected in small time. This main particularity leads to similar issues to solve. Especially, we again must investigate the sensibility of the parametrix expansion w.r.t. the frozen flow.

In [7], the main difficulties encountered were in handling the degeneracy of the operator and its associated anisotropic behavior while the derivation of a Duhamel representation (or first order parametrix expansion) as well as the existence of a solution were the easier parts. We here face different problems, especially when trying to obtain a suitable Duhamel representation or when dealing with the existence part. These difficulties arise from two main features of our framework: the stable operator $\Delta^{\alpha/2}$ induces major integrability issues, and we consider drift terms that are only locally Hölder continuous. To overcome these particularities, we localize the Cauchy problem $\mathscr{C}(T,f,g,\Delta^{\alpha/2},F)$ by multiplying all the coefficients by a suitable *localizing* test function.

Comments and general result

The main result we obtain is the following theorem.

Theorem 2.8.1 ([10]). *There exists a class of functions $\mathscr{S}_b^{\alpha+\beta}([0,T]\times\mathbf{R}^d)$ that contains the set $\mathcal{C}_b^{\alpha+\beta}([0,T]\times\mathbf{R}^d)$ on which the Cauchy problem has a unique solution satisfying:*

$$\|u\|_{L^\infty([0,T],C_b^{\alpha+\beta})} \leq C(\|g\|_{C_b^{\alpha+\beta}} + \|f\|_{L^\infty([0,T],C_b^\beta)}). \tag{2.8.2}$$

An interesting example covered by these assumptions is the non-local Ornstein-Uhlenbeck process with generator:

$$\triangle^{\alpha/2}\varphi(x) + Ax\cdot D_x\varphi(x), \tag{2.8.3}$$

with $F(t,x)=Ax$ and A is any $d\times d$ real matrix. If $\alpha=2$, the Schauder estimates were first proved by Da Prato and Lunardi [DPL95]. After that paper the O.U. operator has been much investigated as a prototype of operator with unbounded coefficients.

Schauder estimates in the α-stable non-local framework have been addressed by several authors. For driftless operators or in the case $\alpha\geq 1$ (but possibly including a general diffusion coefficients), we can refer to *e.g.*[Bas09], [MP14], [IJS]. In the particular elliptic setting, when $\alpha\in[1,2)$ and L^α is a non degenerate symmetric α-stable operator and for bounded Hölder drifts, global Schauder estimates were obtained by Priola, see *e.g.* Section 3 in [Pri12b] and [Pri18b] with respective

applications to the strong well-posedness and Davie's uniqueness for the corresponding SDE. Also, when $\alpha \in [1, 2)$, elliptic Schauder estimates can be proven for more general Lévy-type generators that are invariant for translations, see Section 6 in [Pri18b].

There is also a rather large literature concerning the regularity of the solution of the Cauchy problem when the drift (possibly depending on the solution) is divergence free, *i.e.* $\nabla \cdot F(t, x) = 0$. We can refer *e.g.* to [CV10] or [SVZ13].

However, fewer result were available in the drifted case when $\alpha \in (0, 1)$. When F is Hölder continuous and bounded, Silvestre obtained in [Sil12] sharp Schauder estimates on balls for the fractional Laplacian. Nevertheless, it seems that the above result when $L^\alpha = \Delta^{\alpha/2}$ cannot be obtained from the result by Silvestre using a standard covering argument for a locally Hölder drift F. Indeed the Schauder constant in [Sil12] also depends on the global boundedness of F. We can mention as well the recent work of Zhang and Zhao [ZZ18], who address, through probabilistic arguments, the parabolic Dirichlet problem in the super-critical case for stable-like operators with a non trivial bounded drift. They also obtain interior Schauder estimates and some boundary decay estimates. In the whole space, related estimates in Besov spaces for bounded drifts and potentially singular spherical measures can also be found in Chen *et al.* [CZZ17].

Chapter 3

Weak-regularisation for SDE with distributional drift

In this chapter, we focus on the results obtained in [13]. Therein, we consider the weak well-posedness of a formal SDE *i.e.* a SDE with distributional drift. We investigate the weak well-posedness through the martingale formulation, modifying slightly the original Stroock and Varadhan perspective in order to avoid the issues coming from the distributional drift. Then, we link the martingale solutions to the dynamics suggested by the formal SDE.

This result is thus part of the "regularization by noise" perspective, as defined in Chapter 1. In this latter, we derive the weak heuristic rule which somehow suggests that we could solve, in a weak sense, formal SDE with distributional drift (with regularity parameter $\beta < 0$), provided the noise scales at least at order $\gamma < 1$. The purpose of this work is to investigate what can be said in this regime, considering α-stable Markovian noise for α in $(1, 2]$ (which scales at order $1/\alpha$). The Markovian setting allows us to use the connection with PDEs, and the results, such as those presented in Chapter 2, strongly rely on PDE theory. In this regard, it should be emphasized that the forward parametrix approach, outlined in previous parts, plays a central role, although it is implemented in a simpler way due to the structure of the problem. This enables us to focus on other issues specific to the distributional framework, especially concerning the meaning to give to the dynamics of the solution.

3.1 The formal formulation

The noise The class of noises we consider is the class of d-dimensional symmetric α-stable process $\mathcal{W}$, for some α in $(1, 2]$ (thus including Brownian noise). The semi-group generated by the noise is denoted by $(P_t^\alpha)_t$, and its generator is denoted by $\mathcal{L}^\alpha$. When $\alpha = 2$, $L^2 = (1/2)\Delta$ where Δ stands for the usual Laplace operator on $\mathbf{R}^d$. In the pure-jump stable case $\alpha \in (1, 2)$, for all $\varphi \in C_0^\infty(\mathbf{R}^d, \mathbf{R})$:

$$\mathcal{L}^\alpha \varphi(x) = \text{p.v.} \int_{\mathbf{R}^d} [\varphi(x + z) - \varphi(x)]\nu(dz),$$

where, writing in polar coordinates $z = \rho\xi$, $(\rho, \xi) \in \mathbf{R}_+ \times \mathbb{S}^{d-1}$, the Lévy measure decomposes as $\nu(dz) = \mu(d\xi)d\rho/\rho^{1+\alpha}$ with μ a symmetric non degenerate measure on the sphere $\mathbb{S}^{d-1}$. We assume:

(UE) There exists $\kappa \geq 1$ s.t. for all $\lambda \in \mathbf{R}^d$:

$$\kappa^{-1}|\lambda|^\alpha \leq \int_{\mathbb{S}^{d-1}} |\lambda \cdot \xi|^\alpha \mu(d\xi) \leq \kappa|\lambda|^\alpha.$$

The formal SDE We are interested in providing a well-posedness theory for the following formal d-dimensional stable driven SDE. For a fixed $T > 0$, $t \in [0, T]$:

$$X_t = x + \int_0^t F(s, X_s)ds + \mathcal{W}_t, \qquad \text{(formal SDE)}$$

The crucial point in the above SDE lies in the fact that the drift F is only supposed to belong to the space $L^r([0,T], B^\beta_{p,q}(\mathbf{R}^d, \mathbf{R}^d))$, where $B^\beta_{p,q}(\mathbf{R}^d, \mathbf{R}^d)$ denotes a Besov space. In a nutshell, when $p = q = \infty$, for any non integer $\beta > 0$, Besov spaces coincide with Hölder spaces $B^\beta_{\infty,\infty}(\mathbf{R}^d, \mathbf{R}^d) = \mathcal{C}^\beta_b(\mathbf{R}^d, \mathbf{R}^d)$; when $\beta < 0$, this somehow indicates that the Hölder modulus blows up at rate β. The parameters p and q are related to the integrability of such a modulus. We refer to Section 2.6.4 of [Tri83] for a rigorous definition. The parameters (p, q, β, r) are allowed to be s.t. $-1/2 < \beta < 0$, $p, q, r \geq 1$. Assuming the parameter β to be strictly negative implies that F may not even be a function, but just a distribution, so it is not clear that the integral part in the above equation has any meaning, at least as it is written. This is why, at this stage, we refer to it as the formal d-dimensional stable SDE or "the formal SDE".

The formal PDE As done in the previous part, our approach to tackle the solvability (in a sense to be precised) of the formal SDE turns into investigating the smoothing effect (and hence the well-posedness) of the associated PDE. Here, the Cauchy Problem $\mathscr{C}(T, f, g, \mathcal{L}^\alpha, F)$ associated with writes

$$(\partial_t + F \cdot D + L^\alpha)u = f, \quad u_T = g, \qquad \text{(formal Cauchy problem)}$$

for f and g in large enough classes $\mathcal{F}$ and $\mathcal{G}$ respectively. We again use the term formal as it is not clear wether the above product $F \cdot Du$ is meaningful, since product between two distributions can only be defined under suitable constraints.

3.2 Rigorous formulation of the problem and results for drift as generalized derivative of a Hölder map

From now on, we assume that the parameters satisfy $p, q, r = +\infty$, in order to focus on the parameter β. In this case, as $B^\beta_{\infty,\infty} = C^\beta_b$ can be identified with Hölder space, the drift F is the generalized derivative of a function $\mathbf{F}$ belonging to $B^{\beta+1}_{\infty,\infty} = C^{\beta+1}_b$.

Rigorous formulation for the PDE As at this stage everything remains formal, we need to define appropriate notions to work with. The first one relies on the PDE. The product to handle, between the gradient of the PDE solution and the distributional drift, suggests to work with mild formulation of the problem in order to benefit from the smoothing effect of the semi-group generated by the noise. The natural candidate for the mild solution or, in comparison with Chapter 2, the parametrix expansion around the generator of the noise, formally gives that the mild solution of the formal Cauchy problem writes

$$\forall (t, x) \in [0, T] \times \mathbf{R}^d, \quad u(t, x) = P^\alpha_{T-t}[g](x) - \int_t^T ds P^\alpha_{s-t}[\{f - F \cdot Du\}](s, x).$$

We then need to specify the space in which we are looking for such a solution. Especially, we must take care to ensure that: (i) the space is contained in the space of functions v for which the product

$F \cdot Dv$ makes sense as a distribution (hence with the same regularity as F); (ii) the semi-group indeed maps distribution with the same regularity as F... onto this space.

To do so, we know that for any $\delta > 0$, the semi-group P_t^α maps C_b^δ to $C_b^{\delta+\alpha}$ *w.c.s.*. In our work, we also manage to prove that this remains valid for any $0 > \delta > -1/2$ (up to an infinitesimal loss of regularization). Therefore, we have that for any $t > 0$, $P_t^\alpha F$ belongs to $C_b^{\alpha+\beta}$ *w.c.s.*, which somehow settles point (ii). On the other hand, we need the sum of the regularity indexes of F and Du to be strictly positive in order to define the product between them thanks to the Bony's paraproduct rule. Combining this with the previous point, we obtain the condition $\{\beta\} + \{\beta + \alpha - 1\} > 0 \Leftrightarrow \beta > (1-\alpha)/2$, as if $P_t^\alpha F$ belongs to $C_b^{\alpha+\beta}$ *w.c.s.* then $DP_t^\alpha F$ belongs to $C_b^{\alpha+\beta-1}$ *w.c.s.*

Hence, for maps $f : \mathbf{R}_+ \times \mathbf{R}^d \to \mathbf{R}$ and $g : \mathbf{R}^d \to \mathbf{R}$ and $T > 0$, we define the solutions to the formal Cauchy problem $\mathscr{C}(T, f, g, \mathcal{L}^\alpha, F)$ as the mild solutions u belonging to $\mathcal{C}^{0,1}([0,T] \times \mathbf{R}^d, \mathbf{R})$ with Du in $\mathcal{C}_b^0([0,T], B^\theta_{\infty,\infty}(\mathbf{R}^d, \mathbf{R}^d))$ where $\theta + \beta > 0$ (we mainly have in mind $\theta = \alpha + \beta - 1$ up to a small loss $0 < \epsilon << 1$).

Solving the PDE Given this definition, it remains to solve the PDE. To this aim, we implement the strategy outlined in Chapter 2 to handle both Schauder estimates and Strong regularization. The main idea to handle the distributional case is to work within the framework of Besov spaces. Especially, we use as crucial tools duality results between Besov spaces as well as their thermic characterization. This is done by working on the Cauchy problem with mollified coefficients and large class of data. By doing so, we obtain the following result which gives an affirmative answer to the well-posedness, in the sense previously defined, to the formal Cauchy problem $\mathscr{C}(T, f, g, \mathcal{L}^\alpha, F)$, through a compactness argument.

Theorem 3.2.1. *Let $\beta > (1-\alpha)/2$. For all f in $\mathcal{C}([0,T], B^\beta_{\infty,\infty}(\mathbf{R}^d, \mathbf{R}))$ and $g \in \mathcal{C}^1(\mathbf{R}^d, \mathbf{R})$ with $Dg \in B^{\alpha+\beta-1}_{\infty,\infty}(\mathbf{R}^d, \mathbf{R}^d)$ the formal Cauchy problem $\mathscr{C}(T, f, g, \mathcal{L}^\alpha, F)$ admits a unique mild solution. Moreover it satisfies that for all $(t \leq s)$ in $[0,T]^2$, x in $\mathbf{R}^d$:*

$$|u(t,x) - u(s,x)| \leq C|t-s|^{\frac{\alpha+\beta}{\alpha}}, \quad |Du(t,x) - Du(s,x)| \leq C|t-s|^{\frac{\alpha+\beta-1}{\alpha}}.$$

Rigorous definition of the martingale problem Having a well-posedness theory for the PDE at hand allows us, in turn, to rigorously define the martingale problem. However, the main drawback in our setting lies in the explicit use of the generator in the definition of the martingale problem from Stroock and Varadhan's theory. The generator is required to act on a sufficiently large class of functions, $\mathcal{E}$, which must be chosen to be rich enough to characterize a Markov process through the martingale formulation (see, *e.g.*, [SV79]). In our current setting, the main issue arises from the fact that the generator includes a distributional part (the drift term F). Even if $\mathcal{E}$ is chosen so that the products in $(F \cdot D + L^\alpha)\phi$ are well defined for $\phi \in \mathcal{E}$, this term could only be a distribution with the same regularity as F. To avoid this issue, the idea is to take $\mathcal{E}$ as the set of functions for which $(\partial_t + F \cdot D + L^\alpha)\phi$ is a well-defined function.

Having these considerations in mind, we rewrite the Martingale Problem associated with $(F, \mathcal{L}^\alpha, x)$ for $x \in \mathbf{R}^d$ as follows: find a probability measure $\mathbf{P}^\alpha$ on the space Ω_α (with $\Omega_2 = \mathcal{C}([0,T], \mathbf{R}^d)$ and $\Omega_\alpha = \mathcal{D}([0,T], \mathbf{R}^d)$ when $1 < \alpha < 2$) equipped with its canonical filtration so that

(i) $\mathbf{P}^\alpha(X_0 = x) = 1$,

(ii) $\forall f \in \mathcal{C}([0,T], \mathcal{S}(\mathbf{R}^d))$, $g \in \mathcal{C}^1(\mathbf{R}^d, \mathbf{R})$ with $Dg \in B^{\alpha+\beta-1}_{\infty,\infty}(\mathbf{R}^d, \mathbf{R}^d)$,

$$\left(u(t, X_t) - \int_0^t f(s, X_s)ds - u(0, x)\right)_{0\le t\le T}$$

is a (square integrable if $\alpha = 2$) martingale under $\mathbf{P}^\alpha$ where u is the mild solution of the Cauchy Problem $\mathscr{C}(T, f, g, \mathcal{L}^\alpha, F)$[1].

From the Cauchy to the Martingale problem (i) The first step of the procedure is to mollify the coefficients to ensure that the mollified SDE is well-posed. We then solve the (mollified) Cauchy problem $\mathscr{C}(T, -F_m, 0, \mathcal{L}^\alpha, F_m)$, where the parameter m is associated to the mollifying procedure, and remove the drift by using the Zvonkin transform. This allows in turn to derive tightness of the sequence of laws induces by the solutions on the canonical space Ω_α.

(ii) We identify the limit, denoted by $\mathbf{P}^\alpha$, as a solution of the martingale problem. This follows from the expansion of the solution of the mollified Cauchy problem $\mathscr{C}(T, f, g, \mathcal{L}^\alpha, F_m)$ along the solution of the mollified SDE. Namely, when *e.g.* $\alpha = 2$, this gives that for each m,

$$u_m(t, X^m_t) - u_m(0, x_0) - \int_0^t f(s, X^m_s)ds = \int_0^t Du_m(r, X^m_r) \cdot dB_r.$$

Then, using the convergence of $(u_m, Du_m)_{m\ge1}$ to the solution (u, Du) of $\mathscr{C}(T, f, g, \mathcal{L}^2, F)$ on every compact subsets of $[0,T] \times \mathbf{R}^d$ together with a uniform control of the moment of X^m, we deduce that

$$\left(u(t, X_t) - \int_0^t f(s, X_s)ds - u(0, x)\right)_{0\le t\le T}, \tag{3.2.1}$$

is a $\mathbf{P}^2$ square integrable martingale. This is where we use that the gradients of u_m and u are uniformly bounded.

(iii) We then turn to prove uniqueness. To do so, we come back to the canonical space Ω_α and let $\mathbf{P}^\alpha$ and $\tilde{\mathbf{P}}^\alpha$ be two solutions of the Martingale Problem associated with $(F, \mathcal{L}^\alpha, x)$, $x \in \mathbf{R}^d$. Hence, for all $f \in C^0([0,T], \mathcal{S}(\mathbf{R}^d))$ it holds, denoting by u the solution of the Cauchy problem $\mathscr{C}(T, f, 0, \mathcal{L}^\alpha, F)$,

$$-u(0, x) = \mathbf{E}^{\mathbf{P}^\alpha}\left[\int_0^T f(s, X_s)ds\right] = \mathbf{E}^{\tilde{\mathbf{P}}^\alpha}\left[\int_0^T f(s, X_s)ds\right],$$

thanks to Itô's formula. Therefore, the marginal laws of the canonical process are the same under $\mathbf{P}^\alpha$ and $\tilde{\mathbf{P}}^\alpha$. Classical tools allow to obtain the uniqueness and the (strong) Markov property of the solution.

We have thus proved.

[1] We chose to work with $\mathcal{E} = \mathcal{S}(\mathbf{R}^d)$, where $\mathcal{S}$ stands for the class of Schwartz functions. This is mainly motivated by the fact that such a class is rich enough to characterize the law of the Markov process and also by the fact that our approach is based on Besov spaces. Indeed, this class is continuously embedded into any Besov space $B^s_{l,m}(\mathbf{R}^d, \mathbf{R})$, $s \in \mathbf{R}$, $1 \le l, m \le \infty$, see *e.g.* paragraph 2.3.3 in [Tri83]. Note that it is also dense in $B^s_{l,m}(\mathbf{R}^d, \mathbf{R})$ provided $l, m < \infty$.

Theorem 3.2.2. *Let $\alpha \in (1,2]$ and $\beta > (1-\alpha)/2$. Then, the Martingale Problem associated with (F, L^α, x) for $x \in \mathbf{R}^d$, is well-posed in the above sense. Moreover, the canonical process under $\mathbf{P}^\alpha$ is strong Markov.*

💬 *Compared to the weak heuristic rule (which here writes $\beta > 1-\alpha$), the threshold obtained above appears to be more restrictive: this comes from the distributional setting investigated here and the perturbative/mild approach used to solve the PDE which imposes to define properly the product $F(t,\cdot)Du(t,\cdot)$. As this requires the sum of the regularity exponent of the two terms involved in the product to be strictly positive, we mainly obtain that $\beta + \alpha + \beta - 1 > 0$, which is stronger than the weak heuristic rule.*

3.3 Meaning of the dynamics for a drift as generalized derivative

At this stage, we built a martingale solution associated to the data of the formal SDE. Nevertheless, we said nothing on the dynamics of the process[2]. In the work currently exposed in this chapter, we managed to specify in which way the process evolves. We recall that we still work within the framework of the previous section *i.e.* with a drift component obtained as a generalized derivative of $(\beta+1)$-Hölder map.

Building the dynamics The next step consists in linking the Martingale Problem and the formal SDE. Usually, *i.e.* for sufficiently smooth drifts, the starting point consists in building the noise from the canonical process and the drift. Here, the challenge lies into building the (a) drift, which requires to have a noise at hand (otherwise the problem is ill-posed). To do so, the strategy consists in building simultaneously the Martingale solution and the noise $(X, \mathcal{W})$ as the solution of a kind of enlarged Martingale Problem. This allows in turn to build the drift as the difference between them. Indeed, having such a canonical process at hand, we decompose the increment of the process X as

$$X_{t+h} - X_t = \{\mathbf{E}[X_{t+h} - X_t | \mathcal{F}_t]\} + \{X_{t+h} - X_t - \mathbf{E}[X_{t+h} - X_t | \mathcal{F}_t]\}, \quad 0 < h << 1,$$

where $\mathcal{F}_t := \sigma(X_s, \mathcal{W}_s,\ 0 \le s \le t)$. Clearly, the first difference in the above right hand side stands for a drift term, while the second stands for a martingale part. It thus "suffices" to relate both parts with (i) the original drift F and (ii) the α-stable noise $\mathcal{W}$ which we have now at hand.

To do so, the starting point is the Zvonkin-like transform: for the mild solution of $\mathscr{C}(F, L^\alpha, 0, x, t+h)$, we can write $X_{t+h} - X_t = u(t+h, X_{t+h}) - u(t+h, X_t)$ and using then Itô's formula (up to mollification argument) to obtain that,

$$X_{t+h} - X_t = u(t, X_t) - u(t+h, X_t) + \int_t^{t+h} \int_{\mathbf{R}^d \setminus \{0\}} \{u(r, X_{r^-} + x) - u_m(r, X_{r^-})\} d\tilde{N}(r, x)$$

where $\tilde{N}$ is the compensated Poisson measure associated with the stable noise and where we further assume that $\alpha < 2$ to avoid the dichotomy when writing the martingale part.

[2] Note indeed that the pathological case under consideration breaks the equivalence between weak and martingale solution.

