Electronic Journal of Probability

Electron. J. Probab. **21** (2016), no. 22, 1–41.
ISSN: 1083-6489 DOI: 10.1214/16-EJP4453

# Well-posedness of multidimensional diffusion processes with weakly differentiable coefficients

Dario Trevisan*

**Abstract**

We investigate well-posedness for martingale solutions of stochastic differential equations, under low regularity assumptions on their coefficients, widely extending the results first obtained by A. Figalli in [19]. Our main results are: a very general equivalence between different descriptions for multidimensional diffusion processes, such as Fokker-Planck equations and martingale problems, under minimal regularity and integrability assumptions; and new existence and uniqueness results for diffusions having weakly differentiable coefficients, by means of energy estimates and commutator inequalities. Our approach relies upon techniques recently developed jointly with L. Ambrosio in [6], to address well-posedness for ordinary differential equations in metric measure spaces: in particular, we employ in a systematic way new representations for commutators between smoothing operators and diffusion generators.

**Keywords:** Fokker-Planck equations; martingale problem; DiPerna-Lions flows.
**AMS MSC 2010:** 60J60; 35Q84;.
Submitted to EJP on July 29, 2015, final version accepted on March, 3, 2016.

## 1 Introduction

Aim of this article is to study well-posedness (i.e., existence, uniqueness and stability) for martingale solutions of stochastic differential equations

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \quad t \in (0, T), \tag{1.1}$$

providing in particular new results under low regularity assumptions on the coefficients $b : (0, T) \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : (0, T) \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$.

The classical subject of martingale problems dates back at least to [31], where it was first shown that continuous and uniformly elliptic covariances $a = \sigma\sigma^*$ allow for uniqueness results which had no counterpart in the usual (Itô-)Cauchy-Lipschitz theory, provided that the solution to (1.1) is understood in a sufficiently weak sense. Since then, the theory has been growing, due to its robustness and strong connections with the theory of semigroups and parabolic PDE's, also in abstract (metric) frameworks, see e.g. [17].

*Università degli studi di Pisa, Italy. E-mail: dario.trevisan@unipi.it

Our primary goal here is to show that the techniques originally developed in [6] can be extended to the stochastic theory as well as specialized to the Euclidean setting, to extend in a systematic way the results established in the seminal paper [19]. Actually, most of such techniques, tailored to study well-posedness problems for ordinary differential equations in metric measure spaces (possibly infinite-dimensional) are also well-suited to the study of diffusions in metric measure spaces, as developed in the author's PhD dissertation. However, in this paper, we deal uniquely with Euclidean spaces: among various motivations, besides that a wider audience could be mainly interested in this setting, this allows us to compare new results and techniques with alternative approaches. Finally, Euclidean spaces are a useful "intermediate" step for the infinite dimensional theory, e.g. by cylindrical approximations. Thus, the theory developed here is instrumental to further developments in metric measure spaces.

Therefore, in this article, we adopt the same point of view as in [19], where precise connections between well-posedness of PDE's and martingale problems are settled, in particular for a wide class of diffusion having not necessarily continuous nor elliptic coefficients, provided that some Sobolev regularity holds. Of course, well-posedness has to be understood "in average" with respect to $\mathscr{L}^d$-a.e. initial condition (here and below, $\mathscr{L}^d$ is Lebesgue measure on $\mathbb{R}^d$). More precisely, a formalization akin to that of DiPerna-Lions (see e.g. [3] for an account of the deterministic theory) is introduced, the main objects being Stochastic Lagrangian Flows, i.e., Borel families $(\boldsymbol{\eta}(x))_{x\in\mathbb{R}^d}$ of probability measures on $C([0,T];\mathbb{R}^d)$, such that

i) $\boldsymbol{\eta}(x)$ solves (1.1), starting from $x$ at $t=0$, for $\mathscr{L}^d$-a.e. $x \in \mathbb{R}^d$;

ii) the push-forward measures $(e_t)_\sharp \int \boldsymbol{\eta}(x)\, d\mathscr{L}^d(x)$, where $e_t$ is the evaluation map at $t\in[0,T]$, are absolutely continuous with respect to $\mathscr{L}^d$, with uniformly bounded densities.

Let us stress the fact that, as in the deterministic theory, uniqueness is understood for flows, thus in a selection sense: we are not claiming well-posedness for $\mathscr{L}^d$-a.e. initial datum. Moreover, we remark that, although the conditions above might read as perfect analogues of the notion of Regular Lagrangian flows [3, Definition 13], Stochastic Lagrangian Flows are not necessarily (neither expected to be) deterministic maps of the initial point only; this is evident when $\sigma = 0$ above and any probability concentrated on possibly non-unique solutions to the ODE give rise to a solution to the martingale problem. Despite this discrepancy, such a theory provides rather efficient tools to study stochastic differential equations under low regularity assumptions, in Euclidean spaces, and together with [24], which deals with analogous issues from a PDE point of view, has become the starting point for further developments, among which we quote [28, 26, 18, 36, 37].

Before we proceed with a more detailed description of our results and techniques, let us stress the fact that we are concerned uniquely with martingale problems, so we do not address nor compare our results with those obtained for strong solutions of equations under low regularity assumptions on the coefficients (see the seminal paper [32] and [22, 15] for more recent results). Rigorous correspondences between martingale (or weak) and strong solutions may be provided by the classical Yamada-Watanabe theorem [34] (and extensions, see e.g. [31]). Moreover, the literature on Fokker-Planck equations for general measures is so vast that we must limit ourselves to a comparison of our results only with those which are strongly related and look similar in techniques and mathematical contents: this is done in Section 3.3.

We proceed with a brief description of our contributions developed below, which can be split into two parts, roughly corresponding to Section 2 (together with Appendix A) and Section 3.

In the first part, we investigate the problem of abstract equivalence between "Eulerian" and "Lagrangian" descriptions for multidimensional diffusion processes, where by the former we mean Fokker-Planck equations and the latter consists of solutions to martingale problems. Although such a correspondence can not be considered novel and many ideas can be traced back at least to [1] in the theory of ODE's and DiPerna-Lions flows, as well as [23] for càdlàg martingale problems, to our knowledge, we provide for the first time general results valid under somewhat minimal integrability assumptions on coefficients as well as on solutions. Moreover, we choose to state and prove our results in such a way that they can be translated with a minimal effort to the case of general metric measure spaces, a subject of further research.

In this part, the crucial result is Theorem 2.5, which provides a so-called "superposition principle", i.e., a (non-canonical) way to lift any probability-valued solution of a Fokker-Planck equation to some solution of the corresponding martingale problem. Here, "to lift" means that the 1-marginals of the process which solve the martingale problem coincide with the given solution of the Fokker-Planck equation. Results in a similar spirit appear quite often in the literature (see also the comments just below the statement of Theorem 2.5) and could be traced back to L.C. Young's theory of generalized curves. Technically, one could start from already known results such as [19, Theorem 2.6] or [23, Theorem 4.9.17] to provide a slightly shorter proof, but we preferred to postpone an almost self-contained derivation in Appendix A: indeed, even if we rely on the results quoted above, it turns out that one has to settle non-trivial technical problems. In particular, an underlying result is Theorem A.2, where we establish an estimate for the modulus of continuity of solutions to martingale problems under somewhat minimal integrability assumptions (based on a refined Lévy-type estimate); an alternative but less effective approach, based on fractional Sobolev spaces, was developed in the author's PhD dissertation. Finally, we point out that we exploit a technique originally developed in [6, Theorem 7.1], in case of cylindrical approximations, to move from bounded coefficients to possibly unbounded ones.

In the second part, we address the problem of well-posedness for Fokker-Planck equations, providing sufficient conditions assuming Sobolev regularity of the coefficients. We mainly focus on uniqueness issues, which are settled by means of energy or $L^2$ estimates, formally satisfied by any weak solution, under suitable bounds on the divergence of the driving coefficients: such an approach could be hardly considered novel, as it was already present in [16], for transport equations. However, our main contribution consists in a novel and systematic approach to the estimate of the error terms arising in the approximation procedure, to obtain so-called commutator inequalities: see Section 3.4 for a brief account of the method as well as complete proofs of our crucial resuts. It turns out that, essentially by means of the same technique, we are able to deal with Sobolev derivations (Lemma 3.5), Sobolev diffusions (Lemma 3.6) as well as with time-dependent elliptic diffusions (Lemma 3.7). Such a technique, which ultimately consists in choosing a Markov semigroup as a smoothing operator and relying on duality arguments as well as an interpolation *à la* Bakry-Eméry, has also the advantage of being completely "Eulerian" and "coordinate free". Let us point out that such a technique was first developed in [6] to deal with an analogue problem for derivations in metric measure spaces.

In conclusion, we state and prove two well-posedness results: Theorem 3.1, for diffusions (1.1) having possibly degenerate coefficients, assuming first order Sobolev regularity for the drift $b$ and second order Sobolev regularity for the infinitesimal covariance $a = \sigma\sigma^*$ (together with uniform bounds on their divergence); Theorem 3.3, for the bounded elliptic case, i.e. $\lambda |v|^2 \leq a(v, v) \leq \Lambda |v|^2$ for every $v \in \mathbb{R}^d$, with $t \mapsto a_t$ Lipschitz, where (roughly speaking) regularity assumptions can be reduced of one order (i.e., no assumption on $b$, and first order Sobolev regularity for $a$). Both results entail

Lagrangian counterparts about well-posedness of flows, see e.g. Corollary 3.2. We regard such results as chief examples of the strength and versatility of our techniques for commutator estimates, and we point out that other interesting results could arise in different situations, such as perturbations of elliptic generators which enjoy some ultra- (or hyper-) contractivity features, as well as the case of BV-regular coefficients, that we do not address here.

**Acknowledgments.** The author has been partially supported by PRIN10-11 grant from MIUR for the project Calculus of Variations and by Scuola Normale Superiore, during the project *Studio della buona positura di equazioni di evoluzione deterministiche e stocastiche*. The author is a member of the GNAMPA group of the Istituto Nazionale di Alta Matematica (INdAM). The author thanks his PhD advisors L. Ambrosio and M. Pratelli, for many discussions before and during the writing of this paper, as well as the thesis referees D. Bakry and M. Röckner for their useful comments and constructive criticism, which were taken into great account also while developing this work, in particular with respect to comparison with existing literature. The author thanks an anonymous referee for valuable comments, in particular with regards to details of the proof of Theorem 2.5.

## 2 Diffusion processes and their equivalent descriptions

In this section, we study abstract correspondences between "Eulerian" and "Lagrangian" descriptions for multidimensional diffusion processes, in particular with respect to well-posedness results. The main ideas involved are not entirely novel, but they widely extend those from [19]: here we obtain results under minimal regularity and integrability assumptions. As already remarked in the introduction, on a technical side, a crucial tool is the superposition principle for diffusions, Theorem 2.5, whose proof is deferred to Appendix A.

In Section 2.1, we introduce diffusion operators in $\mathbb{R}^d$, Fokker-Planck equations, martingale problems and flows; in Section 2.2 we study their equivalences.

### 2.1 Definitions and basic facts

Throughout, we use the following notation, for $v$, $w \in \mathbb{R}^d$ ($d \geq 1$) and $A$, $B \in \mathbb{R}^{d\times d}$,

$$v \cdot w = \sum_{i=1}^d v^i w^i, \quad |v|^2 = v\cdot v, \quad (v\otimes w)^{i,j} := v^i w^j, \quad \text{for } i,j \in \{1,\ldots d\},$$

$$A : B = \sum_{i,j=1}^d A^{i,j} B^{i,j}, \quad |A|^2 = A : A, \quad A(v,w) = A : (v\otimes w),$$

and the following notation for differential calculus on $(0,T)\times\mathbb{R}^d$ ($T \in [0,\infty)$):

$$f_t(\cdot) = f(t,\cdot),\ \partial_t f = \frac{\partial f}{\partial t},\ \partial_i f = \frac{\partial f}{\partial x^i},\ \partial_{i,j} f = \frac{\partial^2 f}{\partial x^i \partial x^j},\ \text{for } t\in[0,T],\, i,j\in\{1,\ldots,d\},$$

$$\nabla f = (\partial_i f)_{i=1}^d,\ \nabla^2 f = (\partial_{i,j} f)_{i,j=1}^d,\ \text{thus } b\cdot\nabla f = \sum_{i=1}^d b^i \partial_i f \text{ and } a : \nabla^2 f = \sum_{i,j=1}^d a^{i,j}\partial_{i,j} f,$$

as well as the notation $\mathscr{L}^d$ for Lebesgue measure on $\mathbb{R}^d$ and $\nabla^*$ for the distributional adjoint of $\nabla$ (i.e., $\nabla^* b = -\operatorname{div} b$ on vector fields).