(i) To obtain the drift, we take conditional expectation and obtain that $\mathbf{E}[X_{t+h} - X_t|\mathcal{F}_t] = u(t, X_t) - u(t+h, X_t)$. From the mild representation of u we then derive, expanding one more time the gradient in the mild representation,

$$\begin{aligned}
& \mathbf{E}[X_{t+h} - X_t|\mathcal{F}_t] \\
= \; & u(t, X_t) - u(t+h, X_t) = \int_t^{t+h} ds P^\alpha_{s-t}[F \cdot Du](s, X_t) \\
= \; & \int_t^{t+h} ds P^\alpha_{s-t}[F](s, X_t) + \int_t^{t+h} ds P^\alpha_{s-t}\Big[F \cdot \Big\{ \int_t^s dr P^\alpha_{r-t}[F \cdot Du](r, X_t) \Big\}\Big](s, X_t).
\end{aligned}$$

As for any $t > 0$, P^α_t maps C^β_b to $C^{\beta+\alpha}_b$ *w.c.s.* we understand that the first term in the last equality is or order $h^{(\beta+\alpha)/\alpha}$ while the second has better regularity, due to the successive applications of the semi-group, and is thus of order $h^{1+\varepsilon}$ for some $\varepsilon > 0$. This hints the shape of the drift: in small time, only the first term matters and the drift thus looks like the convolution of the original distributional drift with the density of the noise.

💭 *The idea of iterating the expansion is reminiscent from the parametrix approach. It is indeed usual to iterate the expansion to obtain an infinite series, avoiding thus the implicit representation. The crucial point being that the iteration of the parametrix kernel gives better smoothing effect so that the series indeed converges.*

(ii) To obtain the noise, we subtract the conditional expectation of the increment and get that

$$\begin{aligned}
& X_{t+h} - X_t - \mathbf{E}[X_{t+h} - X_t|\mathcal{F}_t] = \int_t^{t+h} \int_{\mathbf{R}^d \setminus \{0\}} \{u(r, X_{r^-} + x) - u_m(r, X_{r^-})\} d\tilde{N}(r, x) \\
= \; & \int_t^{t+h} \int_{\mathbf{R}^d \setminus \{0\}} \{u(r, X_{r^-} + x) - u_m(r, X_{r^-}) - x\} d\tilde{N}(r, x) + (\mathcal{W}_{t+h} - \mathcal{W}_t).
\end{aligned}$$

From the smoothness of $u(t, \cdot)$ (which almost lies in $C^{\beta+\alpha}$) together with Taylor expansion, we can see that the first term in the last equality have fluctuations of order $h^{1/\alpha + (\beta+\alpha-1)/\alpha}$ (up to a small loss) while the last one has fluctuations of order $h^{1/\alpha}$. Thus, in small time, only the increment $\mathcal{W}_{t+h} - \mathcal{W}_t$ matters. This allows to identify the noise.

At this stage, we deduce that the infinitesimal dynamics of X should reads as

$$dX_t = \mathscr{F}(t, X_t, dt) + d\mathcal{W}_t, \text{ with } \forall x \in \mathbf{R}^d, \forall r, s, \mathscr{F}(r, x, s-r) = \int_r^s dv P^\alpha_{s-r}[F](r, X_r)$$

The result we obtained give an even better description as we were able to define a stochastic calculus, where stochastic integrals are in the spirit of Young (even more in the spirit of non-linear Young integral) called stochastic non-linear Young integral. The above description also provides a natural candidate for being a weak solution. This latter fact, as well as other properties, are summarized in the resume below, which hold in the general case.

3.4 Comments and general results

💭 *We encourage the reader to regularly choose $p = q = r = \infty$ and $\alpha = 2$ to simplify the statements (especially the conditions) and make the connection with the previous part.*

We start with the condition assumed on the parameter p, q, r and β related to the Lebesgue-Besov space $L^r(B^\beta_{p,q})$ where the drift is assumed to live. We say that these parameters satisfy a good relation if

$$p, q, r \geq 1, \quad \alpha \in \left(\frac{1 + [d/p]}{1 - [1/r]}, 2\right], \quad \beta \in \left(\frac{1 - \alpha + [d/p] + [\alpha/r]}{2}, 0\right). \tag{good relation}$$

We also set up once and for all the quantity

$$\theta := \beta + \alpha - \frac{d}{p} - \frac{\alpha}{r}, \tag{Df(θ)}$$

related with the smoothing effect of the noise as it will appear many times below.

Theorem 3.4.1. *Let $p, q, r \geq 1$, $\alpha \in (1, 2]$ and $\beta \in (-1/2, 0)$ satisfy a good relation. For all f in $\mathcal{C}([0, T], B^{\theta-\alpha}_{\infty,\infty}(\mathbf{R}^d, \mathbf{R}))$ and $g \in \mathcal{C}^1(\mathbf{R}^d, \mathbf{R})$ with $Dg \in B^{\theta-1}_{\infty,\infty}(\mathbf{R}^d, \mathbf{R}^d)$, where θ is given by (Df(θ)), the formal Cauchy problem $\mathscr{C}(F, L^\alpha, f, g, T)$ admits a unique mild solution. Moreover it satisfies that for all $(t \leq s)$ in $[0, T]^2$, x in $\mathbf{R}^d$:*

$$\begin{aligned} |u(t, x) - u(s, x)| &\leq C|t - s|^{\frac{\theta}{\alpha}}, \\ |Du(t, x) - Du(s, x)| &\leq C|t - s|^{\frac{\theta-1}{\alpha}}. \end{aligned}$$

We can therefore define the associated Martingale Problem and the corresponding well-posedness result.

Theorem 3.4.2. *Let $p, q, r \geq 1$, $\alpha \in (1, 2]$ and $\beta \in (-1/2, 0)$ satisfy a good relation. Then, the Martingale Problem associated with (F, L^α, x) for $x \in \mathbf{R}^d$, is well-posed. Moreover, the canonical process under $\mathbf{P}^\alpha$ is strong Markov.*

For Brownian noise, a same kind (the drift was supposed to be in fractional Sobolev spaces) of result was obtained by Flandoli, Issoglio and Russo [FIR17]. This latter results extended to the inhomogeneous setting previous results of Bass and Chen [BC01] and Flandoli, Russo and Wolf, [FRW03], [FRW04]). We can also refer to Zhang and Zhao [ZZ17]. In the pure-jump setting, this result has been obtained for homogeneous drift as generalized derivative of Hölder function and in the scalar case by Athreya, Butkovski and Mytnik [ABM20]. In either [FIR17], [ZZ17] or [ABM20], for homogeneous drift as generalized derivative of Hölder function, the threshold for weak well-posedness to hold is $\beta > (1 - \alpha)/2$, which fits ours.

As already discussed, it is not possible to go beyond such thresholds in full generality, as Bony's paraproduct rule comes into play. To bypass such limit, Delarue and Diel [DD15] added some rough path structure to the drift and handled the case of a drift in $L^\infty(B^{(-2/3)^+}_{\infty,\infty})$ in the scalar case. The multi-dimensional case has been done with paracontrolled calculus by Cannizzaro and Choukh in [CC18]. This latter setting has been extending recently in the pure jump case by Kremp and Perkowski in [KP22].

Building the dynamics: general case In the following, we say that $p, q, r \geq 1$, $\alpha \in (1, 2]$ and $\beta \in (-1/2, 0)$ satisfy a good relation for the dynamics if the following relation holds:

$$p, q, r \geq 1, \quad \alpha \in \left(\frac{1 + [d/p]}{1 - [1/r]}, 2\right], \quad \beta \in \left(\frac{1 - \alpha + [2d/p] + [2\alpha/r]}{2}, 1\right). \tag{good relation for the dynamics}$$

We emphasize that the above condition is slightly more constraining than the previous one when p and/or r is not $+\infty$.

We start by defining the non-linear stochastic Young integral. This definition mainly comes from the work [DD15] which somehow extends to the stochastic framework the non-linear Young integral introduced in [CG16].

Definition 3.4.3. *Let $\tau > 0$, $(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{0\le t\le\tau}, \tilde{\mathbf{P}})$ be a filtered probability space and let $(\psi_t)_{0\le t\le\tau}$ be a progressively measurable process on it. Let $(A(s,t))_{0\le s\le t\le\tau}$ be a continuous and progressively measurable map in the sense that for any $0 \le s \le t$,*

$$\tilde{\Omega} \times \{s' \in [0,s],\ t' \in [0,t],\ s' \le t'\} \ni (\omega, s', t') \mapsto A(s', t')$$

is $\tilde{\mathcal{F}}_t \otimes \mathcal{B}(\{s' \in [0,s],\ t' \in [0,t],\ s' \le t'\})$ measurable and

$$\{s' \in [0,\tau],\ t' \in [0,\tau],\ s' \le t'\} \ni (s,t) \mapsto A(s,t)$$

is continuous. For $\ell \ge 1$, we call L^ℓ-stochastic non-linear Young integral of ψ with respect to the pseudo increment A the limit in $L^\ell(\tilde{\Omega}, \tilde{\mathbf{P}})$

$$\lim_{\substack{\Delta \text{ partition of } [0,\tau] \\ |\Delta|\to 0}} \sum_{t_i \in \Delta} \psi_{t_i} A(t_i, t_{i+1}) =: \int_0^\tau \psi_t A(t, t+dt),$$

when it exists.

We are then in position to state the following theorem which concerns the dynamics of the solution $(X, \mathcal{W})$ of the enlarged martingale problem.

Theorem 3.4.4. *For $\alpha \in (1,2]$ and $\beta \in (-1/2, 0)$ satisfying a good relation for the dynamics it holds that there exists a probability measure $\mathbf{P}^\alpha$ on $\mathcal{C}([0,T], \mathbf{R}^{2d})$ when $\alpha = 2$ and $\mathcal{D}([0,T], \mathbf{R}^{2d})$ when $1 < \alpha < 2$ such that the canonical process, denoted by $(X, \mathcal{W})$, satisfies*

(i) The law of X under $\mathbf{P}^\alpha$ is a solution of the Martingale problem associated with (F, L^α, x), $x \in \mathbf{R}^d$ and the law of $\mathcal{W}$ under $\mathbf{P}^\alpha$ is a Brownian motion if $\alpha = 2$ and an α-stable process with generator L^α if $\alpha < 2$.

(ii) The dynamics of the canonical process reads

$$X_t = x + \int_0^t \mathscr{F}(s, X_s, ds) + \mathcal{W}_t, \quad \mathbf{P}^\alpha - a.s. \tag{dynamics}$$

where for any $0 \le v \le s \le T$, $x \in \mathbf{R}^d$,

$$\mathscr{F}(v, x, s-v) \ := \ \int_v^s dr \int_{\mathbf{R}^d} dy F(r,y) p_\alpha(r-v, y-x),$$

with p_α the (smooth) density of $\mathcal{W}$ and where the integral in the dynamics is understood as an L^ℓ-stochastic non-linear Young integral, for any $1 \le \ell < \alpha$, in the sense of the above definition.

As already mentioned, we prove a stronger result concerning the dynamics. We show that it is possible to define a stochastic non-linear Young integral against the process, leading in turn to use Itô calculus. This is done for a suitable class of progressively measurable processes ψ which we now describe. For any $\mathfrak{q}' \geq 1$, any $\delta \in (0,1)$, we set

$$\mathcal{H}^{\delta}_{\mathfrak{q}'}(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P}) := \Big\{ (\psi_t)_{t\in[0,T]} \text{ progressively measurable}, \tag{3.4.1}$$
$$\sup_{t\in[0,T]} \mathbf{E}^{\frac{1}{\mathfrak{q}'}}[|\psi|^{\mathfrak{q}'}] + \sup_{s\neq t\in[0,T]} \frac{\mathbf{E}^{\frac{1}{\mathfrak{q}'}}[|\psi_s - \psi_t|^{\mathfrak{q}'}]}{|t-s|^{\delta}} < +\infty \Big\}.$$

Corollary 3.4.5 (Associated L^{ℓ}-stochastic non-linear Young integral, $1 \leq \ell < \alpha$). *Under the above assumptions, one can define a stochastic non-linear Young integral w.r.t. the quantities in appearing in the dynamics. Namely, for any $1 \leq \ell, \mathfrak{q} < \alpha$, for which there exists $\mathfrak{q}' \geq 1$ satisfying $1/\mathfrak{q}' + 1/\mathfrak{q} = 1/\ell$, one has*

$$\int_0^t \psi_s dX_s = \int_0^t \psi_s \mathscr{F}(s, X_s, ds) + \int_0^t \psi_s d\mathcal{W}_s,$$

for any $\psi \in \mathcal{H}^{1-1/\alpha-\varepsilon_2}_{\mathfrak{q}'}$, for all $0 < \varepsilon_2 < (\theta-1)/\alpha$ and where the first term in the above right hand side is defined as an L^{ℓ}-stochastic non-linear Young integral.

For SDE with distributional drift, only few informations are usually given on the dynamics: the results may relies only the martingale formulation, at the level of the Zvonkin transform, [FIR17]; in terms of Dirichlet process, [ZZ17] and [ABM20]; as the limit of any smooth mollification, [ZZ17]. At the end of the day, the shape of the drift remains very obscur. To the best of our knowledge, this accurate description is due to [DD15] through the clever analysis of the increment exposed before. Our results clearly rely on them, and sometimes precise / clarify some associated questions.

Further properties and weak formulation. The previously described dynamics for the Martingale solution strongly suggests that a notion of weak solution associated with the formal SDE can be considered. This leads to the following definition.

Definition 3.4.6. *We call weak solution of the formal SDE a pair $(Y, \mathcal{Z})$ of adapted processes on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbf{P})$ such that $\mathcal{Z}$ is an $\{\mathcal{F}_t\}_{t\geq 0}$ α-stable process and $(Y, \mathcal{Z})$ satisfies*

$$Y_t = x + \int_0^t \mathscr{F}(s, Y_s, ds) + \mathcal{Z}_t, \quad \mathbf{P}-a.s., \qquad \mathbf{E}|\int_0^t \mathscr{F}(s, Y_s, ds)| < +\infty \tag{3.4.2}$$

for any t in $[0,T]$ and where for any $0 \leq v \leq s \leq T$, $x \in \mathbf{R}^d$,

$$\mathscr{F}(v, x, s-v) \;=\; \int_v^s dr \int_{\mathbf{R}^d} dy F(r,y) p_{\alpha}(r-v, y-x)$$

with p_{α} the (smooth) density of $\mathcal{Z}$.

We say that weak uniqueness holds for the formal SDE if for any two weak solutions $(Y, \mathcal{Z})$, $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbf{P})$ and $(\tilde{Y}, \tilde{\mathcal{Z}})$, $(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}_{t\geq 0}, \tilde{\mathbf{P}})$ with the same initial condition, then $(Y_t)_{t\geq 0} \overset{(\text{law})}{=} (\tilde{Y}_t)_{t\geq 0}$.

We then prove the following well-posedness result

Theorem 3.4.7. *Let $p, q, r \geq 1$, $\alpha \in (1, 2]$ and $\beta \in (-1/2, 0)$ satisfy a good relation for the dynamics. Then,*

(i) the formal SDE admits a unique weak solution;

(ii) if $d = 1$, pathwise uniqueness holds, i.e. the paths of two weak solutions defined on the same probability basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbf{P}, \mathcal{Z})$ coincide a.s. whenever they start from the same initial condition.

Moreover, one can define an associated L^1-stochastic non-linear Young calculus i.e. the above Corollary hold with $\ell = 1$ therein.

The following proposition somehow highlight why we thought that the above description is actually the most accurate description that has been done.

Proposition 3.4.8. *Either the Martingale solution or the weak solution of the formal SDE is: (i) A virtual solution of the formal SDE; (ii) A Dirichlet process; (iii) It holds that for any smooth approximating sequence $(F_m)_{m\geq 1}$ such that*

$$\lim_{m\to+\infty} \|F - F_m\|_{L^r(B^\beta_{p,q})} = 0^3,$$

$$\lim_{m\to\infty} \left\| \int_0^t \mathscr{F}(s, X_s, ds) - \int_0^t F_m(s, X_s) ds \right\|_{L^\ell} = 0, \quad 1 \leq \ell < \alpha,$$

with $\ell = 1$ for the weak solution; (iv) If F is time-space β-Hölder continuous, the previous construction coincides with the "usual" drift:

$$\int_0^t \mathscr{F}(s, X_s, ds) = \int_0^t F(s, X_s) ds, \quad a.s..$$

3.5 (Some) perspectives

- **How robust are the notions of well-posedness for SDE with distributional drift ?** Although we obtain well-posedness for both the Martingale problem and the weak formulation associated with the formal SDE, we did not proved that both formulations are equivalents. What we were able to prove is that weak existence implies the existence of a Martingale solution and that uniqueness for the Martingale problem implies weak uniqueness. The remaining connections are more involved, due to the *a priori* lack of Itô differential calculus. There is a recent work in that direction, from Issoglio and Russo [IR22], but the picture is still not complete.

- **Are thresholds from the weak heuristic rule attainable ?** Another question relies on the heuristic given in Chapter 1. Therein, it is suggested that a for a non degenerate Brownian noise, a drift in $L^\infty(\mathbf{B}^{-1+\varepsilon}_{\infty,\infty})$, $0 < \varepsilon << 1$, is attainable. To do so, the main issue in the above approach relies on multiplication between distributions. As emphasized, the maximal thresholds we obtained precisely allows to define the product in the Duhamel

[3] When p and/or r are/is $+\infty$ the β should be replaced by $\beta - \varepsilon$, for some ε meant to be small.

formula thanks to Bony's paraproduct rule. In this sense, the thresholds are sharp. There are nevertheless other possibilities to bypass such a limit, through rough path or paracontroled calculus. This is precisely what is done by the Authors in [DD15] and then in [CC18] but implies, in counterpart, to add some structure to the drift in the sense that it has to be enhanced into a rough path structure. As an application, such a result may allow to define the characteristics of the PDE studied by Jabin and Wang in [JW18] and allow give a microscopic (at the scale of the particles) interpretation on the associated Propagation of Chaos result. This latter perspective is anything but innocent: we will see at the end of Chapter 5 that the threshold from the weak heuristic rule with a non degenerate Brownian noise can be attained for a certain class of McKean-Vlasov SDEs.

Chapter 4

Non-linear SDEs and associated Mean-Field systems

In the last decade, there has been a real enthusiasm from both the probabilistic and PDE community for non-linear (in McKean sense) and Mean-Field systems. In particular under the impulse of the recent (simultaneous) works of Lasry and Lions, and Caines, Huang and Malhamé, on Mean Field Games [LL06a, LL06b, LL07] and [HMC06] respectively, as well as some recent breakthroughs on fluid mechanics problems [JW18, BJW19]. In both cases, nonlinear systems, in the sense of McKean Vlasov, as well as the associated particles systems interacting in mean field, play a central role.

The objective of this chapter is to present such systems and some associated mathematical tools introduced quite recently by the Mean Field Games community. This part is thus purely bibliographic, no personal results will be presented. It also allows us to make the switch between linear and non-linear (in the McKean sense) SDE in the manuscript, as these systems will appear in all the remaining chapters.

4.1 The systems

Non-linear SDEs (in the sense of McKean) have the particularity to involve the law of the process in their coefficients: the solution of such a system feels, in addition to its own position, its own law. Typically, the system writes

$$dX_t = b\left(t, X_t, \boldsymbol{\mu}_t\right) dt + \sigma(t, X_t, \boldsymbol{\mu}_t) dB_t, \quad X_0 \sim \mu, \quad \mu \in \mathcal{P}(\mathbf{R}^d), \quad t \geq 0, \quad \text{(McKean-Vlasov SDE)}$$

where $\mathcal{P}(\mathbf{R}^d)$ is the space of probability measures on $\mathbf{R}^d$, $(b, \sigma) : \mathbf{R}_+ \times \mathbf{R}^d \times \mathcal{P}(\mathbf{R}^d) \to \mathbf{R}^d \times \mathbf{R}^{d \otimes d}$ and $(\boldsymbol{\mu}_t)_{t \geq 0}$ denotes the family of marginals of the X process. The terminology of non-linear, which could seems obscure from the probabilistic viewpoint, relates on the historical treatment of such processes by McKean [McK66]: the (law) solution of the SDE solves (in a distributional sense) a non-linear Fokker-Planck equation *i.e.* whose operator involves the solution itself. In the following, we call McKean-Vlasov SDE such a system.

Intuitively, these systems can be understood as the asymptotic dynamics of a particle[1][2] evolving

[1] The asymptotic relates to the number of particles in the system.

[2] We emphasize that the terminology of "particle" is purely generic here, and does not refer to any particular physical problem.

within an interacting system, where the interaction is of mean-field type:

$$dX_t^i = b\left(t, X_t^i, \boldsymbol{\mu}_t^N\right) dt + \sigma(t, X_t^i, \boldsymbol{\mu}_t^N) dB_t^i, \quad X_0^i \sim \mu, \quad i = 1, \cdots, N, \qquad \text{(Mean-Field SDE)}$$
$$\boldsymbol{\mu}_t^N = \frac{1}{N} \sum_{i=1}^N \delta_{X_t^i}, \quad t \geq 0, \quad \mu \in \mathcal{P}(\mathbf{R}^d)$$

$\{X_0^i\}_i$ being N independent r.v. independent of the $\{B^i\}_i$ which are N independent Brownian motions of dimension d; the mean-field interaction translates into the dependence of the dynamics w.r.t. the empirical measure of the system $\boldsymbol{\mu}^N$ and the symmetry of the system, in the sense that the law of the above N-uplet is permutation's invariant.

4.2 Well-posedness of the systems and Propagation of Chaos

Well-posedness theory for such systems relies on classical Cauchy-Lipschitz framework. Obviously, as infinite dimensional variable comes into play, the distance used to quantify the Lipschitz property appears to be crucial. A natural candidate is the Wasserstein distance: for all μ, ν in $\mathcal{P}_\ell(\mathbf{R}^d)$, where $\mathcal{P}_\ell(\mathbf{R}^d)$ stands for space of probability measures with finite ℓ^{th} moment,

$$W_\ell(\mu, \nu) = \left(\inf_{X \sim \mu, Y \sim \nu} \mathbf{E}[|X - Y|^\ell] \right)^{1/\ell},$$

especially for $\ell = 1, 2$. This indeed comes from the above form which makes explicitly appears the corresponding L^ℓ distance. On the one hand, this relation allows to implement a fixed point procedure to establish well-posedness results, in a rather similar way than for classical SDEs. On the other hand, it allows to ensure the well-posedness of the associated Mean-Field particle system, viewed as a high dimensional SDE, as the Lipschitz regularity of the coefficients w.r.t. Wasserstein distance for empirical measure gives rise to Lipschitz regularity for Euclidean distance on $\mathbf{R}^{Nd}$. We may refer at this stage to *e.g.* [McK67, Fun84] or [Szn91, CD18a].

Once the system has been shown to be well-posed, the next question concerns the Propagation of Chaos phenomenon relying the McKean-Vlasov SDE and the Mean-Field SDE. When the initial conditions are independent (or asymptotically independent) we can show, as long as the coefficients are Lipschitz (w.r.t. the Euclidean distance for the spatial argument and w.r.t. the Wasserstein distance for the law argument) that the chaos propagates in the system so that, when the number N of particles is large, the particles "become independent": we talk about Propagation of Chaos. Roughly speaking, this can be stated as

$$\forall k, \lim_{N \to +\infty} \mu_0^{1:k,N} \to \mu_0^{\otimes k} \implies \lim_{N \to +\infty} \mu_t^{1:k,N} \to \boldsymbol{\mu}_t^{\otimes k}, t > 0,$$

where for any $t \geq 0$, we denoted by $\mu_t^{1:k,N}$ denotes the joint law of k particles in the Mean-Field SDE. From [Szn91], this is equivalent to a law of large number on the space of probability measures:

$$\boldsymbol{\mu}_0^N \to \mu \implies \boldsymbol{\mu}_t^N \to \boldsymbol{\mu}_t, \quad t > 0.$$

It seems that the terminology of Propagation of Chaos goes back to Kac's work [Kac56] see also the work of McKean [McK67] and one of the probably most known reference for Propagation of

Chaos: the Sznitmann lecture notes [Szn91]. Therein, the Author implements a coupling procedure between the particles in the Mean-Field SDE with an auxiliary system of independent copies of the solution of the McKean-Vlasov SDE in order to estimate the L^2-distance between two (corresponding, through the coupling) elements of both systems and then manage to prove that this distance is of order $N^{-1/2}$, where N stands for the number of particles. The proof again heavily relies on Lipschitz assumptions w.r.t. the 2-Wasserstein distance, the L^2 framework appearing to be quite natural as we somehow look for a rate of convergence for an *ad hoc* LLN.

💬 *We strongly advice the interested reader to have a look on the two recent very complete surveys [CD22, CD21] from Chaintron and Diez for additional references.*

The Propagation of Chaos is crucial from the point of view of applications: on the one hand the McKean-Vlasov SDE provides an approximation of a particle evolving in the Mean-Field SDE in large regime - which allows to work with a d-dimensional system instead of a Nd one (this is, roughly, the idea behind the Mean Field Game); on the other hand the Propagation of Chaos allows to consider the Mean-Field SDE as a (spatial) discretization of the McKean-Vlasov SDE. From this latter numerical perspective, the result in [Szn91] is a strong result, we call it strong Propagation of Chaos as the "approximation error" stays at the level of the path.

4.3 Markov process...

💬 *The Markov property and associated non-linear PDE for the McKean-Vlasov SDE and the Mean-Field SDE goes back to the origin of the study of these processes see e.g. [Bos05] for a survey. We however chose to present things in a slightly different manner here. More in the spirit of the Lions lecture on Mean Field Games at Collège de France, especially the formalism from the books of Carmona and Delarue [CD18a, CD18b]. The interested reader being again strongly advice to consult those pedagogical monographs for further details.*

The Mean-Field SDE, understood as a high dimensional SDE, satisfies the Markov property. A natural question lies in the propagation of this property to the asymptotic system, *i.e.* of McKean-Vlasov type. We can *a priori* guess what could be, or not, possible. Indeed, the Mean-Field SDE verifies the Markov property only as a process taking values in $\mathbf{R}^{Nd}$. The mean-field structure implies that the characterization of the law of the i^{th} particle (among the N) at time $t+s$, for any $s \geq 0$, is done thanks to the position of the said particle at time t and to the statistical distribution of the whole system at this time. Therefore, when "N is large" - *i.e.* in the asymptotic regime $N \to +\infty$ - the law of the process at time $t+s$, for any $s \geq 0$, should be characterized not by the statistical distribution of the particles in space at time t, but by their probabilistic distribution. In other words, the Markov property should be considered on the enlarged space $\mathbf{R}^d \times \mathcal{P}_2(\mathbf{R}^d)$.

To formalize this, a possible strategy consists in introducing the decoupled flow associated with the McKean-Vlasov SDE. Introduce, for Lipschitz coefficients (w.r.t. W_2 for the law variable) and

for an initial condition $\xi \sim \mu$ for the McKean-Vlasov SDE

$$X_s^{t,x,\xi} = x + \int_t^s b(r, X_r^{t,x,\xi}, \boldsymbol{\mu}_r^{t,\mu})dr + \int_t^s \sigma(r, X_r^{t,x,\xi}, \boldsymbol{\mu}_r^{t,\mu})dB_r,$$
$$\boldsymbol{\mu}_t^{t,\xi} = \mu = \mathcal{L}(\xi),\ x \in \mathbf{R}^d, \quad t \in [0,T]. \qquad \text{(decoupled SDE)}$$

This is a classical linear SDE parametrized by the law (the family of marginals $(\boldsymbol{\mu}_s^{t,\mu})_{t\le s\le T}$) of the McKean-Vlasov SDE (the family of marginals $(\boldsymbol{\mu}_s^{t,\mu})_{t\le s\le T}$ depends on ξ only through its law by weak uniqueness, which follows from strong uniqueness).

Consider now the processes $X^{t,x}$ and X^{t,x,δ_x} (respectively solutions of the McKean-Vlasov SDE and of the associated decoupled SDE initialized in x at time t): they solve (in a strong sense) the same SDE and are hence indistinguishable. Thus, since X^{t,x,δ_x} is Markovian, so is $X^{t,x}$. The subtlety, lies in the fact that the inhomogeneous semi-group is, in addition to the time, parameterized by the law of the solution of the McKean-Vlasov SDE. Namely, letting $\xi \sim \mu$ and $(t,x) \in [0,T]\times\mathbf{R}^d$, we know that for all bounded measurable functions $g : \mathbf{R}^d \to \mathbf{R}$, for all s in $[t,T]$, there exists

$$u_{t,\mu}^s : [t,s] \times \mathbf{R}^d \ni (r,x) \mapsto u_{t,\mu}^s(r,x)$$

continuous, such that

$$\mathbf{E}[g(X_s^{t,x,\mu})|\mathcal{F}_r] = u_{t,\mu}^s(r, X_r^{t,x,\mu}).$$

Since, for any r in $[t,T]$, the processes $(X_s^{t,\xi})_{s\in[t,T]}$ and $(X_s^{r,X_r^{t,\xi}})_{s\in[r,T]}$ coincide on $[r,T]$ it is the same for $(\boldsymbol{\mu}_s^{t,\mu})_{s\in[t,T]}$ and $(\boldsymbol{\mu}_s^{r,\boldsymbol{\mu}_r^{t,\mu}})_{s\in[t,T]}$. Thus, for all s in $[t,T]$, for all r in $[t,s]$, $u_{t,\mu}^s(r,\cdot) = u_{r,\boldsymbol{\mu}_r^{t,\mu}}^s(r,\cdot)$, so that

$$\mathbf{E}[g(X_s^{t,x,\mu})|\mathcal{F}_r] = u_{t,\mu}^s(r, X_r^{t,x,\mu}) = u_{r,\boldsymbol{\mu}_r^{t,\mu}}^s(r, X_r^{t,x,\mu}).$$

Finally, if

$$U^s : [0,s] \times \mathbf{R}^d \times \mathcal{P}_2(\mathbf{R}^d) \ni (t,x,\mu) \mapsto U^s(t,x,\mu) := \mathbf{E}[g(X_s^{t,x,\mu})],$$

then for all r in $[t,s]$,

$$\mathbf{E}[g(X_s^{t,\xi})|\mathcal{F}_r] = \mathbf{E}[g(X_s^{t,\xi,\mu})|\mathcal{F}_r] = U^s(r, X_r^{t,\xi}, \boldsymbol{\mu}_r^{t,\mu}).$$

In other words, for the Markov property to be satisfied, the underlying space of $\mathbf{R}^d$ must be enlarged to $\mathbf{R}^d \times \mathcal{P}_2(\mathbf{R}^d)$. This is exactly what we guessed previously.

Before going further and investigating the dynamics of the decoupling field U, let us first focus on the law argument therein, to derive what is the dynamics of such a field w.r.t. this variable. To do so, we come back to the previous discussion where the flow property of the weak solution actually leads to define the semi-group associated with. For any $t \le s$ in $[0,T]$ and for any bounded measurable $\varphi : \mathcal{P}_2(\mathbf{R}^d) \to \mathbf{R}$ we define the family of operators $(\mathscr{P}_{t,s})_{0\le t\le s}$ by:

$$\mathscr{P}_{t,s}\varphi : \mathcal{P}_2(\mathbf{R}^d) \ni \mu \mapsto \varphi(\boldsymbol{\mu}_s^{t,\mu}) \in \mathbf{R}.$$

4.4 ... and related PDE

The dynamics of the (previously defined) semi-group can be derived through differential calculus on space of probability measures. Two approaches, which are linked, will be used below. The

first one is called the flat derivative and is a Gateau derivative on space of signed measure: a map $\varphi : \mathcal{P}(\mathbf{R}^d) \to \mathbf{R}$ is said to have a flat derivative if there exists $\delta_m\varphi \in \mathcal{C}^0(\mathcal{P}_\beta(\mathbf{R}^d)\times\mathbf{R}^d, \mathbf{R})$[3] satisfying $\sup_{\mu\in\mathcal{K}} \delta_m\varphi(\mu)(y) \leq c_\mathcal{K}(1+|y|^\beta)$ such that

$$\lim_{h\to 0} h^{-1}\{\varphi(\mu + h(\nu-\mu)) - \varphi(\mu)\} = \int_{\mathbf{R}^d} \delta_m\varphi(\mu)(y)d(\mu-\nu)(y).$$

The second one is called the Lions (or intrinsic) derivative and is defined on $\mathcal{P}_2(\mathbf{R}^d)$. A map $\varphi : \mathcal{P}_2(\mathbf{R}^d) \to \mathbf{R}$ is said to have a Lions or intrinsic derivative if there exists a map $\partial_\mu\varphi \in \mathcal{C}^0(\mathcal{P}_2(\mathbf{R}^d)\times\mathbf{R}^d, \mathbf{R}^d)$ satisfying $\sup_{\mu\in\mathcal{K}} \partial_\mu\varphi(\mu)(y) \leq c_\mathcal{K}(1+|y|)$ such that

$$\forall \phi \in L^2(\mu), \quad \lim_{h\to 0} h^{-1}\{\varphi(\mu\circ(\mathrm{Id}+h\phi)^{-1}) - \varphi(\mu)\} = \int \partial_\mu\varphi(\mu)(y)\cdot\phi(y)d\mu(y).$$

These two derivatives are, by construction, quite different but may be linked in favorable cases as

$$D\delta_m\varphi = \partial_\mu\varphi,$$

which emphasizes the fact that the Lions derivative has a gradient structure. This can be seen on simple examples: if $\varphi : \mu \mapsto \int \phi d\mu$, with $\phi : \mathbf{R} \to \mathbf{R}$ thus $\delta_m\varphi = \phi$ and $\partial_\mu\varphi = D\phi$.

Having such notions of differentiation, we can then obtain an *ad hoc* Itô's formula for flow of measures induced by an Itô process. This latter makes naturally appear the Lions derivative, which seems to be the more natural object in this framework. We however present it through the prism of the flat derivative, as we feel this is the easiest way to sketch it. If X is the square integrable solution of the McKean-Vlasov SDE and $\boldsymbol{\mu} := (\boldsymbol{\mu}_t)_t$ the associated flow of measures, if $\varphi : \mathcal{P}_2(\mathbf{R}^d) \to \mathbf{R}$ is a "smooth" map, then

$$\varphi(\boldsymbol{\mu}_{t+h}) - \varphi(\boldsymbol{\mu}_t) = \int_{\mathbf{R}^d}\int_0^1 d\lambda\delta_m\varphi(\lambda\boldsymbol{\mu}_{t+h} + (1-\lambda)\boldsymbol{\mu}_t)(y)d(\boldsymbol{\mu}_{t+h} - \boldsymbol{\mu}_t)(y)$$

and we can apply the classical Itô's formula for $\delta_m\varphi(\cdot)(X_{t+h}) - \delta_m\varphi(\cdot)(X_t)$ in the above to deduce

$$d\varphi(\boldsymbol{\mu}_t) = \int_{\mathbf{R}^d} \mathcal{L}[\delta_m\varphi](\boldsymbol{\mu}_t)(y)d\boldsymbol{\mu}_t(y)dt$$

where "$\mathcal{L}$" stands for the generator of the decoupled SDE associated with X (*i.e.* the "classical generator" associated with the linear SDE parametrized by the law of the solution of the McKean-Vlasov SDE). We deduce in turn that the family of generators $(\mathscr{L}_t)_t$ of $(\mathscr{P}_{t,s})_{t\leq s}$ writes, for any smooth functions $\varphi : \mathcal{P}_2(\mathbf{R}^d) \to \mathbf{R}$

$$\begin{aligned}
\mathscr{L}_t\varphi(\mu) &:= \int_{\mathbf{R}^d} \mathcal{L}[\delta_m\varphi](\mu)(y)d\mu(y) \\
&= \int_{\mathbf{R}^d} b(t,y,\mu)D\delta_m\varphi(\mu)(y)d\mu(y) + \int_{\mathbf{R}^d} \frac{1}{2}\mathrm{Tr}\sigma\sigma^*(t,y,\mu)D^2\delta_m\varphi(\mu)(y)d\mu(y) \\
&= \int_{\mathbf{R}^d} b(t,y,\mu)\partial_\mu\varphi(\mu)(y)d\mu(y) + \int_{\mathbf{R}^d} \frac{1}{2}\mathrm{Tr}\sigma\sigma^*(t,y,\mu)D\partial_\mu\varphi(\mu)(y)d\mu(y).
\end{aligned}$$

[3] When no precisions are given, the space $\mathcal{P}_\beta$ is equipped with the distance W_β, $\beta > 0$ with the convention that $\mathcal{P}_0 = \mathcal{P}$ and $W_0 = d_{\mathrm{TV}}$.

We deduce that the related PDE is the Cauchy problem defined on $[0,T] \times \mathcal{P}_2(\mathbf{R}^d)$ for some $T > 0$, by

$$\partial_t \mathfrak{u}(t,\mu) + \mathscr{L}_t \mathfrak{u}(t,\mu) = f(t,\mu),\ (t,\mu) \in [0,T) \times \mathcal{P}_2(\mathbf{R}^d), \quad \mathfrak{u}(T,\mu) = g(\mu) \ \text{ on } \mathcal{P}_2(\mathbf{R}^d)$$

for some data $f : \mathbf{R}_+ \times \mathcal{P}_2(\mathbf{R}^d) \to \mathbf{R}$ and $g : \mathcal{P}_2(\mathbf{R}^d) \to \mathbf{R}$, with the following probabilistic representation (when it holds):

$$\forall (t,\mu) \in [0,T] \times \mathcal{P}_2(\mathbf{R}^d), \quad \mathfrak{u}(t,\mu) = \int_t^T f(s, \boldsymbol{\mu}_s^{t,\mu}) ds + g(\boldsymbol{\mu}_T^{t,\mu}),$$

where $\boldsymbol{\mu}$ stands for the weak solution of the McKean-Vlasov SDE.

We are now in position to deduce the dynamics of the decoupling field U^T. As showed above, the Markov property is achieved provided the decoupled SDE comes into play, the dynamics writes

$$\begin{aligned} &\partial_t U^T(t,x,\mu) + \mathfrak{L} U^T(t,x,\mu) := \partial_t U^T(t,x,\mu) + (\mathcal{L} + \mathscr{L}) U^T(t,x,\mu) = 0, \ \text{ on } [0,T) \times \mathbf{R}^d \times \mathcal{P}_2(\mathbf{R}^d) \\ &U^T(T,x,\mu) = g(x,\mu), \ \text{ on } \mathbf{R}^d \times \mathcal{P}_2(\mathbf{R}^d), \end{aligned}$$

for some data $f : \mathbf{R}_+ \times \mathbf{R}^d \times \mathcal{P}_2(\mathbf{R}^d) \to \mathbf{R}$ and $g : \mathbf{R}^d \times \mathcal{P}_2(\mathbf{R}^d) \to \mathbf{R}$. Non-linear version of the above PDE are known as the Master Equation (of the MFG) and have been thoroughly studied in this perspective, see *e.g.* the work of Cardaliaguet, Delarue, Lasry and Lions [CDLL19].

Chapter 5

Regularization and restoration of chaos for Non-linear SDEs "by noise"

In this chapter, we present the results of the works [3, 9, 12] on non-linear (in a McKean-Vlasov sense) SDEs, related Mean Field systems and associated PDEs. The main purpose of these works is to provide a "robust" theory on the well-posedness of the systems and on the associated Propagation of Chaos for rather irregular coefficients and data, provided the diffusion coefficient is uniformly non degenerate.

These works thus also rely on the first part of the manuscript, as regularization by noise is considered, but in the McKean-Vlasov framework. As such, the main purpose of these works is to understand how (and although it acts only in the state space of the position) the noise in the McKean-Vlasov equation would allow to obtain regularization property w.r.t. the measure argument and restore uniqueness and/or Propagation of Chaos in "non-smooth" cases. It appears that the noise still has some smoothing effects, through the law, and allows to extend the classical well-posedness and Propagation of Chaos theory to a rather large class of equations provided it is non degenerate. To highlight these features, we mainly work under a "favorable" framework *i.e.* the coefficients are somehow Hölder continuous, the metric on the measure argument being precised latter on.

5.1 Back to well-posedness of McKean-Vlasov SDEs

💬 *There are many results on well-posedness for McKean-Vlasov SDEs with drift of the form of an interaction kernel, for rather irregular or singular interaction kernel, as many physical problems rely on this kind of equations. However, to the best of our knowledge, most of these works built unique solution from an ad hoc procedure, taking benefit from the particular structure as an interaction kernel, or as linear maps of the measure, and/or some structural properties on the kernel, e.g. assumed to be divergent free or whatever else. In this sense, we feel that those results do not exactly provide a general well-posedness theory outside the Cauchy-Lipschitz Wasserstein framework. We refer again the interested reader to the surveys [Bos05, JW17, CD22, CD21] for further references. An example of result more in the same spirit of ours is certainly the work of Jourdain [Jou97].*

Yet again, ODEs are McKean-Vlasov SDEs, so that it is unexpected to obtain a general well-posedness theory outside the Cauchy-Lipschitz framework without counterparts. In fact, in full generality, the situation is even worse since, as suggested by the counter-example of Sheutzow

[Sch87], uniqueness may fail for the McKean-Vlasov SDE

$$dY_t = F(Y_t, \boldsymbol{\nu}_t)dt + \sigma dB_t, \quad \boldsymbol{\nu}_t = \mathrm{Law}(Y_t),$$

when $\sigma = 0$ for $F(y, \nu) := \int \tilde{F} d\nu$, $\tilde{F}$ bounded and having only linear growth, while this is not the case for classical SDE. However, when $\sigma > 0$, it seems that the usual Lipschitz property w.r.t. the Wasserstein metric can be weakened to a Lipschitz property w.r.t. the total variation distance defined, for any $\mu, \nu \in \mathcal{P}(\mathbf{R}^d)$, as

$$d_{\mathrm{TV}}(\mu, \nu) = \sup_{A \in \mathcal{B}(\mathbf{R}^d)} |\mu(A) - \nu(A)|.$$

This goes back to the work of Shiga and Tanaka [ST85] for a particular type of measure dependence (linear functions of measure) and then to Jourdain [Jou97] in a rather general framework.

At this stage, we further need to precise what we meant by "weakened". This refers to the fact that, for any probability measures μ, ν with full support on some compact subset $\mathcal{K}$ of $\mathbf{R}^d$ it holds that, *e.g.* $W_\ell(\mu, \nu) \leq \mathrm{diam}(\mathcal{K}) d_{\mathrm{TV}}^{1/\ell}(\mu, \nu)$, and a large class of solutions of non degenerate SDEs lives with high probability in compact subsets of the considered space. Whence, the non degeneracy of the noise allows one to consider stronger topology for the infinite dimensional variable.

While being not exactly formulated as this at that time, this latter fact can be seen as the starting point of the work [9]. Having this kind of phenomenon in mind, we try to carry out a general and meticulous study of these smoothing effects. As in the case of linear SDEs (in the McKean sense), it seems clear to us that such properties would be visible on the associated PDE. The study thus aims at reproducing, in an *ad hoc* framework, an "old" result on smoothing effects of non degenerate linear parabolic PDEs illustrated in a book of Friedman, see [Fri64].