We write $\mathscr{M}(\mathbb{R}^d)$ for the space of signed (real-valued) Borel measures on $\mathbb{R}^d$ (with finite total variation), $\mathscr{M}^+(\mathbb{R}^d) \subseteq \mathscr{M}(\mathbb{R}^d)$ for the cone of finite non-negative measures and $\mathscr{P}(\mathbb{R}^d) \subseteq \mathscr{M}^+(\mathbb{R}^d)$ for the convex set of Borel probability measures on $\mathbb{R}^d$. We say that a curve $\nu = (\nu_t)_{t\in(0,T)} \subseteq \mathscr{M}(\mathbb{R}^d)$ is Borel if, for every Borel set $A \subseteq \mathbb{R}^d$, the curve $t \mapsto \nu_t(A)$ is Borel; we let $|\nu| = (|\nu_t|)_{t\in(0,T)}$ be the curve of total variation measures. A curve $\nu = (\nu_t)_{t\in(0,T)} \subseteq \mathscr{P}(\mathbb{R}^d)$ is narrowly continuous if, for every $f \in C_b(\mathbb{R}^d)$, $t \mapsto \int f d\eta_t$ is continuous.

Most of the quantities that we consider below are integrated with respect to the variable $t \in (0,T)$, with respect to $\mathscr{L}^1|_{[0,T]}$: when $\nu = (\nu_t)_{t\in(0,T)} \subseteq \mathscr{M}(\mathbb{R}^d)$ is a Borel curve, we write $|\nu|\, dt$ for the Borel measure on $(0,T)\times\mathbb{R}^d$, $A \mapsto |\nu|(A) = \int_0^T |\nu_t|(A_t)dt$, for $A \subseteq (0,T)\times\mathbb{R}^d$ Borel. For $p$, $q \in [1,\infty]$ and a Borel curve $\nu = (|\nu_t|)_{t\in(0,T)} \subseteq \mathscr{M}^+(\mathbb{R}^d)$, the space $L^p_t L^q_x(\nu)$ is naturally defined and endowed with the Banach norm

$$\|f\|_{L^p_t L^q_x(\nu)} := \left\| \|f(t,x)\|_{L^q_x(\nu_t)} \right\|_{L^p_t(dt)} < \infty,$$

On the space $C([0,T];\mathbb{R}^d)$ (naturally endowed with the $\sup$ norm and its Borel $\sigma$-algebra), we let $e_t : \gamma \mapsto \gamma_t := \gamma(t) \in \mathbb{R}^d$ be the evaluation map at $t \in [0,T]$. The natural filtration on $C([0,T];\mathbb{R}^d)$ is the increasing family of $\sigma$-algebras $\mathcal{F} = (\mathcal{F}_t)_{t\in[0,T]}$, with $\mathcal{F}_t := \sigma(e_s : s \in [0,t])$. Given $\boldsymbol{\eta} \in \mathscr{P}(C([0,T];\mathbb{R}^d))$, we always let $\eta_t := (e_t)_\sharp \boldsymbol{\eta}$ be the 1-marginal law at $t \in [0,T]$. Notice that the family $\eta := (\eta_t)_{t\in[0,T]} \subseteq \mathscr{P}(\mathbb{R}^d)$ is narrowly continuous.

We let throughout $\mathscr{A} = C^{1,2}_b((0,T)\times\mathbb{R}^d))$ (respectively, $\mathscr{A}_c = C^{1,2}_c((0,T)\times\mathbb{R}^d)))$ be the space of uniformly bounded (respectively, compactly supported) and continuously differentiable functions, once with respect to $t \in (0,T)$ and twice with respect to $x \in \mathbb{R}^d$, with uniformly bounded derivatives (as usual, the superscript $(1,2)$ counts the number of derivatives with respect to $(t,x)$, other superscripts may appear, with natural meaning). We prefer the "abstract" notation $\mathscr{A}$ and large parts of the theory can be developed when "test" functions are replaced by other classes (e.g. as developed throughout the monograph [23]). We endow $\mathscr{A}$ with the norm

$$\|f\|_{C^{1,2}_{t,x}} = \sup_{(t,x)\in(0,T)\times\mathbb{R}^d} \left\{ |f(t,x)| + |\partial_t f(t,x)| + |\nabla f(t,x)| + |\nabla^2 f(t,x)| \right\}.$$

Notice that, by uniform continuity, any $f \in \mathscr{A}$ extends to $[0,T]\times\mathbb{R}^d$.

Throughout, we always let

$$a = (a^{i,j})_{i,j=1}^d : (0,T)\times\mathbb{R}^d \to \mathrm{Sym}_+(\mathbb{R}^d), \quad b = (b^i)_{i=1}^d : (0,T)\times\mathbb{R}^d \to \mathbb{R}^d, \tag{2.1}$$

be Borel, where $\mathrm{Sym}_+(\mathbb{R}^d)$ is the space of symmetric, non-negative definite $d\times d$ matrices.

Next, we define diffusion operators in $\mathbb{R}^d$, measure-valued weak solutions to Fokker-Planck equations and martingale problems on the interval $[0,T]$. Most of these notions are classical (for a brief historical account, see e.g. the introduction of [31]): for the sake of clarity we provide definitions and prove some simple facts.

**Definition 2.1** (diffusion operator)**.** *We let $\mathcal{L}(= \mathcal{L}(a,b))$ be the linear differential operator*

$$\mathscr{A} \ni f \quad \mapsto \quad \mathcal{L}f = \frac{1}{2} a : \nabla^2 f + b \cdot \nabla f,$$

*with values in Borel maps on $(0,T)\times\mathbb{R}^d$.*

We write $\mathcal{L}_t f := (\mathcal{L}f)_t$, for $t \in (0,T)$. As usual, the coefficients $b$, $a$ are referred as the drift of $\mathcal{L}$ and the infinitesimal covariance of $\mathcal{L}$. If $a = 0$, then $\mathcal{L}$ reduces to a linear first-order operator, i.e. a derivation, and we say that we are in the deterministic case.

Given a diffusion operator $\mathcal{L}$, we let the "Eulerian" description of evolution of particles "driven" by $\mathcal{L}$ consist of weak solutions of Fokker-Planck (or forward Kolmogorov) equations, in duality with $\mathscr{A}$. Although our main interest lies in solutions to FPE's that are narrowly continuous curves of probability measures, we introduce more general measure valued solutions, as they are useful, e.g. the space of solutions becomes linear.

**Definition 2.2** (weak solutions of FPE's). *A Borel curve $\nu = (\nu_t)_{t\in(0,T)} \subseteq \mathscr{M}(\mathbb{R}^d)$ is a weak solution of the Fokker-Planck equation (FPE)*

$$\partial_t \nu_t = \mathcal{L}_t^* \nu_t, \quad \text{on } (0,T)\times\mathbb{R}^d, \tag{2.2}$$

*if it holds*

$$\int_0^T \int (|a_t| + |b_t|)\, d|\nu_t| dt < \infty \tag{2.3}$$

*and, for every $f \in \mathscr{A}_c$, it holds*

$$\int_0^T \int \left[\partial_t f(t,x) + \mathcal{L} f(t,x)\right] d\nu_t(x)\, dt = 0. \tag{2.4}$$

With the notation introduced above, condition (2.3) can be restated as $a$, $b \in L^1_{t,x}(|\nu|)$. In what follows, we frequently omit to specify the operator $\mathcal{L}$, that we regard as fixed.

**Remark 2.3.** *A density argument akin to [5, Lemma 8.1.2] allows for proving that any solution $\nu = (\nu_t)_{t\in(0,T)} \subseteq \mathscr{P}(\mathbb{R}^d)$ to (2.2) admits a unique narrowly continuous representative $\tilde\nu = (\tilde\nu)_{t\in[0,T]}$, with $\nu_t = \tilde\nu_t$, for $\mathscr{L}^1$-a.e. $t\in(0,T)$. Thanks to this fact, we may also say that the solution $\nu$ starts from $\nu_0$ (or that $\nu_0$ is the initial law of $\nu$). Moreover, for every $f \in \mathscr{A}$, it holds*

$$\int f_{t_2} d\tilde\nu_{t_2} - \int f_{t_1} d\tilde\nu_{t_1} = \int_{t_1}^{t_2} \int \left[\partial_t f + \mathcal{L}_t f\right] d\nu_t dt, \quad \text{for } t_1,\, t_2 \in [0,T],\, t_1 \le t_2, \tag{2.5}$$

Actually, one first proves that (2.5) holds for $f \in \mathscr{A}_c$ and then extends by density to $f \in \mathscr{A}$. Since this last step requires the introduction of useful cut-off functions, we sketch it here, for later use. For $R \geq 1$, we fix $\chi_R : \mathbb{R}^d \to [0,1]$, a smooth function with $\chi_R(x) = 1$, for $|x| \le R$, $\chi_R(x) = 0$, for $|x| \ge 2R$, such that $|\nabla \chi_R| \le 4R^{-1}$ and $\left|\nabla^2 \chi_R\right| \le 4R^{-2}$. Given $f \in \mathscr{A}$, we let $f_R = f\chi_R \in \mathscr{A}_c$, for which we assume that (2.5) holds. The chain rule entails

$$\mathcal{L}_t f_R = (\mathcal{L}_t f)\chi_R + f_t \mathcal{L}_t \chi_R + a_t(\nabla f_t, \nabla \chi_R), \quad \text{for } t \in (0,T),$$

hence the bound

$$|\mathcal{L}_t f_R| \le |\mathcal{L}_t f| + |f_t|\,|\mathcal{L}_t \chi_R| + |a_t|\,|\nabla f_t|\,|\nabla \chi_R| \le C \left\|f_t\right\|_{C^2_b} \left[|a_t| + |b_t|\right].$$

Letting $R\to\infty$, by dominated convergence, we extend the validity of (2.5).

Next, we introduce solutions of the martingale problem, following [31, Chapter 6]. In particular, we argue directly on the "canonical" space $\Omega = C([0,T];\mathbb{R}^d)$, endowed with the evaluation process $e_t(\gamma) = \gamma(t)$, $t\in[0,T]$, and its natural filtration.

**Definition 2.4** (solution of MP's). *A probability $\boldsymbol{\eta} \in \mathscr{P}(C([0,T];\mathbb{R}^d))$ is a solution of the martingale problem (MP) (associated to $\mathcal{L}$) if it holds*

$$\int \left[\int_0^T (|b_t|\circ e_t + |a_t|\circ e_t)\, dt\right] d\boldsymbol{\eta} < \infty \tag{2.6}$$

*and, for every $f\in\mathscr{A}$, the process*

$$[0,T] \ni t \mapsto f_t \circ e_t - \int_0^t \left[\partial_s f_s + \mathcal{L}_s f\right] \circ e_s ds \tag{2.7}$$

*is a martingale with respect to the natural filtration on $C([0,T];\mathbb{R}^d)$.*

Recall the notation $\eta_t = (e_t)_\sharp \boldsymbol{\eta} \in \mathscr{P}(\mathbb{R}^d)$, $t \in [0,T]$, thus $\eta_0$ is the initial law of $\boldsymbol{\eta}$. As for FPE's, we usually omit to specify $\mathcal{L}$, regarded as fixed. Let us remark that a density argument shows that it makes no difference to require that (2.7) is a martingale only for $f \in \mathscr{A}_c$.

For any solution $\boldsymbol{\eta}$ of the MP, the integrability assumption (2.6), which is equivalent to $a$, $b \in L^1(\eta)$, entails that the process $[0,T] \ni t \mapsto \int_0^t [\partial_s f_s + \mathcal{L}_s f] \circ e_s ds$ is well defined, up to a $\boldsymbol{\eta}$-negligible set, as continuous and progressively measurable process. In particular, it belongs to $L^\infty_{loc}(\boldsymbol{\eta}, (\mathcal{F}_t)_t)$, i.e. there exists an increasing sequence of stopping times $\tau_n$, $\boldsymbol{\eta}$-a.s. converging towards $T$, such that $\int_0^{\tau_n} [\partial_s f_s + \mathcal{L}_s f] \circ e_s ds \in L^\infty(\boldsymbol{\eta})$, for every $n \geq 1$: it is sufficient to let

$$\tau_n := T \wedge \inf\left\{ t \in [0,T] : \int_0^t \left( |b_s| \circ e_s + |a_s| \circ e_s \right) ds \geq n \right\}.$$

We prefer throughout not to enlarge the filtration $\mathcal{F}$ with $\boldsymbol{\eta}$ negligible sets. This causes virtually no harm in the exposition, e.g. a martingale $M = (M_t)_{t\in[0,T]}$ must be understood in the sense that it holds $\mathbb{E}[M_t|\mathcal{F}_s] = M_s$, $\boldsymbol{\eta}$-a.s. for every $s \leq t$ (and the $\boldsymbol{\eta}$-negligible set could not belong to $\mathcal{F}_s$).

When $a = 0$, solutions to the MP reduce to probability measures concentrated on absolutely continuous solutions to the ordinary differential equation

$$\frac{d}{dt}\gamma_t = b_t(\gamma_t), \quad \text{for } \mathscr{L}^1\text{-a.e. } t \in (0,T).$$

Indeed, arguing as in [19, Lemma 3.8], it turns out that the martingale (2.7) is constant. More generally, the quadratic variation process of (2.7) is $t \mapsto \int_0^t a_s(\nabla f_s, \nabla f_s) ds$: this plays a crucial role in estimates for the modulus of continuity of the canonical process, see e.g. Corollary A.5.