💬 *We emphasize that there is no hope to obtain better smoothing effect than the one we chose to investigate, at least in the framework we've tried to handle. Indeed, the following counter-example communicated by F. Delarue is clear: the McKean-Vlasov SDE $dY_t = F(\mathbf{E}[Y_t])dt + \sigma dB_t$ may have several solutions as soon as F is Hölder continuous as the ODE $dy_t = F(y_t)dt$ may have several solutions as well. To regularize the latter equation, we should add some noise on the measure space. This can be achieved by adding a common noise to the system, as in such a situation the dynamics of the measure flow becomes stochastic. It however seems that even in such a case, the finite dimensional noise is not big enough to provide such kind of smoothing effect. This can be seen from the results in Chapter 2 where, as the number of oscillators tends to infinity, the minimal threshold for the drift tends to one. Thus, bigger noise is needed. We refer e.g. to the work [Mar22] of Marx or to his Ph.D. thesis [Mar19] for related discussion and references on that latter topic, see also [DH22].*

5.2 Smoothing effect of the McKean-Vlasov semi-group: primer

💬 *primer, noun, 1. [uncountable, countable] a type of paint that is put on wood, metal, etc. before it is painted to help the paint to stay on the surface. 2. [countable] (North American English) a book that contains basic instructions. 3. [countable] (old-fashioned) a book for teaching children how to read, or containing basic facts about a school subject.*

Let us try to foster what the smoothing effect of a McKean-Vlasov semi-group could be by considering for (at least) bounded and measurable map $(f, \tilde{f}) : \mathcal{P}(\mathbf{R}^d) \times \mathbf{R}^d \to \mathbf{R} \times \mathbf{R}$

$$\begin{aligned}
&\tilde{X}_t^{\xi} = \xi + B_t, \quad \xi \sim \mu \text{ and is independent of } B \\
&\tilde{\mathscr{P}}_t f(\mu) := f(\boldsymbol{\mu}_t^{\mu}) \text{ with } \boldsymbol{\mu}_t^{\mu} = \mathcal{L}(\tilde{X}_t^{\mu}), \\
&\tilde{P}_t \tilde{f}(x) = \mathbf{E}[f(\tilde{X}_t^{x})], \qquad t \geq 0.
\end{aligned}$$

We first assume that f has a bounded flat derivative. Let $t > 0$, for any $\mu, \mu' \in \mathcal{P}(\mathbf{R}^d)$, we can write

$$\begin{aligned}
\tilde{\mathscr{P}}_t f(\mu) - \tilde{\mathscr{P}}_t f(\mu') &= f(\boldsymbol{\mu}_t^{\mu}) - f(\boldsymbol{\mu}_t^{\mu'}) = \int_{\mathbf{R}^d} \int_0^1 d\lambda \delta_m f(\lambda \boldsymbol{\mu}_t^{\mu} + (1-\lambda)\boldsymbol{\mu}_t^{\mu'})(y) d(\boldsymbol{\mu}_t^{\mu} - \boldsymbol{\mu}_t^{\mu'})(y) \\
&= \int_{\mathbf{R}^d} \int_0^1 d\lambda \tilde{P}_t[\delta_m f(\lambda \boldsymbol{\mu}_t^{\mu} + (1-\lambda)\boldsymbol{\mu}_t^{\mu'})](y) d(\mu - \mu')(y).
\end{aligned}$$

The semi-group thus preserves the flat differentiability property. Moreover, it holds that

$$\delta_m \tilde{\mathscr{P}}_t f(\mu) = \mathbf{R}^d \ni y \mapsto \tilde{P}_t[\delta_m f(\boldsymbol{\mu}_t^{\mu})](y).$$

It is well known that for any $t > 0$, $\tilde{P}_t$ maps $\mathcal{C}_b^{\beta}$ to $\mathcal{C}_b^{2+\beta}$ *w.c.s.* so that, provided the flat derivative of f lies in $\mathcal{C}_b^{\beta}$ uniformly in the measure argument, it holds that the flat derivative of $\tilde{\mathscr{P}}_t f$ lies in $\mathcal{C}_b^{2+\beta}$, uniformly in the measure argument, *w.c.s.* This allows to understand the smoothing effect of the (state space) noise: this latter smooths the flat derivative of the entry. In other words, the noise smooths the law of the linear process (which is absolutely continuous w.r.t. Lebesgue measure), which in turn smooths maps on which it acts. As such, the flat differentiation property allows to linearize maps of measures and to take advantage of the smoothing effect of the law.

The smoothing effect as a weakening of the topology A rather natural perspective to see this consists in reasoning by duality. If we start with a map f in $\mathscr{C}^{1,\beta}(\mathcal{P}_\beta(\mathbf{R}^d))$ (space of functions having a β-Hölder continuous flat derivative), then for any μ, μ' in $\mathcal{P}_\beta(\mathbf{R}^d)$ we obtain that there exists C_f such that

$$|f(\mu) - f(\mu')| \leq C_f \sup_{\varphi \in \mathcal{C}^{\beta}} \int \varphi d(\mu - \mu').$$

By composing it with the McKean-Vlasov semi-group, the computations done in the previous paragraph show that there exists $C'_{f,t}$ for which

$$|\tilde{\mathscr{P}}_t f(\mu) - \tilde{\mathscr{P}}_t f(\mu')| \leq C'_{f,t} \sup_{\varphi \in \mathcal{C}^{2+\beta}} \int \varphi d(\mu - \mu').$$

Therefore we end up to be Lipschitz w.r.t. a coarser topology.

Smoothing effect in term of distance of transport Let us now consider this smoothing effect in terms of transportation distance. Let f in $\mathscr{C}_b^{1,\beta}(\mathcal{P}(\mathbf{R}^d))$ (space of function having a bounded β-Hölder continuous flat derivative) and let $\Pi(\mu, \mu')$ be the set of all transport plans from μ to μ'.

Then, for any μ, μ' in $\mathcal{P}(\mathbf{R}^d)$ and any transference plan $\pi \in \Pi(\mu, \mu')$ it holds

$$\begin{aligned} |f(\mu) - f(\mu')| &= |\int_0^1 d\lambda \int_{\mathbb{R}^d} \delta_m f((1-\lambda)\mu + \lambda\mu')(y) d(\mu - \mu')(y)| \\ &= |\int_0^1 d\lambda \int_{\mathbb{R}^d} \delta_m f((1-\lambda)\mu + \lambda\mu')(x) - \delta_m f((1-\lambda)\mu + \lambda\mu')(y) d\pi(x,y)| \\ &\leq C_f \sup_\mu |\delta_m f(\mu)(\cdot)|_{\mathcal{C}_b^\beta} \int_{(\mathbb{R}^d)^2} \left\{ |x-y|^\beta \wedge 1 \right\} d\pi(x,y). \end{aligned}$$

Taking then the infimum over $\Pi(\mu, \mu')$ gives that

$$|f(\mu) - f(\mu')| \leq \bar{W}_\beta(\mu, \mu') := \inf_{\pi \in \Pi(\mu,\mu')} \int_{(\mathbb{R}^d)^2} \left\{ |x-y|^\beta \wedge 1 \right\} \pi(dx, dy).$$

On the other hand, we have proved that that for any $t > 0$, the flat derivative of $\tilde{\mathscr{P}}_t f(\mu)$ is in $\mathcal{C}_b^{2+\beta}$. In particular, we obtain that

$$|\tilde{\mathscr{P}}_t f(\mu) - \tilde{\mathscr{P}}_t f(\mu')| \leq C_{t,f} \bar{W}_1(\mu, \mu')$$

where $\bar{W}_1$ denotes the bounded-Lipschitz distance. It is easily seen that $\bar{W}_1 \leq c\bar{W}_\beta \leq \tilde{c}\bar{W}_1^\beta$. Whence here, the smoothing effect is better seen as what happens in the linear (in McKean-Vlasov sense) case: we start with a Hölder continuous map and end up with a Lipschitz one (w.r.t. $\bar{W}_1$).

Smoothing effect in term of differentiability Another perspective is linked to a differentiability property. Indeed, starting with a map f in $\mathscr{C}_b^{1,\beta}(\mathcal{P}(\mathbf{R}^d))$, we end up with a map differentiable in the Lions sense, with differentiable Lions derivative. Indeed, denoting for $t > 0$ by $\mathfrak{u}(t, \cdot) = \tilde{\mathscr{P}}_t f$, for any $\phi \in L^2(\mu)$ we have, (denoting as well $\mu^{\lambda,\phi} := (\lambda\mu \circ (\mathrm{Id} + h\phi)^{-1} + (1-\lambda)\mu)$ and recalling that $\delta_m \mathfrak{u}(t, \cdot)$ is in $C_b^{2+\beta}$ *w.c.s.*),

$$\begin{aligned} & \mathfrak{u}(t, \mu \circ (\mathrm{Id} + h\phi)^{-1}) - \mathfrak{u}(t, \mu) \\ = & \int_{\mathbf{R}^d} \int_0^1 \delta_m \mathfrak{u}(t, \mu^{\lambda,\phi})(y) d\lambda d(\mu \circ (\mathrm{Id} + h\phi)^{-1} - \mu)(y) \\ = & \int_{\mathbf{R}^d} \int_0^1 \{\delta_m \mathfrak{u}(t, \mu^{\lambda,\phi})(y + h\phi(y)) - \delta_m \mathfrak{u}(t, \mu^{\lambda,\phi})(y)\} d\lambda d\mu(y) \\ = & h \int_{\mathbf{R}^d} \int_0^1 D\delta_m \mathfrak{u}(t, \mu^{\lambda,\phi})(y) \cdot \phi(y) d\lambda d\mu(y) \\ = & h \int_{\mathbf{R}^d} \int_0^1 \partial_\mu \mathfrak{u}(t, \mu^{\lambda,\phi})(y) \cdot \phi(y) d\lambda d\mu(y), \end{aligned}$$

and thus for any $t > 0$, $\partial_\mu \mathfrak{u}(t, \mu) = \partial_\mu \tilde{\mathscr{P}}_t f(\boldsymbol{\mu}_t^\mu) = D\tilde{P}_t \delta_m f(\boldsymbol{\mu}_t^\mu)$ lies in $C_b^{\beta+1}$, so that it is differentiable (in space). This, in particular, implies that $\mathfrak{u}(t, \cdot)$ is Lipschitz w.r.t. W_2.

Schauder like estimates From the previous discussions, it can be deduced that for any $\varphi \in \mathscr{C}_b^{1,\beta}(\mathcal{P}(\mathbf{R}^d))$, for any $t > 0$, $\tilde{\mathscr{P}}_t f \in \mathscr{C}_b^{1,2+\beta}(\mathcal{P}(\mathbf{R}^d))$ *w.c.s.* so that for $f \in L^\infty([0,T], \mathscr{C}_b^{1,\beta}(\mathcal{P}(\mathbf{R}^d)))$, $g \in \mathscr{C}_b^{1,2+\beta}(\mathcal{P}(\mathbf{R}^d))$,

$$\mathfrak{u}(t, \mu) := \int_t^T \tilde{\mathscr{P}}_{s-t} f(s, \mu) ds + \tilde{\mathscr{P}}_{T-t} g(\mu)$$

solves, in a classical sense, the PDE

$$(\partial_t + \tilde{\mathscr{L}}_t)\mathfrak{u}(t,\mu) = f \text{ on } [0,T)\times\mathcal{P}_2(\mathbf{R}^d),\ \mathfrak{u}(T,\cdot) = g,$$

and satisfies the Schauder-like estimates

$$\|\mathfrak{u}\|_{L^\infty([0,T],\mathscr{C}_b^{1,2+\beta})} \le C_T(\|f\|_{L^\infty([0,T],\mathscr{C}_b^{1,\beta})} + \|g\|_{\mathscr{C}_b^{1,2+\beta}}).$$

5.3 Investigating the McKean-Vlasov system

We now come back to the general form of the McKean-Vlasov SDE, as stated in the previous chapter *i.e.* we are given the following system

$$dX_t = b\left(t, X_t, \boldsymbol{\mu}_t\right)dt + \sigma(t, X_t, \boldsymbol{\mu}_t)dB_t, \quad X_0 = \xi \sim \mu, \quad \mu \in \mathcal{P}_2(\mathbf{R}^d), \quad t \ge 0, \tag{McKean-Vlasov SDE}$$

with B a d-dimensional Brownian motion and ξ a square integrable r.v. independent of B; and the associated Cauchy Problem, for $T > 0$,

$$\partial_t U(t,x,\mu) + \mathfrak{L}U(t,x,\mu) := \partial_t U(t,x,\mu) + (\mathcal{L}+\mathscr{L})U(t,x,\mu) = f(t,x,\mu), \text{ on } [0,T)\times\mathbf{R}^d\times\mathcal{P}_2(\mathbf{R}^d), \tag{Cauchy problem}$$

for some data $(f,g) : [0,T]\times\mathbf{R}^d\times\mathcal{P}_2(\mathbf{R}^d) \to \mathbf{R}$ to be specified latter on. As suggested by the discussion done in the previous part, we mainly assume that the coefficients a, b lie in $\mathfrak{C}_b^{0,\beta,(2,\beta)}(\mathbf{R}_+\times\mathbf{R}^d\times\mathcal{P}(\mathbf{R}^d))$ (space of bounded functions being continuous in time, β-Hölder continuous in space and having 2 β-Hölder bounded flat derivatives). The main objective is to somehow reproduce the previous simple calculations in a general setting.

Tools. To investigate the smoothing effect of the operator $\mathfrak{L}$, we rely, as in the previous chapters, on a parametrix approach. However things are here quite more subtle: because of the non-linear (in McKean sense) structure of the problem, the choice of the proxy becomes quite involved. There is no hope to avoid the dependence of the proxy w.r.t. the law argument while we precisely need to investigate smoothness in that direction. To overcome this problem and indeed work with proxy having frozen law argument in its coefficients, the idea consists in building a suitable sequence of proxys defined through Picard iterations. This works as follows. For any $m \ge 0$ we set

$$dX_t^{(m+1)} = b(t, X_t^{(m+1)}, \boldsymbol{\mu}_t^{(m)})dt + \sigma(t, X_t^{(m+1)}, \boldsymbol{\mu}_t^{(m)})dB_t, \quad X_0 = \xi \sim \mu, \quad \mu \in \mathcal{P}(\mathbf{R}^d), \quad t \ge 0,$$

The above system is a linear SDE (in McKean sense) and thus admits a unique smooth weak solution for a large class of coefficients b, σ, provided σ is uniformly non degenerate and bounded, which is obviously assumed here. By smooth, we mean that the law of the process is absolutely continuous and has moreover Gaussian like density, for any positive time. Indeed, from the well-posedness of the SDE, for any $m \ge 0$ and time $t > 0$, it holds that $d\boldsymbol{\mu}_t^{(m+1),\mu}(x) = p^{(m+1)}(\mu;t,\cdot)dx$ and $p^{(m+1)}(\mu,t,\cdot) = \int q^{(m+1)}(\mu,x,t,\cdot)d\mu(x)$, with $q^{(m+1)}(\mu,x,t,\cdot)$ the density of the $(m+1)^{\text{th}}$ Picard proxy initialized at point x, *i.e.* the density of the associated decoupled flow and we know from Friedman [Fri64] that,

$$q^{(m+1)}(\mu,x,t,\cdot) = \tilde{q}^{(m+1)}(\mu,x,t,\cdot) + \sum_{k\ge1}\tilde{q}^{(m+1)}\otimes H^{(m+1)}(\mu,x,t,\cdot)$$

where $\tilde{q}^{(m+1)}$ is the density of the "usual" proxy (frozen in the spatial argument), $\otimes$ stands for a space-time like convolution operator and $H^{(m+1)}$ is the parametrix kernel $(\mathcal{L}^{(m)} - \tilde{\mathcal{L}}^{(m)})\tilde{q}^{(m+1)}$. This is the main tool to have at hand in order to start our program.

On the well-posedness of the McKean-Vlasov SDE. We begin by studying the weak well-posedness of the McKean-Vlasov SDE through the martingale formulation. Our approach relies on Banach fixed point on the space of martingale solutions

$$\mathcal{M}_T := \{\mathbf{P} \in \mathcal{C}^0([0,T], \mathcal{P}(\mathbf{R}^d)) : \mathbf{P}_0 = \mu\},$$

equip with the distance

$$d_{\mathrm{TV}}^{[0,T]}(\mathbf{P}, \mathbf{Q}) := \sup_{t\in[0,T]} d_{\mathrm{TV}}(\mathbf{P}_t, \mathbf{Q}_t).$$

To make the Picard iterations converge on this space, we expand two successive iterations of the scheme through a classical parametrix expansion and investigate the sensitivity of such an expansion w.r.t. the frozen (at the previous step) law argument. To do so, we use the fact that the coefficients are, uniformly in the time and in the spatial argument, in $\mathscr{C}_b^{1,\beta}$, which in turn implies that they are Lipschitz for the distance defined by duality on the functional space $\mathcal{C}_b^\beta$. This space being included in L^∞, we deduce, from the dual representation of the d_{TV}-distance, that they are Lipschitz for the total variation distance, uniformly in time and thus Lipschitz for $d_{\mathrm{TV}}^{[0,T]}$. Let us point out that, along the calculations, the Hölder continuity assumed on the flat derivative of the diffusion coefficient is crucial to preserve the convergence of the series.

Once the weak well-posedness has been proved, the strong uniqueness may be obtained almost for free. Indeed, having at hand a weak solution allows one to consider the McKean-Vlasov SDE as a linear SDE parametrized by the flow of measure associated with the (unique) weak solution. It is therefore possible, at this stage, to take benefit from the results on strong well-posedness for linear SDE. In the current setting, we obtained strong well-posedness by adding only a Lipschitz assumption (in space) on the diffusion coefficient.

On the related Cauchy problem. Thanks to the convergence of the Picard sequence towards the unique solution of the corresponding martingale problems, we can identify the (pointwise) limit of $q^{(m)}$ as the density q of the decoupled flow associated to the McKean-Vlasov process. We then prove that this density is, on the one hand, smooth enough to apply the operator $\mathfrak{L} := \mathcal{L} + \mathscr{L}$ on it; on the other hand, behaves well in the sense that its derivatives have Gaussian type bounds. The first part of the operator, $\mathcal{L}$, is classical and can be found in [Fri64]. The second one, $\mathscr{L}$, is more involved, and relies on the smoothing effect previously exhibited. We thus manage to prove it through the parametrix expansion along the Picard sequence, uniformly in the Picard argument "m". A compactness argument allows to conclude, along a subsequence of the Picard sequence. This thus justifies, through Markov property (stemming from weak well-posedness) and Itô's formula, that the density q of the decoupled flow associated with the McKean-Vlasov SDE is a fundamental solution of $\mathfrak{L}$: it satisfies,

$$(\partial_t + \mathfrak{L})q(\mu, t, s, x, y) = 0, \quad q(\mu, t, s, x, y) \to \delta_x,\ t \uparrow s.$$

From this result, we eventually derive that

$$U(t,x,\mu) := \mathbf{E}[g(X_T^{t,x,\mu}, \boldsymbol{\mu}_T^{t,\mu}) + \int_t^T f(s, X_s^{t,x,\mu}, \boldsymbol{\mu}_s^{t,\mu})ds]$$

is a "classical" solution to the Cauchy problem with data (f,g), provided $f \in \mathfrak{C}^{0,\beta,(2,\beta)}([0,T] \times \mathbf{R}^d \times \mathcal{P}_2(\mathbf{R}^d))$ and $g \in \mathfrak{C}^{2+\beta,(2,2+\beta)}(\mathbf{R}^d \times \mathcal{P}_2(\mathbf{R}^d))$.

5.4 General results on well-posedness and comments

We state here our results, as they are in the works. For all this part, we consider the general McKean-Vlasov SDE and its associated Mean-Field SDE whose diffusion coefficient is assumed to be uniformly elliptic.

Well-posedness results For the well-posedness, the first result has been given in [3]. To the best of our knowledge, this is one of the first general results on the well-posedness of McKean-Vlasov SDEs outside the Cauchy-Lipschitz-Wasserstein framework that includes a diffusion coefficient depending on the law argument and using the Zvonkin approach through PDE on Wasserstein space. It must be indeed stressed that many other results on well-posedness in this setting relied on Girsanov transform. As such, asking for the diffusion coefficient to be law independent appears to be very convenient. We refer to *e.g.* [Jou97, VM21]. This result may be seen as a preliminary result as it can be included in the results in [9].

(A) The coefficients satisfy the the following structural assumptions

$$(b,\sigma)(t,x,\mu) := (b,\sigma)(t,x,\int \varphi d\mu),$$

where $\varphi_i : \mathbf{R}^d \to \mathbf{R}^d$ is a bounded β-Hölder map (the map φ appearing in the drift can be distinct from the map appearing in the diffusion) and b, σ lie in $\mathcal{C}_b^{0,\beta,2}$.

The terminology of "structural assumption" relies on the fact that the dependence w.r.t. the law argument is assumed to be of scalar type (possibly composed with a smooth function). The theorem proved is the following

Theorem 5.4.1 ([3]). *Under assumption* **(A)**, *the McKean-Vlasov SDE has a unique strong solution.*

The following works contain more general results. The first result we prove therein is under the following set of assumptions and regards the well-posedness.

(A1) The drift coefficient b lies in $L^\infty(\mathbf{R}_+ \times \mathbf{R}^d \times \mathcal{P}(\mathbf{R}^d), \mathbf{R}^d)$ and is Lipschitz-continuous in d_{TV}, uniformly with respect to (t,x);

(A2) The diffusion coefficient $a := \sigma\sigma^*$ lies in $\mathfrak{C}_b^{0,\beta,(1,\beta)}(\mathbf{R}_+ \times \mathbf{R}^d \times \mathcal{P}(\mathbf{R}^d), \mathbf{R}^d)$.

Theorem 5.4.2 ([9]). *Under the above, the martingale problem associated with the McKean-Vlasov SDE is well-posed for any initial distribution $\mu \in \mathcal{P}(\mathbf{R}^d)$. In particular, weak uniqueness in law holds.*

Also, as underlined in the previous section:

Corollary 5.4.3 ([9]). *Assume that the assumptions* **(A1)**-**(A2)** *hold and that for all* (t,μ) *in* $\mathbf{R}_+ \times \mathcal{P}(\mathbf{R}^d)$*, the map* $x \mapsto \sigma(t,x,\mu)$ *is Lipschitz continuous uniformly with respect to* t *and* μ*. Then, strong uniqueness holds for the McKean-Vlasov SDE.*

The regularity of the coefficients w.r.t. the measure argument are stated in a quite different way: for the drift b, a rather general Lipschitz continuity w.r.t. the d_{TV} distance is required, while for the diffusion coefficient we rather choose to suppose that it has a Hölder continuous flat derivative. This last assumption implies that a is Lipchitz w.r.t. the distance defined by duality with the space of C_b^β functions. In fact, in the presence of absolutely continuous measures (w.r.t. the Lebesgue measure), it seems to be an *almost* equivalent regularity assumption (in the sense that all the examples we have in mind for Lipschitz continuous functions w.r.t. the distance defined by duality with the space of C_b^β functions have a.e. a Hölder continuous flat derivative). However, assuming the existence of a flat derivative appears to be crucial when dealing with the parametrix expansion used to obtain the result, because it gives more flexibility: to smooth singularities induced by the differentiation of the heat kernel and to investigate the regularity of the law of the McKean-Vlasov process w.r.t. the measure argument. This is the reason why, below, we further assumed such kind of regularity for the drift as well.

We emphasize that the well-posedness result has then been extended by Frikha, Konakov and Menozzi for stable driven McKean-Vlasov SDE in [FKM21]. There also are several recent works on the well-posedness of the McKean-Vlasov SDE for coefficients being Lipschitz continuous w.r.t. to stronger metric than the Wasserstein one, *e.g.* [HW19, HY21, HW22] where they use fractional Wasserstein or weighted total variation distance. In the case of a convolution type (or at least linear in the measure argument) interaction, we can also refer to [Lac18, MV20, RZ21] where singular drift are considered (but without distribution dependent diffusion coefficient). Also, in [HŠS21], the Authors investigate well-posedness for coefficients with polynomial growth in the measure argument, introducing *ad-hoc* Lyapunov functions.

Density estimates

(B1) The diffusion coefficient $a := \sigma\sigma^*$ and drift coefficient b lie in $\mathcal{C}_b^0(\mathbf{R}_+ \times \mathbf{R}^d \times \mathcal{P}(\mathbf{R}^d))$. Moreover, they belong for any (t,μ) in $\mathcal{C}_b^\beta$, uniformly and for any (t,x) it belongs to $\mathscr{C}_b^{2,\beta}$, uniformly.

💬 *At the end, we only require the flat derivatives of the coefficient to be Hölder continuous w.r.t. the additional variable steaming from the flat differential, uniformly in the others. We anyhow chose to state the above more readable version.*

The result we obtain is the following

Theorem 5.4.4 ([9]). *Assume that* **(A1)**, **(A2)** *and* **(B1)** *hold. Let* $T > 0$ *and* $(s,z) \in (0,T] \times \mathbf{R}^d$*. Then, the mapping* $[0,s) \times \mathbf{R}^d \times \mathcal{P}_2(\mathbf{R}^d) \ni (t,x,\mu) \mapsto q(\mu,t,s,x,z)$ *is in* $\mathcal{C}^{1,2,2}([0,s) \times \mathbf{R}^d \times \mathcal{P}_2(\mathbf{R}^d))$ *and is the fundamental solution associated with* $\mathfrak{L}$*. Moreover, for any* $(\mu,t,x,v) \in \mathcal{P}_2(\mathbf{R}^d) \times [0,s) \times (\mathbf{R}^d)^2$*,*

$$
\begin{aligned}
|\partial_v^n[\partial_\mu q(\mu,t,s,x,z)](v)| &\leq \frac{C}{(t-s)^{\frac{1+n-\beta}{2}}} g(c(s-t), z-x),\ n \in \{0,1\},\\
|\partial_t q(\mu,t,s,x,z)| &\leq \frac{C}{s-t} g(c(s-t), z-x).
\end{aligned}
$$

From this result and the key relation

$$p(\mu,s,t,z)=\int_{\mathbf{R}^d} q(\mu,s,t,x,z)\,d\mu(x), \tag{5.4.1}$$

we deduce the following corollary.

Corollary 5.4.5 ([9]). *Assume that* **(B1)** *hold. Let $T>0$ and $(s,z)\in(0,T]\times\mathbf{R}^d$. Then, the mapping $[0,s)\times\mathcal{P}_2(\mathbf{R}^d)\ni(t,\mu)\mapsto p(\mu,t,s,z)$ is in $\mathcal{C}^{1,2}([0,s)\times\mathcal{P}_2(\mathbf{R}^d))$ and is the fundamental solution associated with $\mathscr{L}$. Moreover, for any $(\mu,t,v)\in\mathcal{P}_2(\mathbf{R}^d)\times[0,s)\times\mathbf{R}^d$,*

$$\begin{aligned}
|\partial_v^n[\partial_\mu p(\mu,t,s,z)](v)| &\le C\Big\{\frac{1}{(s-t)^{\frac{1+n}{2}}}g(c(s-t),z-v)\\
&\quad+\frac{1}{(s-t)^{\frac{1+n-\eta}{2}}}\int_{\mathbf{R}^d}g(c(s-t),z-x)d\mu(x)\Big\},\ n\in\{0,1\},\\
|\partial_t p(\mu,t,s,z)| &\le \frac{C}{s-t}\int_{\mathbf{R}^d}g(c(s-t),z-x)d\mu(x).
\end{aligned}$$

Remark. *We importantly emphasize that, as a by product of our proof, Gaussian estimate on the Holder modulus of the density of the decoupled flow as well as on its derivatives w.r.t. all variables are available.*

On the Cauchy problem on $\mathbf{R}^d\times\mathcal{P}_2(\mathbf{R}^d)$. The assumptions on the data f,g slightly differ from what we stated in the "Primer" as we wanted to handle unbounded entries.

(C1) The two maps $[0,T]\times\mathbb{R}^d\times\mathcal{P}_2(\mathbf{R}^d)\ni(t,x,\mu)\mapsto f(t,x,\mu)$ and $\mathbb{R}^d\times\mathcal{P}_2(\mathbf{R}^d)\ni(x,\mu)\mapsto g(x,\mu)$ are continuous and the two maps $\mathcal{P}_2(\mathbf{R}^d)\ni\mu\mapsto f(t,x,\mu)$, $g(x,\mu)$ have a continuous linear functional derivative for any fixed $(t,x)\in[0,T]\times\mathbb{R}^d$. Moreover, the maps $[0,T]\times(\mathbb{R}^d)^2\times\mathcal{P}_2(\mathbf{R}^d)\ni(t,x,y,\mu)\mapsto\delta_m f(t,x,\mu)(y)$, $(\mathbb{R}^d)^2\times\mathcal{P}_2(\mathbf{R}^d)\ni(x,y,\mu)\mapsto\delta_m g(x,\mu)(y)$ are continuous.

(C1) The maps f, g, $\delta_m f$ and $\delta_m g$ satisfy the following regularity and growth assumptions: there exist $C:=C(T)\ge0$ and $r\ge1$ such that for any $(t,x,y,\mu)\in[0,T]\times(\mathbf{R}^d)^2\times\mathcal{P}_2(\mathbf{R}^d)$ and any bounded set $D\subset\mathbb{R}^d$,

$$\sup_{x\neq x',x,x'\in D}\frac{|f(t,x,\mu)-f(t,x',\mu)|}{|x-x'|^\beta}\le C(1+M_2^r(\mu)), \tag{5.4.2}$$

$$\sup_{y\neq y',y,y'\in D}\frac{|\delta_m f(t,x,\mu)(y)-\delta_m f(t,x,\mu)(y')|}{|y-y'|^\beta}\le C\exp\Big(\alpha\frac{|x|^2}{T}\Big)(1+M_2^r(\mu)), \tag{5.4.3}$$

and

$$|f(t,x,\mu)|+|g(x,\mu)|\le C\exp\Big(\alpha\frac{|x|^2}{T}\Big)(1+M_2^r(\mu)), \tag{5.4.4}$$

$$|\delta_m f(t,x,\mu)(y)|+|\delta_m g(x,\mu)(y)|\le C\exp\Big(\alpha\frac{|x|^2}{T}\Big)(1+|y|^2+M_2^r(\mu)), \tag{5.4.5}$$

where $M_2(\mu):=(\int_{\mathbb{R}^d}|x|^2\,d\mu(x))^{1/2}$ and α is any non-negative constant satisfying $\alpha<(2c)^{-1}$, the constant c depending on the constant appearing in the estimates on the transition density q.

Theorem 5.4.6 ([9]). *Assume that* **(A1)**, **(A2)**, **(B1)**, **(C1)** *and* **(C2)** *hold. Then, the function* U *defined by*

$$\begin{aligned} U(t,x,\mu) &:= \int_{\mathbb{R}^d} g(z,\boldsymbol{\mu}_T^{t,\mu}) q(\mu,t,T,x,z)dz - \int_t^T \int_{\mathbb{R}^d} f(s,z,\boldsymbol{\mu}_s^{t,\mu}) q(\mu,t,s,x,z)dzds \\ &= \mathbf{E}\left[g(X_T^{t,x,\mu},\boldsymbol{\mu}_T^{t,\mu}) - \int_t^T f(s,X_s^{t,x,\mu},\boldsymbol{\mu}_s^{t,\mu})ds \right], \end{aligned}$$

is a solution of the Cauchy problem in the strip $[0,T]\times\mathbf{R}^d\times\mathcal{P}_2(\mathbf{R}^d)$ *and*

$$|U(t,x,\mu)| \le C\exp\Big(\frac{k|x|^2}{T}\Big)(1+M_2^q(\mu)), \qquad \text{for } (t,x,\mu)\in[0,T]\times\mathbb{R}^d\times\mathcal{P}_2(\mathbf{R}^d), \tag{5.4.6}$$

where $C := C(T,)$ *and* $k := k(\lambda,\alpha)$ *are positive constants.*

Moreover, U *is unique among all of the classical solutions to the Cauchy problem on* $[0,T]\times\mathbf{R}^d\times\mathcal{P}_2(\mathbf{R}^d)$ *with source term* f *and terminal condition* g*.*

For smooth coefficients and data, the above Cauchy problem was already proved to be well-posed in [BLP09, CCD22] (in the last reference, a non-linear version coming from MFG was investigated). For non-smooth terminal condition, it has been investigated by Crisan and McMurray [CM17], using the smoothing effect of the McKean-Vlasov semi-group through Malliavin Calculus.

5.5 From the PDE on Wasserstein space to Propagation of Chaos: general results and comments

In this section, we focus on the results obtained in [12], where we precisely use the smoothing property of the PDE to provide a Propagation of Chaos theory for the Mean-Field SDE with non-Lipschitz (w.r.t. euclidean and Wasserstein distance) coefficients.

In some sense, this part relies on numerical perspectives (or is inspired by what happens in this setting): for the time discretization of linear (in the McKean sense) SDE, it is well known that smoothing properties of the associated PDE can be used to reinforce, or obtain, the convergence rate of the associated Euler-Maruyama scheme. Either under weaker regularity assumptions on the coefficients, or with better rates. We refer *e.g.* on the one hand to Dareiotis and Gerencsér [DG20] and Lê and Ling [LL21] and on the other hand to the works of Talay and Tubaro [TT90] and Konakov and Menozzi [KM17]. Therein, the two options are investigated. The first one (which comes secondly in the chronological order) consists in using the smoothing properties of the noise, through the Zvonkin transform, to prove that the strong (at the level of the path) convergence still holds for irregular (even singular) drift; the second one consists in using the PDE to estimate the weak convergence rate of the scheme (*i.e.* at the level of the semi-group). The better rates obtained are respectively of the order of the square root of the number of point in the discretization grid for the strong rate and of the order of this number for the weak rate.

In that perspective, the same questions for the spatial discretization scheme arise. Whence here for the Mean-Field SDE and the associated Propagation of Chaos phenomenon. Previous results strongly suggest to investigate both approaches to see wether the smoothing properties allows to "restore" the Propagation of Chaos *i.e.* to prove: that it holds for non-Lipschitz (w.r.t. Wasserstein-1 distance for the law argument) coefficients; that better rates than the usual one of

the order of the square root of the number of particles can be obtained.

💬 *We deliberately chose to introduce the above perspectives through the prism of numerical probability as it seems to be more confortable for us to justify the terminology used and the intuitions. But, obviously, restoration of, or better rate for, Propagation of Chaos may rely to other perspectives.*

Empirical projection We start with useful connection between the discretized and continuous (in space) version of the measure. Let $\varphi : \mathcal{P}_2(\mathbf{R}^d) \to \mathbf{R}$ be a smooth map. For $N \geq 1$, we may define φ^N as the empirical projection of φ on $\mathbf{R}^{dN}$ *i.e.* $\varphi^N : \mathbf{R}^{dN} \ni \boldsymbol{x} = (x_1, \dots, x_N) \mapsto \varphi(\boldsymbol{\mu}_{\boldsymbol{x}}^N)$ where $\boldsymbol{\mu}_{\boldsymbol{x}}^N := N^{-1} \sum_{i=1}^N \delta_{x_i}$. If φ is of class $\mathscr{C}^{1,2}$, then φ^N is in $\mathcal{C}^2$ and for all $(x_1, \dots, x_N) \in \mathbf{R}^{dN}$:

$$\partial_{x_i} \varphi^N(x_1, \dots, x_N) = \frac{1}{N} D\delta_m \varphi^N(\boldsymbol{\mu}_{\boldsymbol{x}}^N)(x_i),$$

and

$$\partial_{x_j} \partial_{x_i} \varphi^N(x_1, \dots, x_N) = \frac{1}{N^2} D\delta_m \left(D\delta_m \varphi(.)(x_i) \right)(\boldsymbol{\mu}_{\boldsymbol{x}}^N)(x_j) + \mathbf{1}_{i=j} \frac{1}{N} D^2 \delta_m(\boldsymbol{\mu}_{\boldsymbol{x}}^N)(x_j).$$

The above calculations show that the difference between $\mathscr{L}\varphi$ and its space discretization $\mathcal{L}^N \varphi^N$ is of order $N^{-1} D\delta_m D\delta_m \varphi$ which also rewrites $N^{-1} \partial_\mu^2 \varphi$.

Propagation of Chaos Thanks to the smoothing properties of the McKean-Vlasov semi-group, we are able to obtain "Propagation of Chaos"[1] results for particles having only Hölder continuous coefficients in space and w.r.t. the measure variable. This, at three levels: at the level of the density, at the level of the semi-group and at the level of the path. The two first levels rely on what we called a weak Propagation of Chaos, from a numerical perspective, whereas they are strong forms of chaos as they rely on stronger topology than the one induces by Wasserstein distance.

To do so, we need to rather assume an additional regularity assumptions on the coefficients. This rely on the fact that we now need to obtain a good estimate on the error term appearing when approximating a map $u : \mathcal{P}_2(\mathbf{R}^d) \to \mathbf{R}$ by its empirical projection. As stated above, this leads us to obtain estimate on the second order Lions derivative on both densities (of the McKean-Vlasov process and its associated decouple flow). To do so, we further assume that.

(D) The coefficients b and σ satisfy the assumptions of the previous section and are now assumed to belong to $\mathfrak{C}_b^{0,\beta,(3,\beta)}(\mathbf{R}_+ \times \mathbf{R}^d \times \mathcal{P}(\mathbf{R}^d))$.

We then compare the densities of the solution of the McKean-Vlasov SDE and of one element of the Mean-Field SDE. The natural idea consists in comparing the density of the McKean-Vlasov process $p(\mu, 0, t, z)$ with the density of the first particle in the Mean-Field SDE $p^{1,N}(\mu, 0, t, z)$ as fundamental solutions of the Kolmogorov Backward equation on the Wasserstein driven by $\mathscr{L}$ and to estimate the error. To this aim, we crucially use the fact that, because of the symmetry of the system, it holds that $\lim_{s \uparrow t} \mathbf{E}[p(\boldsymbol{\mu}_s^N, s, t, z)] = p^{1,N}(\mu, 0, t, z)$. The result is the following.

[1] 💬 *We here used quote mark because all the results we obtained do not rigorously lead to Propagation of Chaos results as it has been shortly defined in the previous chapter. In order to avoid a long discussion, we nevertheless chose to use this terminology as this relies on the approximation of the McKean-Vlasov SDE by the Mean-Field SDE, and vice versa.*

Theorem 5.5.1 ([12]). *Assume that* **(A1)**, **(A2)**, **(B1)** *and* **(D)** *hold. Then, there exist positive constants K_T, c, (not depending on* **(D)***), $T \mapsto K_T$ being non-decreasing, such that for any $(t, \mu, z) \in (0, T] \times \mathcal{P}_2(\mathbf{R}^d) \times \mathbb{R}^d$*

$$|(p^{1,N} - p)(\mu, 0, t, z)| \leq \frac{K^+}{N} \left\{ \frac{1}{t^{\frac{1-\eta}{2}}} \int_{\mathbb{R}^d} g(ct, z - x)|x|\mu(dx) + \frac{1}{t^{1-\frac{\eta}{2}}} \int_{\mathbb{R}^d} g(ct, z - x)\mu(dx) \right\}.$$

While being rather natural from a numerical perspective, this kind of result is, to the best of our knowledge, new in the perspective of the Propagation of Chaos. We do not found pointwise Gaussian estimates on the difference on the two densities in the Literature, but rather estimate in d_{TV}, which follows from the above. Also we obtain, as a by-product of our analysis, the following uniform, in N, gaussian estimate on the density of one element of the Mean-Field SDE.

Theorem 5.5.2 ([12]). *Assume that* **(A1)**, **(A2)**, **(B1)** *and* **(D)** *hold. Then, there exist positive constants K_T, c, (not depending on* **(D)***), $T \mapsto K_T$ being non-decreasing, such that for any $(t, \mu, z) \in (0, T] \times \mathcal{P}_2(\mathbf{R}^d) \times \mathbb{R}^d$*

$$p^{1,N}(\mu, 0, t, z) \leq K_T \int_{\mathbb{R}^d} g(ct, z - x)\mu(dx)$$

The next result we have stand at the level of the semi-group of the McKean-Vlasov and of the Mean-Field system. We compare their actions by testing them along a suitable class of functions, namely any φ lying in $\mathscr{C}^{2,\beta}(\mathcal{P}_2(\mathbf{R}^d))$. As such, for $\beta = 1$, it holds that the above space contains the 1-Lipschitz functions. So that the result below allows to obtain the convergence in Wasserstein-1 distance.

Theorem 5.5.3 ([12]). *Assume that* **(A1)**, **(A2)**, **(B1)** *and* **(D)** *hold. Then, there exists a positive constant K_T, (not depending on* **(D)***), $T \mapsto K_T$ being non-decreasing such that for all φ lying in $\mathscr{C}^{2,\beta}(\mathcal{P}_2(\mathbf{R}^d))$ whose associated norm is bounded by one*

$$\begin{aligned} \mathbf{E}\left[|\varphi(\mu_T^N) - \varphi(\mu_T)|\right] &\leq K_T \left\{ \frac{1}{T^{\frac{1-\beta}{2}}} \mathbb{E}[W_2(\mu_0^N, \mu)^2]^{1/2} + \frac{1}{N^{\frac{1}{2}}} \right\}. \\ |(\mathscr{P}_T^N - \mathscr{P}_T)\varphi(\mu)| := |\mathbf{E}[\varphi(\mu_T^N)] - \varphi(\mu_T)| &\leq \frac{K_T}{T^{1-\frac{\beta}{2}}} \frac{1}{N}, \end{aligned}$$

This result relies on the analysis done by Carmona and Delarue in Chapter 5 of [CD18a]. However, the semi-group approach developed is reminiscent from the works of Mischler and Mouhot [MM13], Mischler, Mouhot and Wennberg [MMW15] with a slightly different formalism and from the work of Cardaliaguet, Delarue, Lasry and Lions [CDLL19].

💬 *We briefly come back to our previous footnote. The second above result quantify the convergence of the empirical measure associated with the Mean-Field SDE to the law of the McKean-Vlasov SDE, as such, it truly provides a (at least) qualitative (without rate) Propagation of Chaos result. The first one is more subtle: for sure, if the class of maps "φ" for which it holds should contain the set of bounded continuous maps, then the convergence in law of the empirical measure follows, and thus the Propagation of Chaos result hold. This is however not the case. As it is stated (i.e. for the class of maps "φ" is shown to hold), it implies the convergence in Wasserstein-1 distance of the law of one particle in Mean-Field SDE to the law of the*

McKean-Vlasov SDE: this is more an approximation result. Also, the rate of convergence obtained is better than the usual one, of order $\sqrt{N}$. Such rates (of order N) have been shown to be optimal in a recent work of Lacker [Lac23].

The last result relies on Zvonkin transform and thus requires the diffusion coefficient to be Lipschitz. While Zvonkin transform was already applied to recover a (strong) convergence rate of time discretization scheme, it seems that it is the first time it is used to obtain strong convergence rate for Propagation of Chaos. We emphasize that we do not exactly recover the result for the McKean-Vlasov SDE with Lipschitz (w.r.t. Wasserstein distance) for the L^2-norm of the sup. This comes from the Zvonkin Transform, which makes appear in the proof an additional term involving the expectation of the supremum in time of the difference between the McKean-Vlasov law and the corresponding i.i.d. approximation. It is quite hard to handle this term in full generality and we handle it through results of Fournier-Guillin [FG15] together with previous result obtained in [11].

Theorem 5.5.4 ([12]). *Assume that* **(A1)**, **(A2)**, **(B1)** *and* **(D)** *hold and that $M_r(\mu) < +\infty$, for some $r > 4$. Assume that for any $t \in [0,T]$, the map $\mathbf{R}^d \times \mathcal{P}_2(\mathbf{R}^d) \ni (x,\mu) \mapsto \sigma(t,x,\mu)$ is Lipschitz continuous, uniformly in time. Then, there exists a positive constant K_T, $T \mapsto K_T$ being non-decreasing, such that*

$$\sup_{0\le t\le T} \mathbf{E}[W_2(\mu_t,\mu_t^N)^2] + \max_{i=1,\dots,N} \sup_{0\le t\le T} \mathbf{E}\Big[|X_t^i - \bar{X}_t^i|^2\Big] \le K_T \varepsilon_N$$

and

$$\max_{i=1,\dots,N} \mathbf{E}\Big[\sup_{0\le t\le T} |X_t^i - \bar{X}_t^i|^2\Big] \le K_T \sqrt{\varepsilon_N}$$

where ε_N is defined by

$$\varepsilon_N := \begin{cases} N^{-1/2} \text{ if } d < 4, \\ N^{-1/2}\log(1+N) \text{ if } d = 4, \\ N^{-2/d} \text{ if } d > 4. \end{cases}$$

.

5.6 (Some) perspectives

We first emphasize that, as our results strongly rely on suitable heat kernel estimates on the density of the decoupled flow associated with the McKean-Vlasov SDE, we may expect to recover such kind of results for another classes of Markovian noise *e.g.* for stable driven McKean-Vlasov SDE. These investigations have been done by Cavallazzi in his PhD thesis and gave rise to the works [Cav22a, Cav22b] where a positive answer is given.

- **From Mean-Field approximation to "true" Propagation of Chaos results.** In comparison with the recent works Lacker [Lac23, LLF23], where the Propagation of Chaos is investigated through a method based on relative entropy and BBGKY Hierarchy, we obtained an optimal rate of convergence for the empirical measure. However, we do not specify a "true" quantitative Propagation of Chaos (at the level of k-uplet). It would be interesting to investigate wether we could obtain such kind of estimates using the Kolmogorov PDE on Wasserstein space.