By integration of (2.7) with respect to $\boldsymbol{\eta}$ (i.e., taking expectation) we deduce that any solution $\boldsymbol{\eta}$ of the MP induces, by means of its 1-marginals $(\eta_t)_{t\in(0,T)}$ a narrowly continuous solution of the FPE (2.2). A converse statement is provided by the following theorem, whose proof is deferred in Appendix A; in the next section, it plays a crucial role to connect various well-posedness results for FPE's and MP's.

**Theorem 2.5** (superposition principle)**.** *Let $\nu = (\nu_t)_{t\in[0,T]} \subseteq \mathscr{P}(\mathbb{R}^d)$ be a narrowly continuous solution of* (2.2)*. Then, there exists $\boldsymbol{\eta}$ which is a solution to the MP (associated to the same diffusion operator $\mathcal{L}$) such that, for every $t \in [0,T]$, it holds $\eta_t = \nu_t$.*

In what follows, we refer to $\boldsymbol{\eta}$ above as a superposition solution for $\nu$.

We refer to this result as the superposition principle for diffusions: the terminology originates in the deterministic literature of ODE's, see [1]: the solution $\boldsymbol{\eta}$ can be non-trivially distributed among the possibly non-unique solutions to the ODE, thus introducing some "randomness" in an otherwise deterministic setting; these probability measures are nevertheless superpositions of deterministic paths. In the setting of diffusion operators, solutions are already expected to be random, thus the term is justified only by extension, although it would be interesting, at least in some cases, to be able to distinguish between the two "sources of randomness": this would require us to introduce concepts such as strong and weak solutions.

As remarked in the introduction, Theorem 2.5 is a quite general result, only the integrability condition (2.3) being required, which is some sense minimal to give sense to FPE's and MP's (although one may slightly relax it by dealing with local martingale problems). Our result extends [19, Theorem 2.6], where only uniformly bounded coefficients are considered; let us mention that results in a similar spirit – that of L.C. Young's theory of generalized curves – appear quite often in the literature, e.g. Echeverria's theorem [17, Theorem 4.9.17] (see [23] for extensions) in the framework of martingale problems in spaces of càdlàg paths, or Smirnov's decomposition of 1-currents [30] (see also [27] for an alternative approach, valid also in the case of metric currents). Our strategy of proof extends that of [19, Theorem 2.6] and should be regarded as a (non-trivial) counterpart of [5, §8.1 and §8.2] in the setting of multi-dimensional diffusions: although rather natural, the derivation is not immediate from the available literature (both from deterministic and stochastic), due to various technical points. The major difficulty in our proof is to provide estimates for the modulus of continuity of the canonical process (a problem that would appear also if we wanted to deduce it from Echeverria's theorem).

Next, we investigate some stability properties enjoyed by solutions of MP's and FPE's, with respect to suitable operations: their proofs are straightforward, so we omit them.

Clearly, all the definitions above can be given with respect to any interval $[S,T]$ in place of $[0,T]$ (when it is not mentioned, we always refer to the interval $[0,T]$): solutions are then well-behaved with respect to the natural restriction map

$$C([0,T];\mathbb{R}^d) \ni \gamma \mapsto \gamma|_{[S,T]} = (\gamma_t)_{t\in[S,T]} \in C([S,T];\mathbb{R}^d).$$

**Proposition 2.6.** *Let $S$, $T \in \mathbb{R}$, with $0 \le S \le T$, and let $\boldsymbol{\eta} \in \mathscr{P}(C([0,T];\mathbb{R}^d))$ be a solution of the MP. Let $\rho : C([0,T];\mathbb{R}^d) \to [0,\infty)$ be a uniformly bounded probability density (with respect to $\boldsymbol{\eta}$), measurable with respect to $\mathcal{F}_S$.*

*Then, the push-forward $(\rho\boldsymbol{\eta})|_{[S,T]} := \left(|_{[S,T]}\right)_\sharp(\rho\boldsymbol{\eta}) \in \mathscr{P}(C([S,T];\mathbb{R}^d)$ is a solution to the MP associated to $\mathcal{L}$ on $[S,T]$.*

The analogous property for FPE's is obvious: if $(\nu_t)_{t\in(S,T)}$ is a solution of (2.2), its restriction $(\nu_t)_{t\in(S,T)}$ is a solution of the FPE on $(S,T)\times\mathbb{R}^d$.

Solutions of FPE's and MP's are clearly stable with respect to convex combinations, as a consequence of Fubini's theorem.

**Proposition 2.7.** *Let $(Z,\mathcal{A},\bar\nu)$ be a probability space and let $(\boldsymbol{\eta}_z)_{z\in Z} \subseteq \mathscr{P}(C[0,T];\mathbb{R}^d)$ be a Borel family, such that, for $\bar\nu$-a.e. $z\in Z$, $\boldsymbol{\eta}_z$ is a solution of the MP (associated to a fixed diffusion operator $\mathcal{L}$). Moreover, let*

$$\int_Z \int_0^T \int \left(|b_t| + |a_t|\right) d\eta_z dt d\bar\nu(z) < \infty \tag{2.8}$$

*hold. Then, $A \mapsto \boldsymbol{\eta}(A) = \int \boldsymbol{\eta}_z(A)\, d\bar\nu(z)$ is a solution of the MP (associated to $\mathcal{L}$).*

A somewhat converse result, for disintegration with respect to the initial law, is a consequence of stability of martingales under conditional expectations with respect to the $\sigma$-algebra $\mathcal{F}_0$.

**Proposition 2.8.** *Let $\boldsymbol{\eta}$ be a solution of the MP and let $(\boldsymbol{\eta}(x))_{x\in\mathbb{R}^d}$ be a regular conditional probability for $\boldsymbol{\eta}$ with respect to $e_0$. Then, for $\eta_0$-a.e. $x\in\mathbb{R}^d$, $\boldsymbol{\eta}(x)$ is a solution of the MP associated to $\mathcal{L}$, with initial law $\delta_x$.*

We conclude this section by introducing a suitable notion of flow associated to a diffusion operator, roughly consisting of Borel families of solutions of the MP, for a (large, in some sense) set of initial conditions in $\mathbb{R}^d$. Our aim is to study flows in the DiPerna-Lions sense (as extended by Figalli to MP's), thus, we introduce the concept of "regular flow", where regularity is usually some growth and/or absolute continuity condition on the 1-marginals, providing a selection criterion, yielding uniqueness in otherwise possibly ill-posed problems. To study this notion in sufficient generality, we formulate such regularity conditions in terms of some set $\mathcal{R} := \mathcal{R}_{[0,T]}$ of narrowly continuous (probability curves that are) solutions of (2.2), which describe the "admissible" class of dynamics. With this notation, we refer to any $\nu\in\mathcal{R}$ as a $\mathcal{R}$-regular solution of (2.2), and we say that solution to the MP is $\mathcal{R}$-regular if the curve of its 1-marginals is a $\mathcal{R}$-regular solution of (2.2). We also let $\mathcal{R}_0 \subseteq \mathscr{P}(\mathbb{R}^d)$ be the set of all initial laws of the solutions belonging to $\mathcal{R}_{[0,T]}$, which we regard as the set of initial distribution of mass that we are allowed to transport.

**Definition 2.9** ($\mathcal{R}$-MF)**.** *A Borel family $(\boldsymbol{\eta}(x))_{x\in\mathbb{R}^d} \subseteq \mathscr{P}(C([0,T];\mathbb{R}^d))$ is said to be a $\mathcal{R}$-regular martingale flow ($\mathcal{R}$-MF) (associated to $\mathcal{L}$) if the initial law of $\boldsymbol{\eta}(x)$ is $\delta_x$, for every $x \in \mathbb{R}^d$, and, for every $\bar{\nu} \in \mathcal{R}_0$, the probability measure $\int \boldsymbol{\eta}(x)d\bar{\nu}(x)$ is a $\mathcal{R}$-regular solution to the MP (associated to $\mathcal{L}$).*

We remark that we are not imposing that, for every $x \in \mathbb{R}^d$, $\boldsymbol{\eta}(x)$ is a $\mathcal{R}$-regular solution to the MP associated to $\mathcal{L}$; the requirement is only *in average*, with respect to every admissible initial density $\bar{\nu} \in \mathcal{R}_0$. Of course, from this condition and Proposition 2.8 we obtain that $\boldsymbol{\eta}(x)$ is a solution of the MP, for $\bar{\nu}$-a.e. $x \in \mathbb{R}^d$, for every $\bar{\nu} \in \mathcal{R}_0$. For example, if we let $\mathcal{R}_{[0,T]}$ be the set of all narrowly continuous solutions of (2.2), then we operate no selection at all, and $\mathcal{R}$-MF's are Borel selections $(\boldsymbol{\eta}(x))_{x\in\mathbb{R}^d}$ of solutions of the MP, with $\boldsymbol{\eta}(x)$ starting at $\delta_x$, for every $x \in \mathbb{R}^d$. The DiPerna-Lions theory is recovered if we let $\mathcal{R}$ be the set of all narrowly continuous solutions $\nu_t = u_t\mathscr{L}^d \in \mathscr{P}(\mathbb{R}^d)$ of (2.2) with $\|u\|_{L^\infty_{t,x}} < \infty$.

We state (without implicitly assuming) some further properties of $\mathcal{R}$-regular solutions of MP's and FPE's that are useful in the next section. The first property is a stability property with respect to pointwise domination: for every $\tilde{\nu}$, $\nu$, narrowly continuous solution of (2.2) such that, for some $C \geq 0$,

$$\tilde{\nu}_t \leq C\nu_t, \text{ for every } t \in [0,T] \text{ and } \nu \in \mathcal{R}_{[0,T]}, \text{ then } \tilde{\nu} \in \mathcal{R}_{[0,T]}. \tag{2.9}$$

A useful property is stability with respect to convex combinations, i.e., for any $\bar{\nu} \in \mathscr{P}(Z)$,

$$\text{if, } \bar{\nu}\text{-a.e. } z \in Z,\ \boldsymbol{\eta}_z \text{ is } \mathcal{R}\text{-regular and (2.8) holds, then } \int \boldsymbol{\eta}_z d\bar{\nu}(z) \text{ is } \mathcal{R}\text{-regular.} \tag{2.10}$$

A reasonable converse should be stability with respect to disintegration, but there are several formulations: given any $\mathcal{R}$-regular $\boldsymbol{\eta}$, writing $(\boldsymbol{\eta}(x))_{x\in\mathbb{R}^d}$ for a regular conditional probability with respect to $e_0$, we may require that

$$\text{for any } \bar{\nu} \in \mathscr{P}(\mathbb{R}^d) \text{ with } \bar{\nu} \leq C\eta_0 \text{ for some } C > 0, \text{ then } \int \boldsymbol{\eta}(x)\bar{\nu}(x) \text{ is } \mathcal{R}\text{-regular,} \tag{2.11}$$

or alternatively that

$$\text{for any } \bar{\nu} \in \mathcal{R}_0 \text{ with } \bar{\nu} \ll \eta_0, \text{ then } \int \boldsymbol{\eta}(x)\bar{\nu}(x) \text{ is } \mathcal{R}\text{-regular,} \tag{2.12}$$

or even that

$$\text{for } \eta_0\text{-a.e. } x \in \mathbb{R}^d,\ \boldsymbol{\eta}(x) \text{ is a } \mathcal{R}\text{-regular solution to the MP,} \tag{2.13}$$

which is a rather strong condition: it formally implies the others whenever (2.10) holds true. Let us also notice that it does not hold when we deal with the DiPerna-Lions class introduced above, while (2.11) as well as (2.12) hold true. Moreover, an application of Theorem 2.5 shows that condition (2.11) is equivalent to (2.9).

Due to technical reasons, we must introduce a slight extension of all the notions above, taking into account a family $(\mathcal{R}_{[s,T]})_{s\in[0,T]}$, where each $\mathcal{R}_{[s,T]}$ consists of narrowly continuous solutions of the FPE associated to $\mathcal{L}$, on $[s,T]$. Then, we let $\mathcal{R}_s$ be the set of all 1-marginals at time $s$ for solutions belonging to $\mathcal{R}_{[s,T]}$, and we refer to $\mathcal{R}$-regular solutions of FPE's and MP's on $[s,T]$, by natural extension of the definitions given on the interval $[0,T]$. We also assume that

$$\text{for any } r,\ s \in [0,T], \text{ with } r \leq s, \text{ if } \nu \in \mathcal{R}_{[r,T]}, \text{ then } (\nu_t)_{t\in[s,T]} \in \mathcal{R}_{[s,T]}, \tag{2.14}$$

In particular, for any $\nu \in \mathcal{R}_{[r,T]}$, one has $\nu_s \in \mathcal{R}_s$. We also accordingly extend the notion of $\mathcal{R}$-MF by considering a family $(\boldsymbol{\eta}(s,x))_{s\in[0,T],x\in\mathbb{R}^d}$, where $(\boldsymbol{\eta}(s,x))_{x\in\mathbb{R}^d}$ is a $\mathcal{R}$-MF, for every $s \in [0,T]$ (notice that we are not requiring joint measurability of $(s,x) \mapsto \boldsymbol{\eta}(s,x)$).