- **About singular drifts.** An intensive field of research regards the Propagation of Chaos for a singular interaction kernel (the dependence of the drift coefficient is linear in the measure argument and often takes the form of a convolution with the kernel). Most of the results are of weak form and rely on the use of relative entropy. On the d-dimensional Torus [JW18] deals with a $W^{-1,\infty}$ interaction kernel, having divergence in this space. We can also refer to [BJS22] and [Han22] for less singular kernels. On the whole $\mathbf{R}^d$, there are results of Lacker [Lac18] for a kernel in L^∞ but without rate, then Jabir in the same setting [Jab19] with explicit rate and Tomasevic [Tom19] for $L^q(\text{time}(L^p(\text{space}))$ kernel with $d/p + 2/q < 1$, without rate as well. For a strong Propagation of Chaos result, still without rate, we can refer to the recent work of Hao, Röckner and Zhang [HRZ22]. As such, the extension of the previously exposed theory to obtain the well-posedness and Propagation of Chaos in singular cases is another natural question.

- **Back to the regularization by noise: to what extent can the drift of the system be "bad" ?** Focusing on the well-posedness part, one may wonder if the well-posedness result can be extended to a larger class of coefficients. When the diffusion coefficient is constant and the drift is a inhomogeneous convolution kernel, we considered this perspective with Jabir and Menozzi in the works [17, 18] of the personal bibliography. Therein, we study the case of a singular in time and distributional in space kernel: $b = (t, x, \mu) \mapsto \int \tilde{b}(t, y - x)d\mu(y)$ with $\tilde{b} \in L^\infty(B^{-\beta}_{\infty,\infty})$ (in fact, we consider a rather larger class letting the integrability exponents of the Besov and Lebegues spaces to be finite natural numbers). Taking advantage of the smoothing effect of the McKean-Vlasov density, we succeed in bypassing the limit coming from the Bony's rule and obtain the well posedness for kernel up to $L^\infty(B^{-1^+}_{\infty,\infty})$, thus (almost) reaching the weak heuristic rule. Assuming further condition on the divergence of the kernel, on the initial condition or by reducing the time interval on which the equation is considered, we also obtain well-posedness in the critical case $\tilde{b} \in L^\infty(B^{-1}_{\infty,\infty})$, see also [HRZ23] for similar (and simultaneous) work. The extension to larger class of interaction remains open.

Chapter 6

The Skorokhod problem on the Wasserstein space

In this chapter, we present the works [5] and [11] where reflected stochastic systems are investigated. The main particularity in the setting considered lies into the fact that the reflection is done according to a constraint on the law of the system, and not on its path, as it was introduced in the so-called Skorokhod problem. In this sense, the equations under consideration are non-linear, in the sense of McKean-Vlasov.

6.1 Mean reflected SDE: scalar system and constraint.

A (toy) financial illustration Consider a financial entity with a (dynamical) strategy $(\pi_t)_t$ in a basket of assets $S = (S^1, \dots, S^d)$. Suppose that, under the constraint of the regulator or at the request of a client, the portfolio manager is allowed to hold this position $X_t = \xi + \int_0^t \pi_r \cdot dS_r$ at time t only if it is acceptable in view of a given risk measure *e.g.* a VaR_α, $\alpha \in (0,1)$ (constraint on the risk of loss) or a utility function (constraint on a minimal gain). To satisfy this constraint, the manager must add a quantity of money K_t to the portfolio at each time t. The dynamics of his wealth is thus

$$dX_t = \pi_t dS_t + dK_t.$$

Obviously, the agent will want to cover the risk minimally so that $\mathbf{E}[h(X_t)]dK_t = 0$, where $h = \mathbf{1}_{\mathbf{R}_+} - (1-\alpha))$ in the case of the VaR_α or $h = u(\cdot) - p$ for some utility function u in the second case. Under the standard assumptions of the Black & Scholes model, we obtain an equation of the form

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t + dK_t,\ X_0 = \xi \sim \mu \qquad t \geq 0, \qquad \text{(Mean Reflected SDE)}$$

associated to the constraint $\mathbf{E}[h(X_t)] \geq 0$ with K deterministic and non decreasing, $K_0 = 0$, $\int_0^t \mathbf{E}[h(X_s)]dK_s = 0$: this is what we called a Mean Reflected SDE. The path of the solution being reflected according to a constraint on its "mean". There is here a striking parallel with the so-called Skorokhod problem introduced in [Sko62] where the path, rather than the expectation of some functional, of the solution is constrained.

💬 *The terminology of "mean reflected" comes from the work of Briand, Elie and Hu [BEH18] where such systems have been, to the best of our knowledge, introduced for the first time in a backward setting i.e. therein, the Authors considered a backward SDE with mean reflection.*

Another point of view: the Skorokhod problem on Fokker-Planck equation As a consequence of the above formulation, there is another way to understand the Mean Reflected SDE. Once the equation is proved to be well-posed, we can derive, through Itô's formula, the dynamics of the law of the system. When doing so with the mean reflected SDE, we end up with the following reflected Fokker-Planck equation:

$$d\boldsymbol{\mu}_t(x) = \left\{ D_x(\boldsymbol{\mu}_t(x)b(x)) + \frac{1}{2} D_x^2(\boldsymbol{\mu}_t(x)a(x)) \right\} dt + D_x\left(\boldsymbol{\mu}_t(x)\right)\, dK_t,$$
$$\int h\, d\boldsymbol{\mu}_t \geq 0 \quad \int_0^t \Big(\int h\, d\boldsymbol{\mu}_s \Big)\, dK_s = 0, \quad t \geq 0,$$

where $a = \sigma^2$. In other words, solving the above system translates into searching for solution to the Skorokhod problem stated on Partial Differential Equation of Fokker-Planck type.

💬 *We emphasize that the above formulation a posteriori justifies the deterministic condition implicitly assumed on the process K. We however emphasize that, without such a restriction, Briand, Elie and Hu proved in [BEH18] that this could lead to ill-posedness issues.*

Shape of the reflection Let us try to foster the shape of the process K added in the system. Introduce for this purpose an auxiliary process $U = (U_t)_t$ so that $X_t = U_t + K_t$. The constraint thus rewrites: $H_t(K_t) := \mathbf{E}[h(U_t + K_t)] \geq 0$ which therefore suggests that $K_t \geq H_t^{-1}(0)$ but as the process K should start from 0 and being non-decreasing this gives that $K_t \geq \sup_{s \leq t}(H_t^{-1})^+(0)$. Eventually, we may deduce from the flatness condition: $\forall t \geq 0$, $\int_0^t \mathbf{E}[h(X_s)]dK_s = 0$ that

$$\forall t \geq 0, \quad K_t := \sup_{s \leq t} \inf\{x \in \mathbf{R}_+ : \mathbf{E}[h(x + U_t)] \geq 0\}.$$

The above form obtained for the reflection process gives a thorough justification to the fact that the Mean Reflected SDE is a non-linear SDE, in the McKean-Vlasov sense. It also strongly suggests that there should be a Mean-Field counterpart.

Reflected interacting particle system with mean-field constraint Having an explicit formula for the reflection process allows to identify a natural Mean-Field counterpart for the Mean Reflected SDE. It writes

$$\begin{cases} X_t^i = \bar{\xi}^i + \displaystyle\int_0^t b(X_s^i)ds + \int_0^t \sigma(X_s^i)dB_s^i + K_t^N, \quad 1 \leq i \leq N, \\ \dfrac{1}{N}\displaystyle\sum_{i=1}^N h(X_t^i) \geq 0, \quad \dfrac{1}{N}\sum_{i=1}^N \int_0^t h(X_t^i)\, dK_s^N = 0, \quad t \geq 0, \end{cases} \tag{MF-RSDE}$$

with

$$K_t^N := \inf\left\{ x \geq 0 : \frac{1}{N}\sum_{i=1}^N h\left(x + U_s^i\right) \geq 0 \right\},$$

where

$$U_s^i = \bar{\xi}^i + \int_0^s b(X_r^i)dr + \int_0^s \sigma(X_r^i)dB_r^i, \quad 1 \leq i \leq N.$$

The main issue with the above definition relies on the initial conditions of the particle system: if we choose to define $(\bar{\xi}^i)_i$ as independent copies of ξ, nothing guaranteed that the event $\Omega_N := \left\{\sum_{i=1}^N h(\bar{\xi}^i) \geq 0\right\}$ is of full measure. To overcome this issue, we can take advantage to the explicit representation of the reflection process K and choose to translate the initial conditions in order to satisfy the constraint by defining

$$\bar{\xi}^i := \xi^i + \Lambda(\boldsymbol{\mu}^N(\boldsymbol{\xi})),\ i = 1, \dots, N,$$

with $(\xi^i)_i$ independent copies of ξ, $\boldsymbol{\mu}^N(\boldsymbol{\xi}) := N^{-1}\sum_{i=1}^N \delta_{\xi^i}$ and

$$\Lambda(\boldsymbol{\mu}^N(\xi)) = \inf\left\{x \geq 0 : \frac{1}{N}\sum_{i=1}^N h\left(x + \xi^i\right) \geq 0\right\}.$$

This choice allows to preserve the fact that $\boldsymbol{\mu}^N(\bar{\boldsymbol{\xi}}) \to \mu$ and also leads to the fact that the MF-RSDE satisfies, for any N, a Skorokhod problem.

Indeed if h is concave the particle system can be seen as a multidimensional reflected SDE in $\mathbf{R}^N$ with oblique reflection in the direction $(1, \dots, 1)$ on the boundary of the constraint

$$\mathcal{O} := \left\{(x_1, \dots, x_N) \in \mathbf{R}^N : H^N(x_1, \dots, x_n) > 0\right\},$$

where H^N is the empirical projection of the map H: $\mathcal{P}_2(\mathbf{R}^d) \ni \mu \mapsto \int h d\mu \in \mathbf{R}$ *i.e.* $H^N(x_1, \dots, x_N) = \sum_{i=1}^N h(x_i)$. This is whence a "classical" Skorokhod problem which enters the framework in *e.g.* [LS84].

Well-posedness of the systems In [5], we rigorously prove all the above facts. The strategy mainly relies on the explicit form obtained for the reflection process K. As such, we ask the map $h : \mathbf{R} \to \mathbf{R}$ appearing in the constraint to be an increasing and bi-Lipschitz function: this ensures, at the one hand, that the process K obtained as a generalized inverse is itself Lipschitz continuous w.r.t. the Wasserstein-1 distance; on the other hand that it has a non degenerate derivative (at least almost everywhere). Having these assumptions at hand, and assuming moreover Lipschitz continuity of the coefficients as well as square integrability of the initial condition, we obtain the following results.

Theorem 6.1.1 ([5]). *Under the above assumptions, the Mean Reflected SDE has a unique solution (X, K).*

This result is nothing else but the forward version of the result in [BEH18]. The next result regards the approximation by the MF-RSDE.

Theorem 6.1.2 ([5]). *Under the above assumptions, for any $N > 0$, any $T > 0$, there exists a constant C_T depending on b, σ and h such that*

$$\max_{1 \leq j \leq N} \mathbf{E}\left[\sup_{s \leq T}\left|X_s^j - \bar{X}_s^j\right|^2\right] \leq C_T N^{-1/2},$$

where $\bar{X}^j$ stands for the copy of the Mean Reflected SDE with Brownian motion B^j and initial condition ξ^j.

Also, we propose a Euler-Maruyama discretization scheme of the interacting particle system and prove that it converges (pathwise) to the unique solution of the system at the optimal rate.

6.2 A general framework

The previous Mean-Field counterpart of the Skokorkhod problem with "weak" constraint (the terminology "weak" relies on the fact that the constraint is on law of the process) suggests some natural generalizations. Some of which are considered in [11]. Let us settle the framework. For a smooth map $H : \mathcal{P}_2(\mathbf{R}^d) \to \mathbf{R}$ for coefficients $(b, \sigma_0, \sigma_1) : \mathbf{R}_+ \times \mathbf{R}^d \to \mathbf{R}^d \times \mathbf{R}^{d\times d} \times \mathbf{R}^{d\times d}$ consider:

$$X_t^i = \xi^i + \int_0^t b(s, X_s^i)\, ds + \int_0^t \sigma_0(s, X_s^i)\, dB_s^i + \int_0^t \sigma_1(s, X_s^i)\, dW_s + \int_0^t \partial_\mu H\left(\boldsymbol{\mu}_s^N\right)(X_s^i)\, dK_s^N,$$
$$\boldsymbol{\mu}_t^N = \frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}, \quad H\left(\boldsymbol{\mu}_t^N\right) \geq 0, \qquad \int_0^t H\left(\boldsymbol{\mu}_s^N\right)\, dK_s^N = 0, \quad t \geq 0,$$

(reflected SDE with normal constraint in mean field)

where the $\{B^i\}_i$ and W are independent Brownian motions, W being a common noise as, contrary to the idyosyncratic noises $\{B^i\}_i$, it acts on all the particles of the system. The initial conditions of the particles $\{\xi^i\}_i$ are (for a while) i.i.d. square integrable random variable with law μ and are independent of the $\{B^i\}_i$ and W. The process K^N is a continuous, non-decreasing process adapted to the filtration $\mathcal{F}^N$ generated by the $\{\xi^i\}_i$, the $\{B^i\}_i$ and W.

On the event $\Omega_N := \left\{H\left(N^{-1}\sum_{i=1}^N \delta_{\xi^i}\right) \geq 0\right\}$, this system indeed reads as a classical reflected SDE in $(\mathbf{R}^d)^N$, with normal reflection in the constraint

$$\mathcal{O}_N = \left\{(x_1, \dots, x_N) \in (\mathbf{R}^n)^N,\ H^N(x_1, \dots, x_n) = H\left(\frac{1}{N}\sum_{i=1}^N \delta_{x_i}\right) > 0\right\}.$$

The reflection is said to be normal as $\partial_{x_i} H^N(x_1, \dots, x_N) = N^{-1}\partial_\mu H\left(\frac{1}{N}\sum_{i=1}^N \delta_{x_i}\right)(x_i)$ therefore, the vector $-(\partial_\mu H\left(\frac{1}{N}\sum_{i=1}^N \delta_{x_i}\right)(x_1), \dots, \partial_\mu H\left(\frac{1}{N}\sum_{i=1}^N \delta_{x_i}\right)(x_N))$ is an outward normal to the set $\mathcal{O}_N$ at the point $(x_1, \dots, x_N) \in \partial\mathcal{O}_N$. As such, well-posedness of the solution is immediate under classical assumptions as it again enters the framework of [LS84].

We may thus guess the asymptotic (McKean-Vlasov) dynamics. The candidate writes:

$$X_t = \xi + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma_0(s, X_s)\, dB_s + \int_0^t \sigma_1(s, X_s) dW_s + \int_0^t \partial_\mu H(\boldsymbol{\mu}_s)(X_s)\, dK_s, \quad (6.2.1)$$
$$H(\boldsymbol{\mu}_t) \geq 0, \quad \int_0^t H(\boldsymbol{\mu}_s)\, dK_s = 0, \quad t \geq 0,$$

(reflected SDE with normal constraint on its conditional law)

where $\boldsymbol{\mu}_t = Law(X_t|W)$ is the conditional law of X at time t and $(K_t)_t$ is a continuous non-decreasing process adapted to the filtration $\mathcal{F}^W$ associated with W. This structure (*i.e.* as a conditional McKean-Vlasov SDE) is reminiscent from the common noise setting, see *e.g.* [CD18b]. The Lions derivative $\partial_\mu H(\mu)(\cdot)$ arises naturally as a gradient of H in the Wasserstein space $\mathcal{P}_2(\mathbf{R}^d)$ (see [CD18a]), so that the "outward normal" to the set $\mathcal{O} := \{\mu,\ H(\mu) > 0\}$ at a point $\mu \in \partial\mathcal{O}$ is, at least formally, $\partial_\mu H(\mu)(\cdot)$. This is a therefore a reflected SDE with normal constraint on its conditional law.

Let us now justify the assumptions we previosuly made on the reflection K. To so so, we again come back to the dynamics of the conditional law through the related Fokker-Planck equation. It

writes

$$\begin{cases} d_t\boldsymbol{\mu}_t(x) = \Big\{\mathrm{div}(\boldsymbol{\mu}_t(x)b(t,x)) + \dfrac{1}{2}\displaystyle\sum_{i,j} D^2_{i,j}(\boldsymbol{\mu}_t(x)a_{ij}(t,x))\Big\}dt \\ \qquad\qquad\qquad\qquad +\mathrm{div}\left(\boldsymbol{\mu}_t(x)\sigma_1(t,x)dW_t\right) + \mathrm{div}\left(\boldsymbol{\mu}_t(x)\partial_\mu H(\boldsymbol{\mu}_t)(x)\right)\, dK_t, \\ H(\boldsymbol{\mu}_t) \geq 0 \quad \displaystyle\int_0^t H(\boldsymbol{\mu}_s)\, dK_s = 0, \quad t \geq 0, \end{cases}$$

where $a = (\sigma_0\sigma_0^* + \sigma_1\sigma_1^*)$. This is a stochastic Fokker-Planck equation with normal reflection, which justifies the fact that K is assumed to be a continuous non-decreasing process adapted to the filtration $\mathcal{F}^W$ associated with W.

6.3 SDE with normal constraint on its conditional law: general results and comments

To tackle the above systems, we assume the following conditions to hold:

(HΩ) The probability space is $(\Omega, \mathbf{P}) = (\Omega^0 \times \Omega^1, \mathbf{P}^0 \otimes \mathbf{P}^1)$, where Ω^0 supports the ξ and B, while Ω^1 supports W with associated filtration $\mathcal{F}^W = \mathcal{F}^1$.

(Hc) The functions $b : \Omega \times \mathbf{R}_+ \times \mathbf{R}^d \to \mathbf{R}^d$ and σ_0, $\sigma_1 : \Omega \times \mathbf{R} + \times \mathbf{R}^d \to \mathbf{R}^{d\times d}$ are measurable with respect to $\mathcal{E} \otimes \mathcal{B}\left(\mathbf{R}^d\right)$ and

(i) For all $T > 0$, there exists L_T such that, $\mathbf{P}$-a.s., for each $t \in [0, T]$,

$$\forall x \in \mathbf{R}^d, \forall y \in \mathbf{R}^d, \qquad |b(t,x)-b(t,y)|+|\sigma_0(t,x)-\sigma_0(t,y)|+|\sigma_1(t,x)-\sigma_1(t,y)| \leq L_T\,|x-y|;$$

(ii) b, σ_0, σ_1 are globally bounded: for all $T > 0$, there exists C_T such that, $\mathbf{P}$-a.s.,

$$\sup_{t\leq T,\ x\in\mathbf{R}^d} \{|b(t,x)| + |\sigma_0(t,x)| + |\sigma_1(t,x)|\} \leq C_T.$$

(Ho) The initial condition $\xi \sim \mu_0$ is independent of B and W, in $L^2(\Omega^0)$ and with $H(\mu_0) \geq 0$;

(HH) The function $H : \mathcal{P}_2(\mathbf{R}^d) \to \mathbf{R}$ is fully $\mathcal{C}^2$ and

(i) there exists $M > 0$ such that: $\forall \mu \in \mathcal{P}_2(\mathbf{R}^d)$,

$$|H(\mu)| + \int_{\mathbf{R}^d} |\partial_\mu H(\mu)|^2(x)\, d\mu(x) + \int_{\mathbf{R}^d} |\partial_y\partial_\mu H(\mu)(y)|\, d\mu(y) + \int_{\mathbf{R}^d\times\mathbf{R}^d} \left|D^2_{\mu\mu}H(\mu)(x,y)\right| d\mu(x)d\mu(y) \leq M^2,$$

(ii) there exist $\beta > 0$ and $\eta > 0$ such that

$$\forall \mu \in \mathcal{P}_2(\mathbf{R}^d) \text{ with } -\eta \leq H(\mu) \leq 0, \qquad \int_{\mathbf{R}^d} |\partial_\mu H(\mu)|^2(x)\, d\mu(x) \geq \beta^2,$$

(iii) there exists $C_1 \geq 0$ such that

$$\forall X, Y \in L^2, \qquad \mathbf{E}\left[|\partial_\mu H([X])(X) - \partial_\mu H([Y])(Y)|^2\right] \leq C_1\,\mathbf{E}\left[|X-Y|^2\right].$$

The above assumptions allow to build a solution by a penalization procedure inspired from [LS84, LMS81]: the SDE with normal constraint on its conditional law can be approached by a sequence of penalized McKean-Vlasov SDE, where the penalization comes into play once the boundary of the constraint set is about to be touched. Yet again, the approach for the well-posedness heavily relies on a suitable bi-Lipschitz property of the map H, which is assumed to hold locally *i.e.* in a neighborhood of the boundary of the constraint set (for the bound from below). The result is the following.

Theorem 6.3.1 ([11]). *Under the above assumptions, the reflected SDE with normal constraint on its conditional law has a unique strong and weak square integrable solution. In addition, we have, for any $T > 0$ and for any $\theta > 0$,*

$$\mathbf{E}\left[\sup_{0\le s\le T}|X_s|^2\right] \le C(T), \qquad \mathbf{E}\Big[\exp\{\theta K_T\}\Big] \le C_\theta(T),$$

for some constants $C(T)$ and $C_\theta(T)$ depending on the data, T and, for $C_\theta(T)$, θ.

The next result rigorously justifies the Mean-Field approximation. As in the previous paragraph, this can be done on the event where the empirical measure associated with the initial conditions satisfies the constraint. Therefore, the reflected SDE with normal constraint in mean field is not exactly a solution to the Skorokhod problem. We may however slightly modify the initial data therein in order to obtain a Mean-Field system that satisfies the Skorokhod problem and which indeed approximates the reflected SDE with normal constraint on its conditional law. To do so, the main idea consists in transporting the initial conditions when they are not far from the (empirical projection of the) constraint set along the gradient flow associated with H and eventually replacing it (by some arbitrary point in the set) when they are too far away. The result we obtain is the following.

Theorem 6.3.2 ([11]). *Let $\bar{X}^i$ denotes the strong solution of the reflected SDE with normal constraint on its conditional law with Brownian motion B^i and initial condition ξ^i, $i = 1, \dots, N$. For $T \ge 0$, there exists a constant $C(T)$ independent of N such that*

$$\sup_{i=1,\dots,N}\sup_{t\in[0,T]}\mathbf{E}\left[|X_t^i - \bar{X}_t^i|^2\mathbf{1}_{\Omega_N}\right] \le C(T)\mathbf{E}^{1/2}\left[\sup_{0\le t\le T} W_2^2(\bar{\mu}_s^N, \bar{\mu}_s)\right]$$

and, for N large enough,

$$\sup_{i=1,\dots,N}\mathbf{E}\left[\sup_{0\le t\le T}|X_t^i - \bar{X}_t^i|^2\mathbf{1}_{\Omega_N}\right] \le C(T)\,\mathbf{E}^{1/4}\left[\sup_{0\le s\le T} W_2^2(\bar{\mu}_s^N, \bar{\mu}_s)\right].$$

Lemma 6.3.3 ([11]). *Assume that the initial data satisfies the following moment condition: there exists an integer $p \ge 8$ such that the Lions derivative of H and the law of the initial condition have finite p-moment. Then, there exists $C_T > 0$ such that*

$$\mathbf{E}\left[\sup_{0\le t\le T} W_2^2(\bar{\mu}_s^N, \bar{\mu}_s)\right] \le C_T\epsilon_N,$$

where ϵ_N has been defined at the end of Chapter 4.

Last, bu not least, it is worth noticing that the parallel with the Skorokhod problem suggests to investigate the associated PDE *i.e.* through the dynamics of the family of semi-groups $(\mathscr{P}_t)_{t\geq 0}$ defined by

$$\forall t \geq 0, \forall g \in \mathcal{C}_b^0(\overline{\mathcal{O}}, \mathbf{R}),\ \mathscr{P}_t g(\mu) = \mathbf{E}[g(\boldsymbol{\mu}_T^{t,\mu})],$$

where $\boldsymbol{\mu}$ denotes the flow of conditional law of the solutions to the reflected SDE with normal constraint on its conditional law. Enlarging naively the definition of viscosity solution on finite dimensional space, we show that the dynamics of $\mathfrak{u} : [0,T] \times \overline{\mathcal{O}} \ni (t,\mu) \mapsto \mathscr{P}_t g(\mu)$ is given (in the viscosity sense) by the the following (backward) Neumann problem in the set $\mathcal{O} = \{\mu,\ H(\mu) > 0\}$:

$$\begin{aligned}
&(\partial_t + \mathscr{L})\mathfrak{u}(t,\mu) + \frac{1}{2}\int_{\mathbf{R}^d\times\mathbf{R}^d} \mathrm{Tr}\left(\partial_\mu^2 \mathfrak{u}(t,\mu)(x,y)\sigma_1(t,x)\sigma_1^*(t,y)\right) d\mu(x)d\mu(y) = 0 \text{ in } [0,T)\times\mathcal{O}\\
&\int_{\mathbf{R}^d} \partial_\mu \mathfrak{u}(t,\mu)(y)\cdot\partial_\mu H(\mu)(y)\mu(dy) = 0 \text{ in } [0,T)\times\partial\mathcal{O}\\
&\mathfrak{u}(T,\mu) = g(\mu) \text{ in } \mathcal{O}.
\end{aligned}$$

The second condition is exactly the Neumann boundary condition associated with the set $\mathcal{O}$, whence on the Wasserstein space. Conversely, any smooth solution to the above Neumann problem can be written as $\mathbf{E}[g(\boldsymbol{\mu}_T^{t,\mu})]$.

6.4 BSDE with normal constraint on its law.

As already mentioned, mean reflected systems were introduced for scalar backward SDE by Briand Elie and Hu in [BEH18]. In the scalar case exposed at the beginning of this chapter, we mentioned that, up to some technical arguments, the tools used to investigate the well-posedness of the mean reflected forward and backward system are somehow similar. Such an analogy leads us to investigate wether the penalization procedure roughly described previously allows to tackle the corresponding general formulation in the backward setting. It appeared that the strategy is robust enough to handle backward systems, up to the additional common noise. The system considered is the following

$$\begin{cases}
Y_t = \xi + \displaystyle\int_t^T f(s,Y_s,Z_s)\,ds - \int_t^T Z_s\,dB_s + \int_t^T \partial_\mu H(\boldsymbol{\nu}_s)(Y_s)dK_s, \quad 0\leq t\leq T,\\
H(\boldsymbol{\nu}_t) \geq 0, \quad 0\leq t\leq T, \qquad \displaystyle\int_0^T H(\nu_s)dK_s = 0,
\end{cases}$$

(BSDE with normal constraint on its law)

where $\boldsymbol{\nu}_t = \mathrm{Law}(Y_t)$, $\xi \in L^2(\mathcal{F}_T)$, where the processes Y, Z are respectively of d and $d\times d$ dimension and where $f : \Omega\times\mathbf{R}_+\times\mathbf{R}^d\times\mathbf{R}^{d\times d} \to \mathbf{R}^d$. Here again, H is a map from $\mathcal{P}_2(\mathbf{R}^d)$ to $\mathbf{R}$ and $\partial_\mu H$ denotes the Lions' derivative so that the "outward normal" to the set $\mathcal{O} := \{\mu,\ H(\mu) > 0\}$ at a point $\mu \in \partial\mathcal{O}$ is again, at least formally, $\partial_\mu H(\mu)(\cdot)$. For this reason the above is now called a BSDE with normal constraint on its law.

Let us now briefly expose how (and why) the assumptions used to handle the backward system slightly differ from the assumptions done in the forward case. Firstly, we work with a map H assumed to be concave, in the Lions sense. Secondly, no common noise comes into play. The point is indeed that in the forward case, in the common noise setting, the process K is no longer deterministic (recalling the associated Fokker-Planck equation), and we crucially need to control

uniformly the diffusion coefficient(s) to control the random K (whereas, even in the forward setting, we were able to obtain the results without boundedness of the diffusion coefficient without common noise). For those who are familiar with BSDE, the problem is now rather clear: it may be quite restrictive to impose conditions so that the process Z above is bounded. This is the reason why we chose to work without the common noise. We eventually conclude by emphasizing that the lack of boundedness (even in L^p-norm) of the process Z explains why we further assumed that H was concave.

We now summarize the results obtained in this setting. We first prove the well-posedness of the BSDE with normal constraint on its law, still through the penalization procedure. Then we investigated the Mean Field counterpart of such a system, extending to the current framework the results obtained in by Briand and Hibon in [BH21]. Namely, we consider the backward system

$$\begin{cases} Y_t^i = \xi^i + \displaystyle\int_t^T f(s, Y_s^i, Z_s^{i,i})ds - \int_t^T \sum_{j=1}^N Z_s^{i,j} dB_s^j + \int_t^T D_\mu H(\boldsymbol{\nu}_s^N)(Y_s^i) dK_s^N, \\ \forall t \in [0,T]: \quad \boldsymbol{\nu}_t^N = \dfrac{1}{N}\displaystyle\sum_{i=1}^N \delta_{Y_t^i}, \quad H(\boldsymbol{\nu}_t^N) \geq 0, \quad \displaystyle\int_0^T H(\boldsymbol{\nu}_s^N) dK_s^N = 0, \quad 1 \leq i \leq N, \end{cases}$$

where for each i, j, $Z_s^{i,j}$ is a $d \times d$ matrix, $\{B^i\}_{1 \leq i \leq N}$ are N independent d-dimensional Brownian motions and K^N is a continuous non decreasing process. Assuming that $H(\boldsymbol{\nu}_T) > 0$ this system is, conditionally to $\Omega_N = \{H(\boldsymbol{\nu}_T^N) \geq 0\}$, a classical reflected BSDE in $(\mathbf{R}^d)^N$, with normal reflection on the boundary of the constraint

$$\mathcal{O}_N = \left\{ (y_1, \ldots, y_N) \in (\mathbf{R}^d)^N, \; H\left(\frac{1}{N}\sum_{i=1}^N \delta_{y_i}\right) > 0 \right\}.$$

We then proved the convergence toward the BSDE with normal constraint on its law.

Eventually, we connect the system with a PDE when the process Y is a real-valued process and in a Markovian set up (*i.e.* when $\xi = g(X_T)$ for some $g : \mathbf{R}^d \to \mathbf{R}$ and some diffusion process X evolving according to coefficients $b, \sigma : \mathbf{R}_+ \times \mathbf{R}^d \to \mathbf{R}$ and to the scalar valued Brownian motion B). In this case, the solution of the BSDE with normal constraint on its law can be written in term of the position X and its law thanks to a decoupling field $U : \mathbf{R}_+ \times \mathbf{R}^d \times \mathcal{P}^2(\mathbf{R}^d) \to \mathbf{R}$. When the constraint is on the path, the dynamics of such a decoupling field is given thanks to an obstacle problem, as shown in [EKKP+97]. Here, assuming that the generator f is deterministic we show that such a decoupling field exists and, when f does not depend on the Z argument, we obtain that U solves, still in the (naïve) viscosity sense, the following obstacle problem on the Wasserstein space:

$$\begin{cases} \min\Big\{ \{(\partial_t + \mathfrak{L})U(t,x,\mu) + f(t,x,U(t,x,\mu))\}; \\ \qquad H(U(t,\cdot,\mu)\sharp\mu) \Big\} = 0, \quad \text{on} \quad [0,T) \times \mathbf{R}^d \times \mathcal{P}_2(\mathbf{R}^d), \\ U(T,\cdot,\cdot) = g, \end{cases}$$

where for any probability measures ν and mesurable function φ, $\varphi\sharp\nu$ denotes the push-forward of the measure ν by the map φ.

6.5 (A) perspective

Before giving some perspectives, we emphasize that the case of BSDE with normal constraint on its law with common noise has been studied by Moreau in its PhD thesis.

Long time behavior. In the "forward case", an interesting question lies in the long-time behavior of such a system when it possesses a stationary law. It could provides a natural candidate to solve a minimization problem on the Wasserstein space with constraint. This fact is somehow suggested by the work of Jabir [Jab17] where, starting from such a minimization problem and building a solution by penalization procedure, the Author ends up with a Mean Reflected SDE. We emphasize that links with other works in that direction are of interest, see *e.g.* [Dau23] for optimal control under constraint.

(rough) Bestiary of functions spaces used within the manuscript

Let $\lambda \geq 0$, $\gamma \in [0,2]$ and $\beta \in (0,1]$.

- We say that $\varphi \in \mathcal{C}^\lambda(\mathbf{R}^m, \mathbf{R}^n)$ if $\varphi : \mathbf{R}^m \to \mathbf{R}^n$ is $\lfloor \lambda \rfloor$-times differentiable and $D^{\lfloor \lambda \rfloor}\varphi$ is $\lambda - \lfloor \lambda \rfloor$-Hölder continuous.

- We say that $\varphi \in \mathcal{C}_b^\lambda(\mathbf{R}^m, \mathbf{R}^n)$ if $\varphi : \mathbf{R}^m \to \mathbf{R}^n$ is $\lfloor \lambda \rfloor$-times differentiable, with bounded derivatives and $D^{\lfloor \lambda \rfloor}\varphi$ is bounded and $\lambda - \lfloor \lambda \rfloor$-Hölder continuous.

  Definition 6.5.1. *We introduce $\mathscr{S}_b^{\alpha+\beta}([0,T] \times \mathbf{R}^d)$ the set of functions $\psi(t,x)$ defined on $[0,T] \times \mathbf{R}^d$ such that:*

  (i) The function ψ is continuous on $[0,T] \times \mathbf{R}^d$.

  (ii) For any $t \in [0,T]$ the function $\psi(t,\cdot) \in C_b^{\alpha+\beta}(\mathbf{R}^d)$ and the norm $|\psi(t,\cdot)|_{C_b^{\alpha+\beta}}$ is bounded w.r.t $t \in [0,T]$, i.e., $\psi \in L^\infty([0,T], C_b^{\alpha+\beta}(\mathbf{R}^d))$.

  (iii) There exists a function $\varphi_\psi : [0,T] \times \mathbf{R}^d \to \mathbf{R}$ s.t. for any smooth and compactly supported function $\eta \in C_0^\infty([0,T] \times \mathbf{R}^d)$, the product $(\varphi_\psi \eta)(t,x)$ is bounded and $\beta + \alpha - 1$-Hölder continuous in space uniformly in $t \in [0,T]$ and for any $x \in \mathbf{R}^d$, $0 \leq t < s \leq T$, it holds that:

  $$\psi(s,x) - \psi(t,x) = \int_t^s \varphi_\psi(v,x)dv.$$

  For $\psi \in \mathscr{C}_b^{\alpha+\beta}([0,T] \times \mathbf{R}^d)$, we write $\partial_t \psi = \varphi_\psi$ which is actually the generalized derivative w.r.t. the time variable of the function ψ.

- When being not specified, the space $\mathcal{P}_\gamma(\mathbf{R}^d)$ is endowed with the $\gamma-$Wasserstein distance and for $\gamma = 0$, $\mathcal{P}_0(\mathbf{R}^d)$ denotes the space of probability measure endowed with the weak topology.

Definition 6.5.2. *We say that $\varphi : \mathcal{P}_\gamma(\mathbf{R}^d) \to \mathbf{R}$ has a flat derivative if there exists $\delta_m \varphi \in \mathcal{C}^0(\mathcal{P}_\gamma(\mathbf{R}^d) \times \mathbf{R}^d, \mathbf{R})$ satisfying $\sup_{\mu \in \mathcal{K}} \delta_m \varphi(\mu)(y) \leq c_\mathcal{K}(1 + |y|^\gamma)$, where $\mathcal{K}$ is any compact set of $\mathcal{P}_\gamma(\mathbf{R}^d)$ such that*

$$\lim_{h \to 0} h^{-1}\{\varphi(\mu + h(\nu - \mu)) - \varphi(\mu)\} = \int_{\mathbf{R}^d} \delta_m \varphi(\mu)(y) d(\mu - \nu)(y).$$

Definition 6.5.3. *We say that $\varphi : \mathcal{P}_\gamma(\mathbf{R}^d) \to \mathbf{R}$ has two flat derivatives if for all y the map $\delta_m \varphi(\cdot)(y)$ admits a flat derivative $\delta_m^2 \varphi(\cdot)(y,\cdot)$ such that $\delta_m^2 \varphi$ is in $\mathcal{C}^0(\mathcal{P}_\gamma(\mathbf{R}^d) \times \mathbf{R}^d \times \mathbf{R}^d)$ and satisfies that for any compact $\mathcal{K} \subset \mathcal{P}_\gamma(\mathbf{R}^d)$, $\sup_{\mu \in \mathcal{K}} \delta_m^2 \varphi(\mu)(y,y') \leq c_\mathcal{K}(1 + |y|^\gamma + |y'|^\gamma)$.*

Definition 6.5.4. *And so on for n flat derivatives, $n \geq 3$...*

- We say that $\varphi \in \mathscr{C}^{n,\beta}$ if $\varphi : \mathcal{P}_\beta(\mathbf{R}^d) \to \mathbf{R}$ is continuous, admits n flat derivatives which are all β-Hölder continuous.

- We say that $\varphi \in \mathscr{C}_b^{n,\beta}$ if $\varphi : \mathcal{P}_\beta(\mathbf{R}^d) \to \mathbf{R}$ is bounded continuous, admits n flat derivatives which are all bounded and β-Hölder continuous w.r.t. all spatial arguments.

Definition 6.5.5. *Let ℓ, m be two integers. We denote by $\mathfrak{C}_b^{\ell,\lambda,(m,\beta)}(\mathbf{R}_+ \times \mathbf{R}^d \times \mathcal{P}_\gamma(\mathbf{R}^d))$ the set of maps $\varphi : \mathbf{R}_+ \times \mathbf{R}^d \times \mathcal{P}_\gamma(\mathbf{R}^d)$ such that, for any $(x,\mu) \in \mathbf{R}_+ \times \mathbf{R}^d$ the map $\varphi(\cdot, x, \mu)$ is in $\mathcal{C}_b^\ell(\mathbf{R}_+, \mathbf{R})$ and its derivatives ly in $\mathcal{C}_b^0(\mathbf{R}_+ \times \mathbf{R}^d \times \mathcal{P}_\gamma(\mathbf{R}^d))$; for any $(t,\mu) \in \mathbf{R}_+ \times \mathcal{P}_\gamma(\mathbf{R}^d)$, the map $\varphi(t, \cdot, \mu)$ lies in $\mathcal{C}_b^\lambda(\mathbf{R}^d, \mathbf{R})$, each derivative being (if $\lambda \geq 1$) in $C_b^0(\mathbf{R}_+ \times \mathbf{R}^d \times \mathcal{P}_\gamma(\mathbf{R}^d))$; for any $(t,x) \in \mathbf{R}_+ \times \mathbf{R}^d$, the map $\varphi(t,x,\cdot)$ lies in $\mathscr{C}^{m,\beta}$, each of its n^{th} flat derivative being in addition in $\mathcal{C}_b^0(\mathbf{R}_+ \times \mathbf{R}^d \times \mathcal{P}_\gamma(\mathbf{R}^d) \times \mathbf{R}^{n \times d})$.*

Definition 6.5.6. *For integers ℓ, m we say that the continuous map $\varphi : \mathbf{R}_+ \times \mathbf{R}^d \times \mathcal{P}_2(\mathbf{R}^d)$ is in $\mathcal{C}^{\ell,2,2}(\mathbf{R}_+ \times \mathbf{R}^d \times \mathcal{P}_2(\mathbf{R}^d))$ if the following conditions hold:*

(i) For any $\mu \in \mathcal{P}_2(\mathbf{R}^d)$, the map $\mathbf{R}_+ \times \mathbf{R}^d \ni (t,x) \mapsto \varphi(t,x,\mu)$ is in $\mathcal{C}^{\ell,2}(\mathbf{R}_+ \times \mathbf{R}^d)$ and the functions $\mathbf{R}_+ \times \mathbf{R}^d \times \mathcal{P}_2(\mathbf{R}^d) \ni (t,x,\mu) \mapsto \partial_t^\ell \varphi(t,x,\mu)$, $\partial_x \varphi(t,x,\mu)$, $\partial_x^2 \varphi(t,x,\mu)$ are continuous.

(ii) For any $(t,x) \in \mathbf{R}_+ \times \mathbf{R}^d$, the map $\mathcal{P}_2(\mathbf{R}^d) \ni \mu \mapsto \varphi(t,x,\mu)$ is continuously L-differentiable and for any $\mu \in \mathcal{P}_2(\mathbf{R}^d)$, we can find a version of the mapping $\mathbf{R}^d \ni v \mapsto \partial_\mu \varphi(t,x,\mu)(v)$ such that the mapping $\mathbf{R}_+ \times \mathbf{R}^d \times \mathcal{P}_2(\mathbf{R}^d) \times \mathbf{R}^d \ni (t,x,\mu,v) \mapsto \partial_\mu \varphi(t,x,\mu)(v)$ is locally bounded and is continuous at any (t,x,μ,v) such that $v \in \text{supp}(\mu)$.

(iii) For the version of $\partial_\mu \varphi$ mentioned above and for any (t,x,μ) in $\mathbf{R}_+ \times \mathbf{R}^d \times \mathcal{P}_2(\mathbf{R}^d)$, the mapping $\mathbf{R}^d \ni v \mapsto \partial_\mu \varphi(t,x,\mu)(v)$ is continuously differentiable and its derivative $\partial_v[\partial_\mu \varphi(t,x,\mu)](v) \in \mathbf{R}^{d \times d}$ is jointly continuous in (t,x,μ,v) at any point (t,x,μ,v) such that $v \in \text{supp}(\mu)$.

The continuous function $\varphi : \mathbf{R}_+ \times \mathbf{R}^d \times \mathcal{P}_2(\mathbf{R}^d)$ is in $\mathcal{C}_f^{\ell,2,2}(\mathbf{R}_+ \times \mathbf{R}^d \times \mathcal{P}_2(\mathbf{R}^d))$ if $\varphi \in \mathcal{C}^{\ell,2,2}(\mathbf{R}_+ \times \mathbb{R}^d \times \mathcal{P}_2(\mathbf{R}^d))$ in the above sense and the following additional condition holds:

(iv) For each $v \in \mathbf{R}^d$, the version $\mathcal{P}_2(\mathbf{R}^d) \ni \mu \mapsto \partial_\mu \varphi(t,x,\mu)(v)$ discussed in (ii) is L-differentiable (component by component) with a derivative given by $(\mu, v, v') \mapsto \partial_\mu^2 \varphi(t,x,\mu)(v)(v') \in \mathbf{R}^{d \times d}$ such that for any $\mu \in \mathcal{P}_2(\mathbf{R}^d)$ and $X \in \mathbb{L}_2$ with $[X] = \mu$, the $\mathbb{R}^{d \times d}$-valued random variable $\partial_\mu^2 \varphi(t,x,\mu)(v)(X)$ gives the Fréchet derivative of the map $\mathbb{L}_2 \ni X' \mapsto \partial_\mu^2 \varphi(t,x,[X'])(v)$ for every $v \in \mathbf{R}^d$. Denoting $\partial_\mu^2 \varphi(t,x,\mu)(v)(v')$ by $\partial_\mu^2 \varphi(t,x,\mu)(v,v')$, the map $\mathbf{R}_+ \times \mathbf{R}^d \times \mathcal{P}_2(\mathbf{R}^) \times (\mathbf{R}^d)^2 \ni (t,x,\mu,v,v') \mapsto \partial_\mu^2 \varphi(t,x,\mu)(v,v')$ is also assumed to be continuous for the product topology.

The spaces $\mathcal{C}^{1,2}(\mathbf{R}_+ \times \mathcal{P}_2(\mathbf{R}^d))$ and $\mathcal{C}_f^{1,2}(\mathbf{R}_+ \times \mathcal{P}_2(\mathbf{R}^d))$, are defined through the above, where we adequately remove the space variable.

- We say that $\varphi \in \mathcal{C}^{1,1}(\mathbf{R}_+ \times \mathcal{P}_2(\mathbf{R}^d))$ if φ is continuous, $t \mapsto \varphi(t,\mu) \in \mathcal{C}^1(\mathbf{R}_+)$ for any $\mu \in \mathcal{P}_2(\mathbf{R}^d)$, $(t,\mu) \mapsto \partial_t \varphi(t,\mu)$ being continuous and if for any $t \in \mathbf{R}_+$, $\mu \mapsto \varphi(t,\mu)$ is continuously L-differentiable such that we can find a version of $v \mapsto \partial_\mu \varphi(t,\mu)(v)$ satisfying: $(t,\mu,v) \mapsto \partial_\mu \varphi(t,\mu)(v)$ is locally bounded and continuous at any (t,μ,v) satisfying $v \in \text{supp}(\mu)$.