**Remark 2.10** (Markov property)**.** *With the notation introduced above, we can state the Markov property via Chapman-Kolmogorov equations, for a $\mathcal{R}$-MF $(\boldsymbol{\eta}(s,x))_{s\in[0,T],x\in\mathbb{R}^d}$,*

$$\eta(r,x)_t = \int \eta(s,y)_t \, \eta(r,x)_s(dy), \quad \bar{\nu}\text{-a.e. } x \in \mathbb{R}^d, \text{ for every } \bar{\nu} \in \mathcal{R}_r \tag{2.15}$$

*for every $r$, $s$, $t \in [0,T]$ with $r \leq s \leq t$.*

*We obtain this property as a consequence of uniqueness, arguing e.g. as in [19, Proposition 3.10]. However, let us remark that it could be be of independent interest to study regular flows that are also Markov, extending e.g. the approach in [31, Chapter 12]. Finally, much less is known about the strong Markov property for DiPerna-Lions flows, i.e., the validity of* (2.15) *with stopping times in place of deterministic times – perhaps one has to introduce some notion of "regular" stopping times.*

## 2.2 Equivalence between FPE's, MP's and flows

The superposition principle provided by Theorem 2.5 allows for establishing a neat correspondence between "Eulerian" and "Lagrangian" descriptions, transferring well-posedness results both ways. Such a connection is firmly established in the deterministic case, see e.g. [2, §4], and in the stochastic setting has been investigated e.g. in [19, §2], in case of a DiPerna-Lions theory, and in [17, §4], for the classical theory (i.e., not in a selection sense). In this section, we provide a complete equivalence between well-posedness results for $\mathcal{R}$-regular solutions of FPE's and MP's.

**FPE's $\Leftrightarrow$ MP's.** Equivalence between existence result is straightforward, by lifting any solution $\nu$ of the FPE, we obtain existence of solutions of the MP, so we focus on uniqueness. A simple result which transfers "uniqueness" is the following one: the non trivial implication *ii)* $\Rightarrow$ *i)* follows from lifting two different solutions $\nu^1$, $\nu^2$ (see also [19, Theorem 2.3]).

**Lemma 2.11.** *Let $\bar{\nu} \in \mathcal{R}_0$. Then, the following conditions are equivalent:*

*i) there exists at most one $\mathcal{R}$-regular solution $\nu$ of* (2.2) *with $\nu_0 = \bar{\nu}$;*

*ii) if $\boldsymbol{\eta}^1$, $\boldsymbol{\eta}^2$ are $\mathcal{R}$-regular solutions of the MP with $\eta^1_0 = \eta^2_0 = \bar{\nu}$, then $\eta^1_t = \eta^2_t$, for $t \in [0,T]$.*

A stronger uniqueness result, for processes, can be obtained arguing as in [31, Theorem 6.2.3] or [19, Proposition 5.5]. Let us point out that here there appears a small gap with the deterministic literature, since a different argument [2, Theorem 9] shows uniqueness for MP's assuming only (2.9), while we must consider also intermediate $s \in [0,T]$ (the argument employed therein uses some conditioning which may not preserve the martingale property in general, but it does when the martingale is deterministic).

**Lemma 2.12** (transfer of uniqueness)**.** *Let $\mathcal{R} = (\mathcal{R}_{[s,T]})_{s\in[0,T]}$ satisfy* (2.14) *and* (2.9)*, with $s$ in place of $0$, for every $s \in [0,T]$. Then, the following conditions are equivalent:*

*i) for every $s \in [0,T]$ and $\bar{\nu} \in \mathcal{R}_s$, there exists at most one $\nu \in \mathcal{R}_{[s,T]}$ with $\nu_s = \bar{\nu}$.*

*ii) for every $s \in [0,T]$, if $\boldsymbol{\eta}^1$, $\boldsymbol{\eta}^2$ are $\mathcal{R}$-regular solutions of the MP on $[s,T]$, with $\eta^1_s = \eta^2_s$, then $\boldsymbol{\eta}^1 = \boldsymbol{\eta}^2$.*

*Proof. ii)* $\Rightarrow$ *i).* As in Lemma 2.11, for $\nu \in \mathcal{R}_{[s,T]}$ with $\nu_s = \bar{\nu}$ we consider a ($\mathcal{R}$-regular) superposition solution $\boldsymbol{\eta}$: the uniqueness assumption entails that its 1-marginals are uniquely identified. *i)* $\Rightarrow$ *ii)*. The proof relies (implicitly) on the Markov property. Let $s \in [0,T]$ and $\boldsymbol{\eta}^1$, $\boldsymbol{\eta}^2$ be solutions of the MP on $[s,T]$, with $\eta^1_s = \eta^2_s$. To deduce that $\boldsymbol{\eta}^1 = \boldsymbol{\eta}^2$, we show that, for every $n \geq 1$, the $n$-marginals of $\boldsymbol{\eta}^1$ an $\boldsymbol{\eta}^2$ coincide, i.e., for any $s \leq t_1 < \ldots < t_n \leq T$ and $A_1, \ldots, A_n \subseteq \mathbb{R}^d$ Borel, it holds

$$\boldsymbol{\eta}^1(e_{t_1} \in A_1, \ldots, e_{t_n} \in A_n) = \boldsymbol{\eta}^2(e_{t_1} \in A_1, \ldots, e_{t_n} \in A_n). \tag{2.16}$$

We argue by induction on $n \geq 1$, the case $n = 1$ being a consequence of *i)* $\Rightarrow$ *ii)* in Lemma 2.11 and property (2.14), i.e. we use the fact that $(\eta^i_t)_{t\in[s,T]}$ for $i \in \{1,2\}$ are $\mathcal{R}$-regular solutions, with $\eta^1_s = \eta^2_s$. To perform the step from $n$ to $n+1$, we argue as follows. For fixed $s \leq t_1 < \ldots < t_n < t_{n+1} \leq T$ and $A_1, \ldots, A_n, A_{n+1} \subseteq \mathbb{R}^d$ Borel sets, we let

$$\rho := \frac{\prod_{i=1}^n \chi_{A_i}(e_{t_i})}{\boldsymbol{\eta}^1(e_{t_1} \in A_1, \ldots, e_{t_n} \in A_n)} : C([s,T];\mathbb{R}^d) \to [0,\infty),$$

i.e., the density of $\boldsymbol{\eta}^1$ conditioned with respect to $\bigcap_{i=1}^n \{e_{t_i} \in A_i\}$. We assume that the denominator above is not null: otherwise there is nothing to prove. Notice also that the inductive assumption gives $(e_{t_n})_\sharp(\rho\boldsymbol{\eta}^1) = (e_{t_n})_\sharp(\rho\boldsymbol{\eta}^2)$, since it amounts to (2.16) with $A_n \cap B$ in place of $A_n$, for every $B \subseteq \mathbb{R}^d$ Borel.

For $i \in \{1,2\}$, we let $\boldsymbol{\eta}^i_\rho$ be the push-forward of the measure $\rho\boldsymbol{\eta}^i$ with respect to the natural restriction from $[s,T]$ to $[t_n,T]$, and notice that both are $\mathcal{R}$-regular solutions of the MP on $[t_n,T]$, with identical laws at $t_n$,

$$(\eta^1_\rho)_{t_n} = (e_{t_n})_\sharp(\rho\boldsymbol{\eta}^1) = (e_{t_n})_\sharp(\rho\boldsymbol{\eta}^2) = (\eta^2_\rho)_{t_n},$$

by Lemma 2.6 and (2.14). By the implication *i)* $\Rightarrow$ *ii)* in Lemma 2.11, we deduce in particular that $(\eta^1_\rho)_{t_{n+1}} = (\eta^2_\rho)_{t_{n+1}}$, thus

$$\frac{\boldsymbol{\eta}^1(e_{t_1} \in A_1, \ldots, e_{t_n} \in A_n, e_{t_{n+1}} \in A_{n+1})}{\boldsymbol{\eta}^1(e_{t_1} \in A_1, \ldots, e_{t_n} \in A_n)} = \frac{\boldsymbol{\eta}^2(e_{t_1} \in A_1, \ldots, e_{t_n} \in A_n, e_{t_{n+1}} \in A_{n+1})}{\boldsymbol{\eta}^2(e_{t_1} \in A_1, \ldots, e_{t_n} \in A_n)},$$

hence we deduce the case $n+1$ of (2.16). □

**MP's $\Leftrightarrow$ flows.** In this case, both notions are "Lagrangian", thus there is no need of the superposition principle here: most of the argument are just consequences of convexity and disintegration of measures.

Although our actual well-posedness results are in the DiPerna-Lions case, where uniqueness is understood up to $\mathfrak{m}$-a.e. equivalence, where $\mathfrak{m}$ is some "reference" $\sigma$-finite Borel measure on $\mathbb{R}^d$ (i.e., $\mathfrak{m} = \mathscr{L}^d$), for the sake of completeness, we provide a result assuming (2.13).

**Proposition 2.13.** *Consider the following conditions:*

*i) for every $\bar\nu \in \mathcal{R}_0$, there exists a unique $\mathcal{R}$-regular solution $\boldsymbol{\eta}^{\bar\nu}$ to the MP, with initial law $\bar\nu$, and the map $\bar\nu \mapsto \boldsymbol{\eta}^{\bar\nu}$ is Borel;*

*ii) for every $\bar\nu \in \mathcal{R}_0$ and $\mathcal{R}$-MF's $(\boldsymbol{\eta}^1(x))_{x\in\mathbb{R}^d}$, $(\boldsymbol{\eta}^2(x))_{x\in\mathbb{R}^d}$, one has $\boldsymbol{\eta}^1 = \boldsymbol{\eta}^2$, $\bar\nu$-a.e. on $\mathbb{R}^d$.*

*Then, it always holds* i) $\Rightarrow$ ii)*, while* ii) $\Rightarrow$ i) *holds true provided that some $\mathcal{R}$-MF exists and both* (2.10) *and* (2.13) *hold.*

*Proof.* *i)* $\Rightarrow$ *ii)* is straightforward, since regular conditional probabilities are essentially unique (a $\mathcal{R}$-MF is in particular a regular conditional probability of $\int \boldsymbol{\eta}(x)d\bar\nu(x)$ with respect to $e_0$).

To show the implication *ii)* $\Rightarrow$ *i)*, let $\bar\nu \in \mathcal{R}_0$ and $\boldsymbol{\eta}^1$, $\boldsymbol{\eta}^2$ be $\mathcal{R}$-regular solutions of the MP, with initial law $\bar\nu$. By disintegrating with respect to $e_0$ and (2.13) we may assume that $\bar\nu = \delta_{\bar x}$, for some $\bar x \in \mathbb{R}^d$. Let $(\boldsymbol{\eta}(x))_{x\in\mathbb{R}^d}$ be a $\mathcal{R}$-MF (here we use the existence assumption) and define

$$\boldsymbol{\eta}^i(x) := \chi_{\{x\neq\bar x\}}\boldsymbol{\eta}(x) + \chi_{\{x=\bar x\}}\boldsymbol{\eta}^i, \quad \text{for } i\in\{1,2\},$$

which are two different $\mathcal{R}$-MF's since, for any $\bar\mu \in \mathcal{R}_0$, it holds, by (2.10), $\int \boldsymbol{\eta}^i(x)d\bar\mu(x) = \int_{\{x\neq\bar x\}} \boldsymbol{\eta}(x)d\bar\mu(x) + \bar\mu(\bar x)\boldsymbol{\eta}^i \in \mathcal{R}$. □

The result above is rather unsatisfactory in terms of existence of $\mathcal{R}$-MF's, which seems a delicate problem, in general. For example, existence may follow if one assumes the validity of assumption *i)*, (2.10), (2.13) and that $\mathcal{R}_0$ is a Borel of probability measures. Then, for every $x\in\mathbb{R}^d$ such that $\delta_x\in\mathcal{R}_0$, there exists a unique $\mathcal{R}$-regular solution of the MP, and by suitable definition for $x$ not in such a set, we obtain a $\mathcal{R}$-MF (which is then unique in the sense above). An easier existence result follows if we assume the domination condition

$$\text{for some } \sigma\text{-finite measure } \mathfrak{m}, \text{ it holds } \bar\nu \ll \mathfrak{m}, \text{ for every } \bar\nu\in\mathcal{R}_0, \tag{2.17}$$

as in the DiPerna-Lions case. We also assume that $\mathfrak{m}$ is minimal in the sense that, for every $A\subseteq\mathbb{R}^d$ Borel with $\mathfrak{m}(A)>0$, there exists some $\bar\nu\in\mathcal{R}_0$ with $\bar\nu(A)>0$.

**Proposition 2.14.** *Let* (2.17) *hold, and consider the following conditions:*

*i) for every $\bar\nu\in\mathcal{R}_0$, there exists a unique $\mathcal{R}$-regular solution $\boldsymbol{\eta}^{\bar\nu}$ to the MP, with initial law $\bar\nu$, and the map $\bar\nu\mapsto\boldsymbol{\eta}^{\bar\nu}$ is Borel;*

*ii) there exists a $\mathcal{R}$-MF's $(\boldsymbol{\eta}(x))_{x\in\mathbb{R}^d}$ and $\mathcal{R}$-MF's are $\mathfrak{m}$-a.e. unique, i.e. if $(\boldsymbol{\eta}^1(x))_{x\in\mathbb{R}^d}$ and $(\boldsymbol{\eta}^2(x))_{x\in\mathbb{R}^d}$ are $\mathcal{R}$-MF's, then $\boldsymbol{\eta}^1=\boldsymbol{\eta}^2$, $\mathfrak{m}$-a.e. in $\mathbb{R}^d$.*

*If* (2.10) *and* (2.11) *holds, then* i) $\Rightarrow$ ii). *If* (2.10) *and* (2.12) *are satisfied, then* ii) $\Rightarrow$ i).

*Proof. i) $\Rightarrow$ ii).* We have only to settle existence of some $\mathcal{R}$-MF, as uniqueness is trivial. For any probability $\bar\nu = u\mathfrak{m}\in\mathcal{R}_0$ we consider the unique $\mathcal{R}$-regular solution of the MP $\boldsymbol{\eta}^u$ with initial law $\bar\nu$ and a regular conditional probability with respect to $e_0$, $(\boldsymbol{\eta}^u(x))_{x\in\mathbb{R}^d}$. Then, for any $v\mathfrak{m}\in\mathcal{R}_0$, it holds

$$\boldsymbol{\eta}^u(x) = \boldsymbol{\eta}^v(x), \quad \mathfrak{m}\text{-a.e. } x\in X \text{ such that } u(x)>0 \text{ and } v(x)>0. \tag{2.18}$$

Indeed, it is sufficient to show that, for every $\varepsilon>0$ and every $\rho\mathfrak{m}$ probability density, concentrated on $\{u>\varepsilon, v>\varepsilon\}$, with $\rho$ uniformly bounded, it holds

$$\int \boldsymbol{\eta}^u(s)\rho(x)d\mathfrak{m}(x) = \int \boldsymbol{\eta}^v(x)\rho(x)d\mathfrak{m}(x).$$

This, in turn, follows from uniqueness and (2.11): both members above are $\mathcal{R}$-regular solutions to the MP, with initial law $\rho\mathfrak{m}\le\varepsilon^{-1}Cu\mathfrak{m}$.

Next, we notice that there must exists some $u\mathfrak{m}\in\mathcal{R}_0$ equivalent to $\mathfrak{m}$, i.e., such that $u>0$ $\mathfrak{m}$-a.e. in $\mathbb{R}^d$, since $\mathfrak{m}$ is equivalent to the supremum of all the measures in $\mathcal{R}_0$ (appropriately rescaled). Then, we define $\boldsymbol{\eta}(x) := \boldsymbol{\eta}^u(x)$, for $x\in\mathbb{R}^d$. To conclude that $\boldsymbol{\eta}(x)$ is a $\mathcal{R}$-MF, we use (2.18): given any probability $v\mathfrak{m}\in\mathcal{R}_0$, it holds

$$\int \boldsymbol{\eta}(x)v(x)d\mathfrak{m} = \int \boldsymbol{\eta}^v(x)v(x)d\mathfrak{m} = \boldsymbol{\eta}^v.$$

To prove *ii) $\Rightarrow$ i)*, existence of $\mathcal{R}$-regular solutions to the MP, given the existence of a $\mathcal{R}$-MF is trivial, so we focus on uniqueness. We let $\tilde{\boldsymbol{\eta}}$, be a $\mathcal{R}$-regular solution of the MP with some initial law and show that it must coincide with the one induced by the (unique)

$\mathcal{R}$-MF $(\boldsymbol{\eta}(x))_{x\in\mathbb{R}^d}$, i.e., $\tilde{\boldsymbol{\eta}} = \int \boldsymbol{\eta}(x) d\tilde{\eta}_0(x)$. To this aim, we let $u\mathfrak{m} \in \mathcal{R}_0$ be a probability measure equivalent to $\mathfrak{m}$, and consider the measure

$$\overline{\boldsymbol{\eta}} := \frac{1}{2}\tilde{\boldsymbol{\eta}} + \frac{1}{2}\int \boldsymbol{\eta}(x)u(x)d\mathfrak{m}(x),$$

which is a $\mathcal{R}$-regular solution to the $MP$ by (2.10), whose initial law is again equivalent to $\mathfrak{m}$. By disintegration with respect to $e_0$, we obtain a Borel family of probability measures $(\overline{\boldsymbol{\eta}}(x))_{x\in\mathbb{R}^d}$, which, by (2.12), provides a $\mathcal{R}$-MF and so by uniqueness it coincides with $\boldsymbol{\eta}(x)$, for $\mathfrak{m}$-a.e. $x \in \mathbb{R}^d$, yielding

$$\int \boldsymbol{\eta}(x)\left[\frac{1}{2}d\tilde{\eta}_0(x) + \frac{1}{2}u(x)d\mathfrak{m}(x)\right] = \frac{1}{2}\tilde{\boldsymbol{\eta}} + \frac{1}{2}\int \boldsymbol{\eta}(x)u(x)d\mathfrak{m}(x)$$

from which we conclude. □

We end this section with some remarks on standard consequences of uniqueness: the Markov property and stability with respect to approximation.

As in Remark 2.10, we consider $\mathcal{R}$-regular flows with respect to a family $(\mathcal{R}_{[s,T]})_{s\in[0,T]}$ such that (2.14) holds.

**Proposition 2.15** (Markov property)**.** *Assume that uniqueness holds for $\mathcal{R}$-regular MP's, in the sense that, for every $s \in [0,T]$, $\bar{\nu} \in \mathcal{R}_s$, there exists a unique $\mathcal{R}$-regular solution to the MP on $[s,T]$, with initial law $\bar{\nu}$. Then, for every $\mathcal{R}$-MF $(\boldsymbol{\eta}(s,x))_{s\in[0,T],x\in\mathbb{R}^d}$,* (2.15) *holds true, for every $r$, $s$, $t \in [0,T]$, with $r \le s \le t$.*

The proof is straightforward from the following identity between measures on $C([s,T];\mathbb{R}^d)$:

$$(|_{[s,T]})_\sharp\left[\int \boldsymbol{\eta}(r,x)\bar{\nu}(dx)\right] = \int \boldsymbol{\eta}(s,y)\left[\int \eta(r,x)_s\bar{\nu}(dx)\right](dy),$$

which, in turn, holds true because both terms define $\mathcal{R}$-regular solutions of the MP on $[s,T]$, with initial law $\int \eta(r,x)_s\bar{\nu}(dx)$: this is obvious for the right hand side, while for the left hand side it is a consequence Proposition 2.6 and condition (2.14).

Another well understood, but rather technical, property that sometimes follows from existence and uniqueness is a non-quantitative version of stability with respect to approximations, which in this setting would read as follows.

**Proposition 2.16** (stability)**.** *For $n \ge 1$, let $a^n$, $b^n$ be Borel maps as in* (2.1)*, let $\mathcal{L}^n := \mathcal{L}(a^n,b^n)$ and let $\boldsymbol{\eta}^n$ solve the MP associated to $\mathcal{L}^n$. If*

*i) there exists a unique $\mathcal{R}$-regular solution $\boldsymbol{\eta}$ of the MP associated to $\mathcal{L} = \mathcal{L}(a,b)$ with $\eta_0 = \bar{\nu}$,*

*ii) it holds $\eta_0^n \to \bar{\nu}$ narrowly, $a^n \to a$ and $b^n \to b$ pointwise as $n \to \infty$,*

*iii) for some convex, l.s.c functions $\Theta_1$, $\Theta_2$ as in Theorem A.2 it holds*

$$\limsup_{n\to\infty}\int_0^T\int \Theta_1\left(|b_t^n|\right) + \Theta_2\left(|a_t^n|\right)d\eta_t^n dt \le \int_0^T\int \Theta_1\left(|b_t^n|\right) + \Theta_2\left(|a_t^n|\right)d\eta_t dt,$$

*iv) every limit point in $C([0,T];\mathscr{P}(\mathbb{R}^d))$ of $(\eta^n)_{n\ge1}$ belongs to $\mathcal{R}$,*

*then $\boldsymbol{\eta}^n \to \boldsymbol{\eta}$ narrowly in $\mathscr{P}(C([0,T];\mathbb{R}^d))$.*

Notice that we do not require that $\boldsymbol{\eta}^n$ are $\mathcal{R}$-regular: in general it does not even make sense, since $\mathcal{R}$ is a class of solutions to the FPE associated to $\mathcal{L}$, not to $\mathcal{L}^n$. A proof of the result above would not be difficult, but it would require us to combine some technical results, such as those established in Section A.2 and [3, Lemma 23] to establish that $(\boldsymbol{\eta}^n)$ is a tight sequence and any limit point provides a $\mathcal{R}$-regular solution to the MP associated to $\mathcal{L}$; the conclusion is then straightforward from uniqueness. If (2.17) holds, then $\mathfrak{m}$-a.e. convergence in place of pointwise convergence of the coefficients is sufficient, if we also restrict to solutions $\boldsymbol{\eta}^n$ whose marginals are absolutely continuous (as done, e.g. in [19, Theorem 3.7]).

## 3 Well-posedness results

In this section, we state and prove two results (Theorem 3.1 and Theorem 3.3) about existence and uniqueness for solutions of the FPE (2.2), belonging to suitable classes of probability measures. In particular, as we are interested in the DiPerna-Lions theory, we deal with absolutely continuous with respect to the $d$-dimensional Lebesgue measure, $\mu_t = u_t \mathscr{L}^d$, satisfying some bounds on their density $u : [0,T] \times \mathbb{R}^d \to \mathbb{R}$. Besides such integrability conditions on the solution, we require (Sobolev) regularity assumptions on the coefficients $a$, $b$.

In Section 3.1, we give some formal derivation of the energy estimates which eventually lead to well-posedness for FPE's; in Section 3.2, we introduce the notation for Sobolev spaces and related basic facts; in Section 3.3 we state our results and compare their with some (related) existing literature; the technical heart of the matter is developed in Section 3.4, where crucial commutator inequalities are proved; in Section 3.5 we give a proof of main our results.

### 3.1 Energy estimates and renormalized solutions

As in the classical DiPerna-Lions theory (as well as in [19]), we rely on energy inequalities satisfied by an absolutely continuous solution $u = (u_t)_{t\in[0,T]}$ of (2.2) (i.e. $\mu_t = u_t \mathscr{L}^d$). Let us briefly sketch a formal derivation, where we assume all the quantities involved being smooth (solutions and coefficients).

The main idea is to write the equation satisfied by $t \mapsto \int \beta(u_t(x))dx$, where $\beta : \mathbb{R} \mapsto \mathbb{R}$ is a smooth function (from a Lagrangian viewpoint, this amounts in choosing, as a test function, an expression involving the density of the solution $u$ itself). The chain rule gives $\partial_t \beta(u) = \beta'(u)\mathcal{L}^*(u)$ and, by linearity, we consider separately the drift and diffusion terms. Straightforward calculus gives

$$\beta'(u)(b\cdot\nabla)^* u = \beta'(u)\nabla^*(bu) = \beta'(u)u\nabla^* b + \beta'(u) b\cdot\nabla u = (b\cdot\nabla)^*\beta(u) + \left[\beta'(u)u - \beta(u)\right]\nabla^* b.$$

For the diffusion part, we first notice the identity $(a:\nabla^2)^* u = -\nabla^*(a\cdot\nabla u) + \nabla^*((\nabla^* a)u)$, where $\nabla^* a$ is the vector field $(\nabla^* a)_i = -\sum_{j=1}^d \partial_j a_{i,j}$, for $i \in \{1,\ldots,d\}$. Therefore, we obtain

$$\begin{aligned}
&\beta'(u)(a:\nabla^2)^* u = \\
&\quad = -\beta'(u)\nabla^*(a\cdot\nabla u) + \beta'(u)\nabla^*((\nabla^* a)u) \\
&\quad = -\beta'(u)\nabla^*(a\cdot\nabla u) + \nabla^*((\nabla^* a)\beta(u)) + \left[\beta'(u)u - \beta(u)\right](\nabla^*)^2 a \\
&\quad = -\nabla^*(a\cdot\nabla\beta(u)) - \beta''(u) a(\nabla u,\nabla u) + \nabla^*((\nabla^* a)\beta(u)) + \left[\beta'(u)u - \beta(u)\right](\nabla^*)^2 a \\
&\quad = (a:\nabla^2)^*\beta(u) - \beta''(u)a(\nabla u,\nabla u) + \left[\beta'(u)u - \beta(u)\right](\nabla^*)^2 a
\end{aligned}$$

where $(\nabla^*)^2 a = \nabla^*(\nabla^* a) = \sum_{i,j}\partial^2_{i,j} a_{i,j}$. Summing up, we have the identity

$$\partial_t\beta(u) = \mathcal{L}^*(\beta(u)) - \frac{\beta''(u)}{2} a(\nabla u,\nabla u) - \left[\beta'(u)u - \beta(u)\right](\operatorname{div}\mathcal{L}), \tag{3.1}$$

where $\operatorname{div}\mathcal{L} := -\nabla^* b - (\nabla^*)^2 a/2$. By integrating over $\mathbb{R}^d$, (formally $\mathcal{L}1 = 0$), we deduce

$$\partial_t \int \beta(u) d\mathscr{L}^d = -\int \frac{\beta''(u)}{2} a(\nabla u,\nabla u) d\mathscr{L}^d - \int \left[\beta'(u)u - \beta(u)\right](\operatorname{div}\mathcal{L})\, d\mathscr{L}^d.$$