- We will say that $\varphi \in \mathcal{C}_f^{1,2}(\mathbf{R}_+ \times \mathcal{P}_2(\mathbf{R}^d))$ if $\varphi \in \mathcal{C}^{1,2}(\mathbf{R}_+ \times \mathcal{P}_2(\mathbf{R}^d))$ and for the version of $\partial_\mu \varphi$ previously considered, for any $(t, v) \in [0, T] \times \mathbf{R}^d$, the mapping $\mathcal{P}_2(\mathbf{R}^d) \ni \mu \mapsto \partial_\mu \varphi(t, \mu)(v)$ is L-differentiable with a derivative given by $(t, \mu, v, v') \mapsto \partial_\mu \varphi(t, \mu)(v, v') \in \mathbf{R}^{d \times d}$ such that for any $\mu \in \mathcal{P}_2(\mathbf{R}^d)$ and $X \in \mathbb{L}_2$ with $[X] = \mu$, $\partial_\mu \varphi(t, \mu)(v, X)$ gives the Fréchet derivative of the map $\mathbb{L}_2 \ni X' \mapsto \partial_\mu \varphi(t, [X'])(v)$ for every $(t, v) \in \mathbf{R}_+ \times \mathbf{R}^d$. Moreover, the map $\mathbf{R}_+ \times \mathcal{P}_2(\mathbf{R}^d) \times (\mathbf{R}^d)^2 \ni (t, \mu, v, v') \mapsto \partial_\mu^2 \varphi(t, \mu)(v, v')$ is assumed to be continuous for the product topology.

Bibliographie - Bibliography

[ABM20] A. Athreya, O. Butkovsky, and L. Mytnik, Strong existence and uniqueness for stable stochastic differential equations with distributional drift, The Annals of Probability (2020).

[Bas09] R. F. Bass, Regularity results for stable-like operators, Journal of Functional Analysis (2009).

[BC01] R.F. Bass and Z.Q. Chen, Stochastic differential equations for Dirichlet processes, Probability Theory and Related Fields (2001).

[BC03] ———, Brownian motion with singular drift, The Annals of Probability (2003).

[BEH18] Ph. Briand, R. Elie, and Y. Hu, Bsdes with mean reflection, The Annals of Applied Probability (2018).

[BFGM19] L. Beck, F. Flandoli, M. Gubinelli, and M. Maurelli, Stochastic ODEs and stochastic linear PDEs with critical drift: regularity, duality and uniqueness, Electronic Journal of Probability (2019).

[BH21] Ph. Briand and H. Hibon, Particles systems for mean reflected bsdes, Stochastic Processes and their Applications (2021).

[BJS22] D. Bresch, P.-E. Jabin, and J. Soler, A new approach to the mean-field limit of vlasov-fokker-planck equations, 2022.

[BJW19] D. Bresch, P.-E. Jabin, and Z. Wang, On mean field limit and quantitative estimates with a large class of singular kernels: Application to the patlak-keller-segel model, Comptes Rendus Mathematique (2019).

[BLP09] R. Buckdahn, J. Li, and S. Peng, Mean-field backward stochastic differential equations and related partial differential equations, Stochastic Processes and their Applications (2009).

[Bos05] M. Bossy, Some stochastic particle methods for non-linear parabolic pdes, ESAIM Proceedings and Survey (2005).

[Cav22a] Thomas Cavallazzi, Itô's formula for the flow of measures of poisson stochastic integrals and applications, 2022.

[Cav22b] ———, Quantitative weak propagation of chaos for stable-driven mckean-vlasov sdes, 2022.

[CC18] G. Cannizzaro and K. Chouk, Multidimensional SDEs with singular drift and universal construction of the polymer measure with white noise potential, The Annals of Probability (2018).

[CCD22] F. Chassagneux, D. Crisan, and F. Delarue, A probabilistic approach to classical solutions of the master equation for large population equilibria, Memoirs of the AMS, 2022.

[CD18a] R. Carmona and F. Delarue, Probabilistic theory of mean field games with applications i: Mean field fbsdes, control, and games, Springer International Publishing, 2018.

[CD18b] ———, Probabilistic theory of mean field games with applications ii: Mean field games with common noise and master equations, Springer International Publishing, 2018.

[CD21] L.-P. Chaintron and A. Diez, Propagation of chaos: a review of models, methods and applications. ii. applications, 2021.

[CD22] ———, Propagation of chaos: a review of models, methods and applications. i. models and methods, 2022.

[CDLL19] P. Cardaliaguet, F. Delarue, M. Lasry, and P.-L. Lions, The master equation and the convergence problem in mean field games, Princeton University Press, 2019.

[CG16] R. Catellier and M. Gubinelli, Averaging along irregular curves and regularisation of ODEs, Stochastic Processes and their Applications (2016).

[Che05] A. S. Cherny, Singular Stochastic Differential Equations, Springer, 2005.

[CM17] D. Crisan and E. McMurray, Smoothing properties of mckean–vlasov sdes, Probability Theory and Related Fields (2017).

[CV10] L. A. Caffarelli and A. Vasseur, Drift diffusion equations with fractional diffusion and the quasi-geostrophic equation, Annals of Mathematics. Second Series (2010).

[CZZ17] Z.-Q. Chen, X. Zhang, and G. Zhao, Well-posedness of supercritical sde driven by Lévy processes with irregular drifts, https://arxiv.org/abs/1709.04632 (2017).

[CZZ21] Z.Q. Chen, X. Zhang, and G. Zhao, Well-posedness of supercritical SDE driven by Lévy processes with irregular drifts, Transaction of the American Mathematical Society (2021).

[Dau23] S. Daudin, Mean-field limit for stochastic control problems under state constraint, 2023.

[Dav07] A.M. Davie, Uniqueness of solutions of stochastic differential equations, International Mathematics Research Notices (2007).

[DD15] F. Delarue and R. Diel, Rough paths and 1d SDE with a time dependent distributional drift: application to polymers, Probability Theory and Related Fields (2015).

[DF14] F. Delarue and F. Flandoli, The transition point in the zero noise limit for a 1d Peano example, Discrete and Continuous Dynamical Systems (2014).

[DG20] K. Dareiotis and M. Gerencsér, On the regularisation of the noise for the euler-maruyama scheme with irregular drift, Electronic Journal of Probability (2020).

[DH22] F. Delarue and W.R.P. Hammersley, Rearranged stochastic heat equation, 2022.

[DL89] R.J. DiPerna and P.-L. Lions, Ordinary differential equations, transport theory and sobolev spaces, Inventiones Mathematicae (1989).

[DM10] F. Delarue and S. Menozzi, Density estimates for a random noise propagating through a chain of differential equations, Journal of Functional Analysis (2010).

[DPL95] G. Da Prato and A. Lunardi, On the ornstein-uhlenbeck operator in spaces of continuous functions, Journal of Functional Analysis (1995).

[EKKP+97] K. El Karoui, C. Kapoudjian, E. Pardoux, S. Peng, and M.-C. Quenez, Reflected solutions of backward sde's, and related obstacle problems for pde's, The Annals of Probability (1997).

[FF11] F. Fedrizzi and F. Flandoli, Pathwise uniqueness and continuous dependence of SDEs with non-regular drift, Journal of Probability and Stochastic Processes (2011).

[FFPV17] F. Fedrizzi, F. Flandoli, E. Priola, and J. Vovelle, Regularity of stochastic kinetic equations, Electronic Journal of Probability (2017).

[FG15] N. Fournier and A. Guillin, n the rate of convergence in wasserstein distance of the empirical measure, Probability Theory and Related Fields (2015).

[FIR17] F. Flandoli, E. Issoglio, and F. Russo, Multidimensional stochastic differential equations with distributional drift, Transactions of the American Mathematical Society (2017).

[FKM21] N. Frikha, V. Konakov, and S. Menozzi, Well-posedness of some non-linear stable driven sdes, Discrete and Continuous Dynamical Systems (2021).

[Fri64] A. Friedman, Partial differential equations of parabolic type, Prentice-Hall Inc., Englewood Cliffs, N.J., 1964.

[FRW03] F. Flandoli, F. Russo, and J. Wolf, Some SDEs with distributional drift. I. General calculus, Osaka Journal of Mathematics (2003).

[FRW04] ———, Some SDEs with distributional drift. II. Lyons-Zheng structure, itô's formula and semimartingale characterization, Random Operators and Stochastic Equations (2004).

[Fun84] T. Funaki, A certain class of diffusion processes associated with nonlinear parabolic equations, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete (1984).

[Ger22] M. Gerencser, Regularisation by regular noise, Stochastics and Partial Differential Equations: Analysis and Computations (2022).

[GO13] M. Gradinaru and Y. Offret, Existence and asymptotic behaviour of some time-inhomogeneous diffusions, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques (2013).

[Han22] Y. Han, Entropic propagation of chaos for mean field diffusion with l^p interactions via hierarchy, linear growth and fractional noise, 2022.

[HMC06] M. Huang, R.P. Malhamé, and P.E. Caines, Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle, Communications in Information & Systems (2006).

[Hör67] L. Hörmander, Hypoelliptic second order differential operators, Acta Mathematica (1967).

[HRZ22] Z. Hao, M. Röckner, and X. Zhang, Strong convergence of propagation of chaos for mckean-vlasov sdes with singular interactions, 2022.

[HRZ23] ———, Second order fractional mean-field sdes with singular kernels and measure initial data, 2023.

[HŠS21] W. Hammersley, D. Šiška, and L. Szpruch, Mckean-vlasov sdes under measure dependent lyapunov conditions, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques (2021).

[HW19] X. Huang and F.-Y. Wang, Distribution dependent sdes with singular coefficients, Stochas-tic Processes and their Applications (2019).

[HW22] ———, Singular mckean-vlasov (reflecting) sdes with distribution dependent noise, Journal of Mathematical Analysis and Applications (2022).

[HY21] X. Huang and F.-F. Yang, Distribution dependent sdes with hölder continuous drift and α-stable noise, Numerical Algorithms (2021).

[IJS] C. Imbert, T. Jin, and R. Shvydkoy.

[IM21] C. Imbert and C. Mouhot, The Schauder estimate in kinetic theory with application to a toy nonlinear model, Annales Henri Lebesgue (2021).

[IR22] E. Issoglio and E. Russo, Sdes with singular coefficients: The martingale problem view and the stochastic dynamics view, 2022.

[Jab17] J/.-F. Jabir, Diffusion processes with weak constraint through penalization approximation, 2017.

[Jab19] J.-F. Jabir, Rate of propagation of chaos for diffusive stochastic particle systems via girsanov transformation, 2019.

[Jou97] B. Jourdain, Diffusions with a nonlinear irregular drift coefficient and probabilistic interpretation of generalized burgers' equations, ESAIM P.&S. (1997).

[JW17] P.-E. Jabin and Z. Wang, Mean field limit for stochastic particle systems, Springer International Publishing (2017).

[JW18] _____, Quantitative estimates of propagation of chaos for stochastic systems with $W^{-1,\infty}$ kernels, Inventiones mathematicae (2018).

[Kac56] M. Kac, Foundations of kinetic theory, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, vol. 3, University of California Press Berkeley and Los Angeles, California, (1956).

[KM17] V. Konakov and S. Menozzi, Weak error for the euler scheme approximation of diffusions with non-smooth coefficients, Electronic Journal of Probability (2017).

[Kol34] A.N. Kolmogorov, Zufällige Bewegungen (zur Theorie der Brownschen Bewegung), Annals of Mathematics (1934).

[KP10] N.V. Krylov and E. Priola, Elliptic and parabolic second-order PDEs with growing coefficients, Communications in Partial Differential Equations (2010).

[KP22] H. Kremp and N. Perkowski, Multidimensional sde with distributional drift and lévy noise, Bernoulli (2022).

[KR05] N.V. Krylov and M. Röckner, Strong solutions of stochastic equations with singular time dependent drift, Probability Theory and Related Fields (2005).

[Lac18] D. Lacker, On a strong form of propagation of chaos for mckean-vlasov equations, Electronic Communications in Probability (2018).

[Lac23] _____, Hierarchies, entropy, and quantitative propagation of chaos for mean field diffusions, Probability and Mathematical Physics (2023).

[LL06a] P.-L. Lions and J.-M. Lasry, eux à champ moyen. I. Le cas stationnaire, Comptes Rendus Mathématique. Académie des Sciences. Paris (2006).

[LL06b] _____, Jeux à champ moyen. II. Horizon fini et contrôle optimal, Comptes Rendus Mathématique. Académie des Sciences. Paris (2006).

[LL07] _____, Mean field games, Japanese Journal of Mathematics (2007).

[LL21] K. Lê and C. Ling, Taming singular stochastic differential equations: A numerical method, 2021.

[LLF23] D. Lacker and L. Le Flem, Sharp uniform-in-time propagation of chaos, Probability Theory and related Fields (2023).

[LMS81] P.-L. Lions, J.L. Menaldi, and A.-S. Sznitman, onstruction de processus de diffusion réfléchis par pénalisation du domaine, Comptes Rendus Mathématique. Académie des Sciences. Paris (1981).

[Lor05] L. Lorenzi, Schauder estimates for degenerate elliptic and parabolic problems with unbounded coefficients in $\setminus^n$, Differential Integral Equations (2005).

[LS84] P.-L. Lions and A.-S. Sznitman, Stochastic Differential Equations with Reflecting Boundary Conditions, Communication on Pure and Applied Mathematics (1984).

[Lun97] A. Lunardi, Schauder estimates for a class of degenerate elliptic and parabolic operators with unbounded coefficients in $\mathbb{R}^n$, Annali della Scuola Normale Superiore di Pisa - Classe di Scienze (1997).

[Mar19] V. Marx, Diffusive processes on the wasserstein space: Coalescing models, regularization properties and mckean-vlasov equations, Theses, Université Côte d'Azur, 2019.

[Mar20] L. Marino, Schauder estimates for degenerate stable kolmogorov equations, Bulletin des Sciences Mathématiques (2020).

[Mar22] V. Marx, A bismut–elworthy inequality for a wasserstein diffusion on the circle, Stochastics and Partial Differential Equations: Analysis and Computations (2022).

[McK66] H.P. McKean, A class of markov processes associated with nonlinear parabolic equations, Proc Natl Acad Sci U S A (1966).

[McK67] ———, Propagation of chaos for a class of non-linear parabolic equations, Stochastic Differential Equations (1967).

[Men11] S. Menozzi, Parametrix techniques and martingale problems for some degenerate kolmogorov equations, Electronic Communications in Probability (2011).

[Men18] ———, Martingale problems for some degenerate Kolmogorov equations, Stochastic Processes and Their Applications (2018).

[MM13] S. Mischler and C. Mouhot, Kac's program in kinetic theory, Inventiones Mathematicae (2013).

[MM21] L. Marino and S. Menozzi, Weak well-posedness for degenerate sdes driven by lévy processes, 2021.

[MMW15] S. Mischler, C. Mouhot, and B. Wennberg, A new approach to quantitative propagation of chaos for drift, diffusion and jump processes, Probability Theory and Related Fields (2015).

[MP14] R. Mikulevicius and H. Pragarauskas, On the Cauchy problem for integro-differential operators in Hölder classes and the uniqueness of the martingale problem, Potential Analysis (2014).

[MS67] H.P. McKean and I.M. Singer, Curvature and the eigenvalues of the Laplacianh.p., Journal of Differential Geometry (1967).

[MV20] Y. Mishura and A. Veretennikov, Existence and uniqueness theorems for solutions of mckean–vlasov stochastic equations, Theory of Probability and Mathematical Statistics (2020).

[Pri09] E. Priola, Global Schauder estimates for a class of degenerate Kolmogorov equations, Studia Mathematica (2009).

[Pri12a] ———, Pathwise uniqueness for singular SDEs driven by stable processes, Osaka Journal of Mathematics (2012).

[Pri12b] ———, Pathwise uniqueness for singular SDEs driven by stable processes, Osaka Journal of Mathematics (2012).

[Pri18a] ———, Davie's type uniqueness for a class of sdes with jumps, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques (2018).

[Pri18b] ———, Davie's type uniqueness for a class of SDEs with jumps, Annales de l'Institut Henri Poincaré Probabilités et Statistiques (2018).

[Pri21] ———, On Davie's uniqueness for some degenerate SDEs, Theory of Probability and Mathematical Statistics (2021).

[RZ21] M. Röckner and X. Zhang, Well-posedness of distribution dependent sdes with singular drifts, Bernoulli (2021).

[Sch87] M. Scheutzow, niqueness and non-uniqueness of solutions of vlasov-mckean equations, Journal of the Australian Mathematical Society. Series A. Pure Mathematics and Statistics (1987).

[Sil12] L. Silvestre, On the differentiability of the solution to an equation with drift and fractional diffusion, Indiana University Mathematical Journal (2012).

[Sko62] A.V. Skorokhod, Stochastic equations for diffusion processes in a bounded region, Theory of Probability & Its Applications (1962).

[ST85] T. Shiga and H. Tanaka, Central limit theorem for a system of markovian particles with mean field interactions, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete (1985).

[SV79] D.W. Stroock and S.R.S. Varadhan, Multidimensional diffusion processes, Springer-Verlag Berlin Heidelberg New-York, 1979.

[SVZ13] L. Silvestre, V. Vicol, and A. Zlatoš, On the loss of continuity for super-critical drift-diffusion equations, Archive for Rational Mechanics and Analysis (2013).

[Szn91] A.-S. Sznitman, Topics in propagation of chaos, Ecole d'Eté de Probabilités de Saint-Flour XIX — 1989 (Paul-Louis Hennequin, ed.), Springer Berlin Heidelberg, 1991.

[Tom19] M. Tomasevic, Propagation of chaos for stochastic particle systems with singular mean-field interaction of $l^q - l^p$ type, 2019.

[Tri83] H. Triebel, Theory of function spaces, ii, Birkhauser, 1983.

[TT90] D. Talay and L. Tubaro, Expansion of the global error for numerical schemes solving stochastic differential equations, Stochastic Analysis and Applications (1990).

[TTW74] H. Tanaka, M. Tsuchiya, and S. Watanabe, Perturbation of drift-type for lévy processes, Journal of Mathematics of Kyoto University (1974).

[Ver80] A.Y. Veretennikov, Strong solutions and explicit formulas for solutions of stochastic integral equations, Matematicheski\u\ Sbornik. Novaya Seriya (1980).

[VM21] A. Veretennikov and Y. Mishura, Existence and uniqueness theorems for solutions of mckean–vlasov stochastic equations, Theory of Probability and Mathematical Statistics (2021).

[WZ16] F.Y. Wang and X. Zhang, Degenerate SDE with Holder-Dini drift and non-Lipschitz noise coefficient, SIAM Journal on Mathematical Analysis (2016).

[ZH20] X. Zhang and Z. Hao, Schauder estimates for nonlocal kinetic equations and applications, ournal de Mathématiques Pures et Appliquées (2020).

[Zha10] X. Zhang, Well-posedness and large deviation for degenerate SDEs with Sobolev coefficients, Revista Matematica Iberoamericana (2010).

[Zha18] ———, Stochastic Hamiltonian flows with singular coefficients, Science China. Mathematics (2018).

[Zvo74] A. K. Zvonkin, A transformation of the phase space of a diffusion process that will remove the drift, Mat. Sb. (N.S.) (1974).

[ZZ17] X. Zhang and G. Zhao, Heat kernel and ergodicity of sdes with distributional drifts, 2017.

[ZZ18] ———, Dirichlet problem for supercritical non-local operators, arXiv:1809.05712 (2018).

Nantes Université

Titre. ***Bal(l)ade entre EDP et probabilités.***

Résumé. Ce manuscrit vise à retracer les travaux effectués depuis ma soutenance de thèse au LJAD, à l'université de Nice Côte d'Azur, poursuivis durant mon affectation au LAMA à l'Université Savoie Mont Blanc puis au LMJL à Nantes Université. Ces travaux s'inscrivent dans le domaine de *l'analyse stochastique* et se situent, de fait, à la frontière de la théorie des probabilités (plus particulièrement celle des Équations Différentielles Stochastiques, EDS) et des Équations aux Dérivées Partielles, EDP.

Deux grandes classes d'EDS y sont abordées : les EDS "classiques" dont les coefficients sont peu réguliers, voire singuliers; les EDS de McKean-Vlasov, qui ont la particularité d'intégrer la loi de sa solution à l'équation. Dans tous les cas, les EDP associées sont étudiées. Il s'agit d'EDP du second ordre, avec parfois une certaine structure non linéaire particulière, dont l'espace d'état sous-jacent est éventuellement augmenté de l'espace de Wasserstein.

Ces systèmes sont étudiés en grande partie du point de vue de la régularisation par le bruit, qui peut agir sous différentes formes. Il s'agit de comprendre comment les fluctuations du bruits de l'EDS permettent de s'affranchir des conditions de régularité usuelles pour garantir l'existence, et surtout l'unicité, d'une solution. Dans tous les cas, une étude fine de l'EDP associée s'avère cruciale. Pour la seconde classe, le manuscrit traite aussi d'une généralisation infinie dimensionnelle du problème de Skorokhod.

Mots clés : EDS, EDP, résolubilité, schéma numérique, Mckean-Vlasov, champ moyen

---

Title. ***Bal(l)ade entre EDP et probabilités.***

Absctract. This manuscript aims to trace the work done since my thesis defense at LJAD, Université Nice Côte d'Azur, continued during my assignment at LAMA, Université Savoie Mont Blanc, and then at LMJL, Nantes Université. This work falls within the field of stochastic analysis and lies at the intersection of probability theory (especially Stochastic Differential Equations, SDEs) and Partial Differential Equations, PDEs.

Two main classes of SDEs are addressed: "classical" SDEs with coefficients that are irregular or even singular, and McKean-Vlasov SDEs, which have the peculiarity of integrating the law of their solution into the equation. In all cases, the associated PDEs are studied. These are second-order PDEs, sometimes with a specific nonlinear structure, whose underlying state space may be augmented with the Wasserstein space.

These systems are studied primarily from the perspective of regularization by noise, which can act in various forms. The aim is to understand how the noise fluctuations of the SDEs allow one to overcome the usual regularity conditions to ensure the existence, and especially the uniqueness, of a solution. In all cases, a detailed study of the associated PDE is crucial. For the second class, the manuscript also deals with an infinite-dimensional generalization of the Skorokhod problem.

Key words: SDEs, PDEs, numerical scheme, McKean-Vlasov, mean-field