If $\beta$ is convex with $\beta(0)=0$, so $\beta''(z)\geq 0$ and $\beta'(z)z-\beta(z)\geq 0$, for $z\in\mathbb{R}$, we obtain

$$\partial_t \int \beta(u) d\mathscr{L}^d \leq \int \left[\beta'(u)u-\beta(u)\right](\operatorname{div}\mathcal{L})^- \, d\mathscr{L}^d, \tag{3.2}$$

which is the key inequality we employ to show existence as well as uniqueness, for suitable choices of $\beta$. For example, letting $\beta(z)=|z|^+$, we deduce that if $u_0\geq 0$, then $u_t\geq 0$ for $t\in[0,T]$ (thus, for simplicity, we assume that $u_t\geq 0$ in what follows). In particular, to deduce uniqueness for solutions in $L^\infty_t(L^r_x)$ (for some $r>1$), we show that the difference between any solutions $u$, $v$ satisfies (3.2), with $\beta(z)=|z|^r$, and Gronwall lemma entails a uniform bound with respect to $t\in(0,T)$ for $\|u_t-v_t\|_{L^r_x}$. Let us also notice that, with the choice $\beta(z)=|z|^2$, keeping track of the non-negative terms dropped above, we would obtain a bound for the "Sobolev energy" $\int_{\mathbb{R}^d} a_t(\nabla u_t,\nabla u_t) d\mathscr{L}^d dt$ and, for $\beta(z)=|z|^r$, with $r>2$, of the energy

$$r(r-1)\int_0^T\int_{\mathbb{R}^d} u_t^{r-2} a_t(\nabla u_t,\nabla u_t) d\mathscr{L}^d dt = \frac{r(r-1)}{(r/2-1)^2}\int_0^T\int_{\mathbb{R}^d} a_t(\nabla u_t^{r/2},\nabla u_t^{r/2}) d\mathscr{L}^d dt$$

In the elliptic case, i.e. if there exists some constant $\lambda>0$ with $a(v,v)\geq\lambda|v|^2$ for every $v\in\mathbb{R}^d$, uniformly in $(0,T)\times\mathbb{R}^d$, we would deduce that any weak solution $u$ actually belongs to the Sobolev space $L^2_t(W^{1,2}_x)$. Moreover, if we have no bounds on $\operatorname{div}\mathcal{L}$ but only on $(\nabla^*)^2 a$, and $b\in L^1_t(L^\infty_x)$, we may still deduce some bound with respect to the energy $z\mapsto|z|^2$,

$$2\int u_t b_t\cdot\nabla u_t d\mathscr{L}^d \leq \frac{\lambda}{2}\int|\nabla u_t|^2 d\mathscr{L}^d + \frac{4\|b_t\|_\infty}{\lambda}\int|u_t|^2 d\mathscr{L}^d,$$

so that

$$\partial_t\int|u_t|^2 d\mathscr{L}^d \leq \int|u_t|^2\left[\left((\nabla^*)^2 a_t\right)^+ + \frac{4\|b_t\|_\infty}{\lambda}\right] d\mathscr{L}^d - \frac{\lambda}{2}\int|\nabla u_t|^2 d\mathscr{L}^d,$$

and again Grownwall inequality leads to a bound for $L^2_x$, uniform in $t\in[0,T]$. Similarly, if $r>2$, we use the inequality $2ab\leq a^2+b^2$ thus, for every $\varepsilon>0$, the term $r\int u_t^{r-1} b_t\cdot\nabla u_t d\mathscr{L}^d$ (assume for simplicity that $u$ is non-negative) is estimated with

$$\frac{r}{r/2-1}\int u_t^{r/2} b_t\cdot\nabla u_t^{r/2} d\mathscr{L}^d \leq \frac{\varepsilon}{2}\int|\nabla u_t^{r/2}|^2 d\mathscr{L}^d + \frac{r^2\|b_t\|_\infty}{2(r/2-1)^2\varepsilon}\int|u_t|^r d\mathscr{L}^d,$$

and letting $\varepsilon=\lambda$, we may conclude again by a Grownall argument that

$$\sup_{t\in[0,T]}\|u_t\|_{L^r_x}\leq\|u_0\|_{L^r_x}\exp\left\{\left(1-\frac{1}{r}\right)\left\|((\nabla^*)^2 a)^+\right\|_{L^1_tL^\infty_x} + \frac{r}{2(r/2-1)^2\lambda}\|b\|_{L^1_tL^\infty_x}\right\} \tag{3.3}$$

Let us finally remark that if we integrate (3.1) with respect to some function $f\in\mathcal{A}$, with $f\geq 0$ (instead of $f=1$), we would deduce

$$\partial_t\int f\beta(u) d\mathscr{L}^d \leq \int(\partial_t+\mathcal{L}f)\beta(u) d\mathscr{L}^d + \int f\left[\beta'(u)u-\beta(u)\right](\operatorname{div}\mathcal{L})^- \, d\mathscr{L}^d. \tag{3.4}$$

The inequality above is so useful that weak solutions $u$ of the FPE, which also satisfy (3.4) for every $f\in\mathcal{A}$, $f\geq 0$, for (many) smooth convex functions $\beta$, are called *renormalized solutions* [19, Definition 4.9]. There are abstract results connecting well-posedness for FPE's and the fact that every weak solution is renormalized, e.g. [19, Lemma 4.10] (but see also [12] for a somewhat converse result, in the deterministic framework); here, for brevity, we limit ourselves to a direct proof of uniqueness of FPE's from the validity of (3.1), e.g. with the special choice $\beta(z)=|z|^r$.

### 3.2 Sobolev spaces

Before we state and prove our main results, we briefly introduce Sobolev spaces associated to the operators $\partial_t$ and $\mathcal{L}$, together with some useful facts; we use throughout a compact notation extending that in Section 2.1.

For $p$, $q \in [1,\infty]$, the space $W^{1,p}_t(L^q_x)$ is defined as the space of functions $u \in L^p_t(L^q_x)$ such that the distributional derivative $\partial_t u$ is represented by a (unique) $g \in L^p_t(L^q_x)$,

$$\int_0^T \int_{\mathbb{R}^d} (\partial_t f) u_t d\mathscr{L}^d dt = -\int_0^T \int_{\mathbb{R}^d} f_t g_t d\mathscr{L}^d dt, \quad \text{for every } f \in \mathcal{A}_c.$$

We endow $W^{1,p}_t(L^q_x)$ with the Banach norm $\|u\|_{L^p_t L^q_x} + \|\partial_t u\|_{L^p_t L^q_x}$. A standard mollification argument, with respect to the variable $t \in (0,1)$, gives that $\mathcal{A}$ is dense in $W^{1,p}_t(L^q_x)$, for $p$, $q < \infty$ (for a proof of this and the following results, we refer e.g. to [29, §III.1]). In particular, the chain rule for $\partial_t$ extends to $W^{1,p}_t(L^q_x)$, thus

$$\partial_t(f\beta(u)) = (\partial_t f)\beta(u) + f\beta'(u)\partial_t u, \quad \text{for every } f \in \mathcal{A},\ u \in W^{1,p}_t(L^q_x),\ \beta \in C^1_b(\mathbb{R}).$$

Another straightforward consequence of the density of $\mathcal{A}$ is the fact that any $u \in W^{1,p}_t(L^q_x)$ enjoys an absolutely continuous representative, i.e. there exists some $\tilde{u} \in AC^p([0,T]; L^q(\mathscr{L}^d))$ such that $\tilde{u}_t = u_t$, for $\mathscr{L}^1$-a.e. $t \in (0,T)$. In particular the map: $T_0(u) := \tilde{u}_0$ (trace at 0) is linear and continuous from $W^{1,p}_t(L^q_x)$ to $L^q_x$. Moreover, $t \mapsto \tilde{u}_t$ is strongly differentiable at $\mathscr{L}^1$-a.e. $t \in (0,T)$ and it holds $\frac{d}{dt}\tilde{u} = \partial_t u$.

We associate to the diffusion operator $\mathcal{L}$ some "Sobolev spaces". An important role in our deductions is played by $D^p(\mathcal{L})$ (for $p \in [1,\infty)$), defined as the abstract completion of $\mathcal{A}$ with respect to the norm $\|f\|_{D^p(\mathcal{L})} := \|f\|_{L^1_t L^p_x} + \|\mathcal{L} f\|_{L^1_t L^p_x}$, which is well defined whenever $a$, $b \in L^1_t L^p_x$ (actually, a more consistent notation for $D^p(\mathcal{L})$ would be $L^1_t(D^p(\mathcal{L}_t))$). Let us remark however that, without further regularity assumptions, the extended operator $D^p(\mathcal{L}) \ni f \mapsto \mathcal{L} f \in L^1_t(L^p_x)$ may be multi-valued, but the assumptions on $\mathcal{L}$ that we impose in our results entail that the extension is single-valued.

A useful fact is the following: if $f \in W^{1,1}_t(L^p_x) \cap D^p(\mathcal{L})$, then one can provide a sequence $(f_n)_{n\geq 1} \subseteq \mathcal{A}$ converging towards $f$ both in $W^{1,1}_t(L^p_x)$ and $D^p(\mathcal{L})$. Indeed, it is sufficient to consider first a sequence $(g_n)_{n\geq 1} \subseteq \mathcal{A}$ converging towards $f$ in $D^p(\mathcal{L})$, let $\rho$ be a smooth probability density on $\mathbb{R}$, and consider the approximation $g_{n,m} := g_n * \rho_m$ (where we let $\rho_m(t) = m\rho(mt)$, $t \in \mathbb{R}$, and we carefully extend $g_n$ to a continuous function outside the set $[0,T] \times \mathbb{R}^d$). For $(n,m) \to \infty$, the sequence $g_{n,m}$ converges towards $f$ in $D^p(\mathcal{L})$, because $g \mapsto g * \rho_m$ is a contraction in $D^p(\mathcal{L})$, as convolution with respect to $t$ and the operator $\mathcal{L}$ commute; for fixed $m \geq 1$, the sequence $g_{n,m}$ converges towards $f * \rho_m$, because $g \mapsto g * \rho_m$ is continuous from $L^1_t(L^p_x)$ into $W^{1,1}_t(L^p_x)$, with norm smaller than $\|\rho_m\|_\infty$. Moreover, as $m \to \infty$, $f * \rho_m$ converges towards $f$ in $W^{1,1}_t(L^p_x)$, since $f \in W^{1,1}_t(L^p_x)$ (this is exactly the standard mollification argument providing density of $\mathcal{A}$ in $W^{1,1}_t(L^p_x)$). By a diagonal argument, we finally extract a sequence $(f_n)_{n\geq 1}$ as required. As a consequence, if $u \in L^\infty_t L^r_x$ ($r > 1$) is a narrowly continuous solution of (2.2), with $a$, $b \in L^1_t L^p_x$, then the weak formulation (2.4) extends to $f \in W^{1,1}_t(L^{r'}_x) \cap D^{r'}(\mathcal{L})$:

$$\int_0^T \int \left[(\partial_t + \mathcal{L}_t) f\right] u_t d\mathscr{L}^d dt = \int f_T u_T d\mathscr{L}^d - \int f_0 u_0 d\mathscr{L}^d, \tag{3.5}$$

where by $f_T \in L^{r'}_x$ and $f_0 \in L^{r'}_x$ we mean the continuous representative of $f$ evaluated at $T$ and 0.

Similarly, we introduce the space $D^p(\mathcal{L}, a\nabla \otimes \nabla)$ as the abstract completion of $\mathcal{A}$ with respect to the norm $\||f| + |\mathcal{L} f| + a(\nabla f, \nabla f)\|_{L^1_t L^p_x}$. Clearly, this could be a space smaller than $D^p(\mathcal{L})$ and it is useful because the following chain rule holds, for $D^p(\mathcal{L}, a\nabla \otimes \nabla)$, and $\gamma \in C^2(\mathbb{R})$, with $\gamma'$ and $\gamma''$ uniformly bounded:

$$\mathcal{L}(\gamma(u)) = \gamma'(u)\mathcal{L}(u) + \gamma''(u)a(\nabla u, \nabla u).$$

As in the previous case, it might be that $f \mapsto \mathcal{L}(f)$ and $f \mapsto a(\nabla f, \nabla f)$ are not single-valued, but the identity above holds true with the natural interpretation (and in our results we introduce assumptions ensuring that these are well-defined functions).

We always consider the divergence $\operatorname{div} \mathcal{L}$ be defined in the sense of distributions, i.e., as the linear operator $\mathcal{A}_c \ni f \mapsto \int_0^T \int \mathcal{L}_t f d\mathscr{L}^d dt$ (provided that $a$, $b$ are locally integrable). We say that $\operatorname{div} \mathcal{L} \in L^1_t(L^p_x)$ if there exists a (necessarily unique) $g \in L^1_t(L^p_x)$ such that

$$\int_{(0,T)\times\mathbb{R}^d} \mathcal{L} f d\mathscr{L}^{1+d} = -\int_{(0,T)\times\mathbb{R}^d} f g d\mathscr{L}^{1+d}, \quad \text{for every } f \in \mathcal{A}_c.$$

Similarly, $\operatorname{div} \mathcal{L}^- \in L^1_t(L^p_x)$ if, for some $g \in L^1_t(L^p_x)$, the inequality $\leq$ in place of equality above holds, for every $f \in \mathcal{A}$, with $f \geq 0$. If $\operatorname{div} \mathcal{L}^- \in L^1_t(L^p_x)$, then we can prove the following inequality, for $u \in D^{1,p}(\mathcal{L}, a\nabla \otimes \nabla)$, and $\beta \in C^3(\mathbb{R})$ convex, with bounded derivatives as well as $\beta'(z)z - \beta(z)$ bounded:

$$\int_{(0,T)\times\mathbb{R}^d} \mathcal{L}(\beta'(u)) u d\mathscr{L}^{1+d} \leq \int_{(0,T)\times\mathbb{R}^d} [\beta'(u)u - \beta(u)] \operatorname{div} \mathcal{L}^- d\mathscr{L}^{1+d}. \tag{3.6}$$

Indeed, let $\rho$ be a smooth convolution kernel on $\mathbb{R}^d$ and consider the diffusion operator $\mathcal{L}^m$ with smooth coefficients $a * \rho^m$ and $b * \rho^m$ (where we let $\rho^m(x) = m^d \rho(mx)$). If we also assume $u \in \mathcal{A}$, then the inequality above holds true by the derivation as in Section 3.1 above. The general case follows by approximation, letting first $m \to \infty$ and then choosing $u^n \in \mathcal{A}$ converging towards $u$ in $u \in D^p(\mathcal{L}, a\nabla \otimes \nabla)$.

Besides these spaces associated with $\mathcal{L}$, let us recall some features of standard Sobolev spaces and the smoothing properties of the standard heat semigroup $(P^\alpha)_{\alpha \geq 0}$ on $\mathbb{R}^d$. For $p \in [1, \infty]$, we consider spaces

$$W^{1,p}_x := \{ f \in L^p_x \, : \, \nabla f \in L^p_x \}, \quad W^{2,p}_x = \left\{ f \in W^{1,p}_x \, : \, \nabla^2 f \in L^p_x \right\},$$

endowed with the usual norms.

A crucial fact for our deductions are quantitative inequalities for the smoothing effect of the heat semigroup $(\mathsf{P}^\alpha)_{\alpha \geq 0}$, which can be deduced by straightforward computations from the heat kernel in $\mathbb{R}^d$. Of course, $\mathsf{P}^\alpha$ is a contraction semigroup in $W^{1,p}_x$ as well as $W^{2,p}_x$; moreover, integration by parts and Hölder inequality give

$$\sqrt{\alpha} \left\| \nabla \mathsf{P}^\alpha f \right\|_{L^p_x} \leq c \left\| f \right\|_{L^p_x} \quad \text{for every } \alpha \in (0, +\infty), \tag{3.7}$$

with $c$ depending on $p \in [1, \infty]$ only (possibly also on the dimension $d$). Such inequalities, called $L^p$-$\Gamma$ inequalities in [6], play a fundamental role for our approach to continuity equations in metric measure spaces: their validity in abstract setups as well as in Riemannian manifolds follow e.g. from uniform lower bounds on the Ricci curvature.

Arguing similarly, it holds for $p \in [1, \infty]$, $i, j \in \{1, \ldots d\}$,

$$\alpha \left\| \partial^2_{i,j} \mathsf{P}^\alpha f \right\|_{L^p_x} \leq c \left\| f \right\|_{L^p_x} \quad \text{for every } \alpha \in (0, \infty). \tag{3.8}$$

Let us also notice that, as $\alpha \downarrow 0$, the left hand side in the two inequalities above are infinitesimal, for a standard density and uniform boundedness argument applies.

Finally, another property that we occasionally use below is that, for $p \in (1, \infty)$, one has $W^{2,p}_x = \{ f \in L^p_x \, : \, \Delta f \in L^p_x \}$, because of the $L^p_x$-boundedness of the second order Riesz transform $f \mapsto \nabla^2 \Delta^{-1} f$, see e.g. [20].

### 3.3 Well-posedness: statement of results

We are in a position to state our main well-posedness results, which we split in two theorems: the first one deals with possibly degenerate diffusions, with Sobolev regular coefficients.

**Theorem 3.1** (degenerate case)**.** *Let $p \in (1, \infty]$, $r \geq 2p/(p-1)$, and $a$, $b$ be as in* (2.1)*, with*

$$a \in L^1_t(W^{2,p}_x), \quad b \in L^1_t(W^{1,p}_x), \quad \textit{and} \quad \operatorname{div} \mathcal{L}^- \in L^1_t L^\infty_x.$$

*Then, for every probability density $\bar{u} \in L^r_x$, there exists a unique narrowly continuous solution $u = (u_t)_{t \in [0,T]}$ of the FPE* (2.2) *with $u_0 = \bar{u}$ and $u \in L^\infty_t(L^r_x)$.*

Actually, the technique employed provides (existence and) uniqueness even without the assumption that $\bar{u}$ is a probability density. As a straightforward consequence of the result above and the equivalence established in the previous section, we have in particular the following

**Corollary 3.2.** *Let $p$, $r$, $a$, $b$ be as in the theorem above, let $\mathcal{R}$ be the class of narrowly continuous solutions $u$ of the FPE* (2.2)*, with $u \in L^\infty_t(L^r_x)$. Then, there exists a unique $\mathcal{R}$-regular martingale flow associated to $\mathcal{L}(a, b)$. Moreover, such flow satisfies the Chapman-Kolmogorov equations* (2.15)*.*

Our second statement deals with non-degenerate (elliptic) diffusions, i.e. if it holds, for some $\lambda > 0$, $a(v, v) \geq \lambda |v|^2$, for every $v \in \mathbb{R}^d$, a.e. in $(0, T) \times \mathbb{R}^d$. In such a case, we can remove one order of Sobolev regularity assumption from both coefficients, but we introduce Lipschitz regularity for $t \mapsto a_t$.

**Theorem 3.3** (bounded elliptic case)**.** *Let $p \in [2, \infty]$, $r \geq 2p/(p-2) \in [2, \infty]$, and $a$, $b$ be as in* (2.1)*, with $a \in L^\infty_t(L^\infty_x)$ and elliptic,*

$$\partial_t a \in L^\infty_{t,x}, \quad a \in L^1_t(W^{1,p}_x), \quad b \in L^1_t L^\infty_x \quad \textit{and} \quad ((\nabla^*)^2 a)^+ \in L^1_t(L^\infty_x).$$

*Then, for every probability density $\bar{u} \in L^r_x$, there exists a unique narrowly continuous solution $u = (u_t)_{t \in [0,T]}$ of the FPE* (2.2) *with $u_0 = \bar{u}$ and $u \in L^\infty_t(L^r_x)$.*

Also in this case, as a consequence of the equivalence between Eulerian and Lagrangian descriptions, we obtain well-posedness for $\mathcal{R}$-regular flows: we omit the formal statement, which reads exactly as Corollary 3.2.

**Remark 3.4** (comparison with existing literature)**.** *The literature on the subject of Fokker-Planck equations and martingale problems is so vast and growing that we limit ourselves to a direct comparison only with very closely related and recent works. In particular, we stress some aspects which are different from the results appearing in [19], [24].*

*In [24], the approach is mostly Eulerian, dealing with FPE's in divergence form*

$$\partial_t u_t + \nabla^*(u_t b) = \frac{1}{2} \nabla^*(\sigma \sigma^* \nabla u_t), \quad \textit{on } (0, T) \times \mathbb{R}^d,$$

*with $\sigma : \mathbb{R}^d \to \mathbb{R}^{d \times k}$. The main result in [24] provides existence and uniqueness for the equation above, provided that*

$$b \in L^1_t(W^{1,1}_{loc}), \quad \frac{b}{1 + |x|} \in L^1_t(L^1_x + L^\infty_x), \quad \nabla^* b \in L^1_t(L^\infty_x)$$

$$\sigma \in L^2_t(W^{1,2}_{loc}), \ \frac{\sigma}{1 + |x|} \in L^2_t(L^2_x + L^\infty_x).$$

*To compare these assumptions, we must notice as in [24, §5.1] that with our notation $a = \sigma \sigma^*$ and the drift is actually $b - \frac{1}{2} \nabla^* a$. In view of this correspondence, it might* *seem that Theorem 3.1 follows from their weaker assumptions: this follows in principle from a result of the type $\sigma := a^{1/2} \in L^2_t(W^{1,2}_{loc})$, if $a \in L^1_t(W^{2,2}_{loc})$, extending the well-known result [31, Lemma 3.2.3] that $a^{1/2}$ is Lipschitz whenever $a \in C^2$. However, their conclusions are in fact weaker, and actually insufficient in order to obtain correspondent Lagrangian results: they prove existence and uniqueness in the class of narrowly continuous probability densities $u \in L^\infty_t(L^\infty_x)$ such that $\sigma \nabla u \in L^2_t(L^2_x)$: the latter (weak) regularity condition then prevents from a straightforward application of the results in Section 2.2. In conclusion, our result has (apparently) stronger regularity conditions on the coefficients, but draws stronger results and leads directly to well-posedness of regular martingale problems and flows.*

*The problem arising from the condition $\sigma\nabla u \in L^2_t(L^2_x)$, which prevents a Lagrangian theory, is well understood in [19], where much effort is put in showing, for the bounded elliptic case, uniqueness in the class of narrowly continuous probability densities $u \in L^2_t(L^2_x)$ [19, Theorem 4.3]. When compared with the assumptions of Theorem 3.3, an evident difference is that we require a first order condition $a \in L^1_t(W^{1,p}_x)$, while no such requirement appear in [19, Theorem 1.3], besides (with our notation) $\nabla^* a$, $\operatorname{div} \mathcal{L}^- \in L^\infty_t(L^\infty_x)$. The technique we employ – approximation by the semigroup associated to the Dirichlet form $f \mapsto \int a(\nabla f, \nabla f)$ – is the same as Figalli's one, and in the elliptic case the novelty is more conceptual, providing a much cleaner derivation of commutator estimates, essentially by the same abstract arguments in the elliptic and the degenerate case. However, in the possibly degenerate case, our results are stronger, compare e.g. with [19, Theorem 1.4], as we allow for much more general diffusion coefficients, and possibly unbounded terms – obtaining as well Lagrangian counterparts.*

*In more recent years, further developments along these research lines appeared in the literature, as well as different techniques (e.g. Crippa-DeLellis' technique [14] was extended to SDE's in [35, 28]): of course, novelties and improvements appear in these developments, but to the author's knowledge that they address different aspects (such as strong solutions, equations with jumps, quantitative estimates, etc.) and there is no substantial overlap with our two results above. On the other side, thanks to the superposition principle (Theorem 2.5), it may be possible to extend such results and prove well-posedness for Fokker-Planck equations with unbounded coefficients starting from correspondent Lagrangian results (we thank the anonymous referee for this comment).*

*We also point out that the theory of measure-valued solutions (i.e., not necessarily absolutely continuous) Fokker-Planck equations, at least in the elliptic case, is well-developed and some results may be compared with ours. For example, [8, Proposition 3.1] entails uniqueness if, for some $p \geq d+2$, $a \in L^\infty_t H^{1,p}_x$ is elliptic, $b \in L^p_t L^p_x$ and $t \mapsto a_t$ is Hölder continuous (locally uniformly in $x$). It is immediate to see that there is no inclusion between such class of coefficients and that of Theorem 3.3, and in particular the hypothesis of our result are dimension-free (indeed, we are specializing a theory tailored for infinite dimensional spaces). However, the uniqueness class is smaller in our case, since we restrict from the very beginning to absolutely continuous solutions, which is nevertheless sufficient to entail a reasonable Lagrangian theory. Let us point out some recent developments [7, 9, 11] and in particular [10] which also contains a survey of known results and methods for the degenerate case. Finally, we point out the monograph in preparation [33], which contains a detailed study and a vast bibliography on the subject.*

Let us briefly discuss some features of the two theorems above and their proof. First, existence of weak solutions in the hypothesis stated above is a much easier task than uniqueness: for example, one can argue by approximation via convolution of the coefficients (and the initial law) with a smooth kernel, so that the estimates on the coefficients are preserved, and one gains enough regularity (e.g. $C^2$ coefficients) so existence is available even at the Lagrangian level. Then, we have enough regularity so that the deductions which lead to inequality (3.1) apply and by a Gronwall argument we deduce a bound in $L^\infty_t(L^r_x)$, in terms of $\operatorname{div} \mathcal{L}^-$ only, and uniform in the approximation (in the elliptic case, we argue with (3.3) instead). By extracting a weakly convergent sequence and by strong convergence of the approximations of the coefficients, we deduce that any weak limit point in $L^\infty_t(L^r_x)$ is a weak solution to the FPE (2.2). In the elliptic case, we deduce as well existence for a solution $u \in L^\infty_t(L^r_x) \cap L^2_t(W^{1,2}_x)$. Let us also recall the approach [19, Theorem 4.3], which is completely Eulerian (i.e., it relies on PDE's techniques only), and has the advantage of yielding easily uniqueness, for solutions belonging to such a (smaller) space, which does not allow for applications of the theory developed in Section 2.2.

In order to establish uniqueness of solutions, our aim is to rigorously establish (3.4) and (3.3), for (any difference of) solutions $u \in L^\infty_t(L^r_x)$. As already remarked, the main problem is related to the regularity of $u$, in order to employ the standard calculus rules. Our strategy relies the well-known smoothing scheme, which dates back at least to [16]: for $\alpha \in (0,1)$ we introduce some linear operator $\mathsf{P}^\alpha$, acting on functions defined on $(0,T) \times \mathbb{R}^d$ such that, by defining $u^\alpha := \mathsf{P}^\alpha u$, we obtain an approximation of $u$ sufficiently regular to rigorously obtain (3.4). Of course, the price that we pay is that $u^\alpha$, in general, is not a solution of (2.2) and one has to carefully estimate the "error terms" thus appearing: our novel contribution indeed provides a systematic approach to such inequalities.

To be more precise, in the cases that we consider, the operators $(\mathsf{P}^\alpha)_{\alpha \geq 0}$ form a strongly-continuous Markov symmetric semigroup on $L^2((0,T) \times \mathbb{R}^d, \mathscr{L}^{d+1})$, so that, in particular, $\mathsf{P}^\alpha$ preserves all $L^p_t L^q_x$ spaces, for $p, q \in [1, \infty]$. If we also prove that $\mathsf{P}^\alpha$ maps $W^{1,1}_t(L^{r'}) \cap D^{r'}(\mathcal{L})$ into itself, we may write, for $f$ belonging to such space,

$$\int_0^T \int [(\partial_t + \mathcal{L}_t) f] \, u^\alpha_t d\mathscr{L}^d dt = \int f^\alpha_T u_T d\mathscr{L}^d - \int f^\alpha_0 u_0 d\mathscr{L}^d + \int_0^T \int u \, [\mathsf{P}^\alpha, (\partial_t + \mathcal{L}_t)] \, f d\mathscr{L}^d dt, \tag{3.9}$$

since the weak formulation (3.5) extends by density of $\mathcal{A}$ in $W^{1,1}_t(L^{r'}) \cap D^{r'}(\mathcal{L})$. The commutator term appears as an algebraic way to highlight the identity as an equation for $u^\alpha$, and all the issue is to show that it is infinitesimal as $\alpha \downarrow 0$.

Next, we prove that $\mathsf{P}^\alpha$ has a "smoothing effect", in a sense that we can choose $\beta'(u^\alpha)$ as a test function, and apply the chain rule with respect to $\partial_t$ and (3.6), so

$$\partial_t \int \beta(u^\alpha_t) d\mathscr{L}^d \leq \int_{\mathbb{R}^d} [\beta'(u^\alpha_t) u^\alpha_t - \beta(u^\alpha_t)] \operatorname{div} \mathcal{L}^-_t d\mathscr{L}^d + \int_{\mathbb{R}^d} u_t \, [\mathsf{P}^\alpha, (\partial_t + \mathcal{L}_t)] \, \beta'(u^\alpha_t) d\mathscr{L}^d,$$

$\mathscr{L}^1$-a.e. $t \in (0,T)$ and in the sense of distributions on $(0,T)$. Finally, we let $\alpha \downarrow 0$, and by strong convergence of $u^\alpha$ towards $u$ in $L^1_t(L^r_x)$, we are able to conclude, provided

$$\int_{\mathbb{R}^d} u_t \, [\mathsf{P}^\alpha, (\partial_t + \mathcal{L}_t)] \, \beta'(u^\alpha_t) d\mathscr{L}^d \leq \varepsilon(\alpha) \to 0, \quad \text{in } L^1(0,T) \text{ as } \alpha \downarrow 0.$$

### 3.4 Commutator inequalities

In this section, we estimate the "error terms" involving the commutator between $\mathsf{P}^\alpha$ and $\partial_t + \mathcal{L}$. Our general strategy is a further development of that first introduced in [6], in the framework of continuity equations in metric measure spaces, and it is completely "coordinate free" and depends on an interpolation argument *à la* Bakry-Émery, namely

$$\begin{aligned}\int u\left[\mathsf{P}^\alpha, \partial_t + \mathcal{L}\right] f d\mathscr{L}^d &= \int \left[\mathsf{P}^\alpha u(\partial_t + \mathcal{L}) f - u(\partial_t+\mathcal{L})\mathsf{P}^\alpha f\right] d\mathscr{L}^d\\ &= \int\int_0^\alpha \frac{d}{ds}\left[(\mathsf{P}^s u)(\partial_t+\mathcal{L})\mathsf{P}^{\alpha-s} f\right] ds d\mathscr{L}^d\\ &= \int_0^\alpha \int \mathsf{P}^s u\left[\Delta, \partial_t+\mathcal{L}\right]\mathsf{P}^{\alpha-s} f d\mathscr{L}^d ds,\end{aligned}$$

where we let $\Delta$ be the generator of $(\mathsf{P}^\alpha)_{\alpha\geq 0}$. It turns out that the commutator between $\Delta$ and $\partial_t + \mathcal{L}$, reflecting the "relative regularity" between the chosen approximation and the target diffusion, depends upon natural quantities such as Sobolev regularity of the coefficients.

In principle, this method provides very general results but, for the ease of exposition, we address separately only three cases, which are of particular interest: the case of a commutator between the Euclidean heat semigroup and a Sobolev derivation, which is a specialization of [6, Lemma 5.8] in the Euclidean case; that of a commutator between the Euclidean heat semigroup and a second-order Sobolev diffusion, which is apparently novel, that we settle by performing a "second order" interpolation argument; and finally that of the commutator between $\partial_t$ and a non-degenerate diffusion acting only the variable $x \in \mathbb{R}^d$, with $t \mapsto a_t$ Lipschitz, which provides an alternative approach to Step 2 in [19, Theorem 4.3].

We let throughout $q \in (1,\infty]$, $r$, $s \in (1,\infty)$, with $q^{-1} + r^{-1} + s^{-1} = 1$ (one can deal with endpoint cases at the price of more delicate approximations).

**Lemma 3.5.** *Let $b \in L^1_t(W^{1,q}_x)$, $u \in L^\infty_t(L^r_x)$ and $f \in L^\infty_x(W^{1,s}_x)$. It holds*

$$\int_0^T \left|\int u_t[\mathsf{P}^\alpha, b_t\cdot\nabla] f_t d\mathscr{L}^d\right| dt \leq c\left\|\nabla b\right\|_{L^1_t L^q_x}\left\|u\right\|_{L^\infty_t L^r_x}\left\|f\right\|_{L^\infty_t L^s_x}, \quad \text{for } \alpha\in(0,1), \tag{3.10}$$

*where $c \in \mathbb{R}$ is some constant (depending on the dimension $d$ only).*

Actually, the proof below shows that $\nabla b$ can be replaced with the symmetric part of the derivative (also called deformation) $D^{sym} b := (\nabla b + (\nabla b)^\tau)/2$, where $\tau$ denotes the transpose operator.

As a consequence of (3.10), the commutator operator $L^\infty_x(W^{1,s}_x) \ni f \mapsto [\mathsf{P}^\alpha, b\cdot\nabla] f \in L^1_t L^{r'}_x$ extends to a linear continuous operator on $L^\infty_t L^s_x$. Moreover, a standard density and uniform boundedness argument entails that, for $f \in L^\infty_t L^s_x$,

$$[\mathsf{P}^\alpha, b\cdot\nabla] f \to 0, \quad \text{strongly in } L^1_t(L^{r'}_x) \text{ as } \alpha \downarrow 0.$$

*Proof.* It is sufficient to argue assuming that $b$, $u$ and $f$ are sufficiently smooth, e.g., $u \in \mathcal{A}_c$, $f \in \mathcal{A}$, as well as $b^i \in \mathcal{A}$, for $i \in \{1,\ldots,d\}$, as the general inequality will follow by approximation (e.g. by convolution with a smooth kernel). Moreover, we argue at $t \in (0,T)$ fixed and then integrate over the interval $(0,T)$: thus we omit to specify $t \in (0,T)$ in what follows.

The curve $s \mapsto F(s) = \int u^s b\cdot\nabla f^{\alpha-s} d\mathscr{L}^d$ is then $C^1_b(0,\alpha)$, with

$$\frac{d}{ds}F(s) = \int (\Delta u^s) b\cdot\nabla f^{\alpha-s} - u^s b\cdot\nabla(\Delta f^{\alpha-s}) d\mathscr{L}^d,$$

By straightforward integration by parts, we obtain the following alternative expression for the right hand side above:

$$\frac{d}{ds}F(s) = \int\left[\left((\nabla b + (\nabla b)^\tau)\nabla u^s, \nabla f^{\alpha-s}\right) + (\nabla^* b)\left(\nabla u^s\cdot\nabla f^{\alpha-s} - u^s\Delta f^{\alpha-s}\right)\right] d\mathscr{L}^d.$$

If $\nabla^* b = 0$, the conclusion is immediate, since we may estimate $|F(\alpha) - F(0)| \leq \int_0^\alpha \left|\frac{d}{ds}F(s)\right| ds$ and, by Hölder inequality,

$$\left|\int u[\mathsf{P}^\alpha, b\cdot\nabla] f d\mathscr{L}^d\right| \leq 2\int_0^\alpha \left\|D^{sym}b\right\|_{L^q_x} \left\|\nabla u^s\right\|_{L^r_x} \left\|\nabla f^{\alpha-s}\right\|_{L^s_x} ds$$
$$\leq \left\|D^{sym}b\right\|_{L^q_x} \left\|u\right\|_{L^r_x} \left\|f\right\|_{L^s_x} \int_0^\alpha \frac{2ds}{\sqrt{s(\alpha-s)}}$$
$$\leq 2\pi \left\|D^{sym}b\right\|_{L^q_x} \left\|u\right\|_{L^r_x} \left\|f\right\|_{L^s_x}.$$

by (3.7) and using $\int_0^1 (s(1-s))^{-1/2} ds = \pi$.

The general case $\nabla^* b \in L^q_x$ is slightly more involved: let us first notice that the term $(\nabla^* b)\nabla u^s \cdot \nabla f^{\alpha-s}$ can be estimated as above, adding a contribution $\pi \left\|\nabla^* b\right\|_{L^q_x}$ to the inequality. Finally, to estimate the contribution of $(\nabla^* b) u^s \Delta f^{\alpha-s}$ we do not put the absolute value inside integration with respect to $s \in (0,\alpha)$, but exchange integration with respect to $x$ and $s$, exploiting the identity

$$\int_0^\alpha (\nabla^* b) u^s \Delta f^{\alpha-s} ds = -(\nabla^* b)\int_0^\alpha u^s \frac{d}{ds} f^{\alpha-s} ds.$$

Next, to integrate by parts only "half of the derivative" with respect to $s$, we simply add $(\nabla^* b)u^\alpha$ times the quantity

$$f^0 - f^\alpha - \int_0^\alpha \frac{d}{ds} f^{\alpha-s} = 0,$$

thus

$$\left|\int_0^\alpha u^s \frac{d}{ds} f^{\alpha-s} ds\right| \leq \left|u^\alpha \left(f^0 - f^\alpha\right)\right| + \int_0^\alpha \left|(u^s - u^\alpha)\Delta f^{\alpha-s}\right| ds,$$

which, once integrated with respect to $x \in \mathbb{R}^d$, by Hölder inequality and (3.8) is bounded from above by

$$\left\|\nabla^* b\right\|_{L^q_x} \left\|u\right\|_{L^r_x} \left\|f\right\|_{L^s_x} \left(2 + c\int_0^\alpha \frac{ds}{\sqrt{s(s-\alpha)}}\right).$$

This settles an analogue of (3.10) at fixed $t \in (0,T)$, and by integration with respect to $t \in (0,T)$, we obtain (3.10). □

The constant $c$ can be even independent of the dimension $d$ of the underlying space, provided that assume some bound directly on $\left\|\nabla^* b\right\|_{L^q_x}$, and use a refined, dimension independent estimate for $\left\|\Delta f^{\alpha-s}\right\|_{L^s_x}$: these are the key observation that lead to well-posedness on possibly infinite dimensional spaces, as developed in [6].

**Lemma 3.6.** *Let $a \in L^1_t(W^{2,q}_x)$, $u \in L^\infty_t(L^r_x)$ and $f \in L^\infty_x(W^{2,s}_x)$. For $\alpha \in (0,1)$, it holds*

$$\int_0^T \left|\int u_t[\mathsf{P}^\alpha, a_t : \nabla^2] f_t d\mathscr{L}^d - \alpha\int u_t[\Delta, a_t : \nabla^2]\mathsf{P}^\alpha f_t d\mathscr{L}^d\right| dt \leq$$
$$\leq \left\|\nabla^2 a\right\|_{L^1_t L^q_x} \left\|u\right\|_{L^\infty_t L^r_x} \left\|f\right\|_{L^\infty_x L^s_t} \tag{3.11}$$

*where $c$ is some constant (depending on $d$ only). Moreover, for $u \in L^\infty_t(L^r_x \cap L^s_x)$, it holds*

$$\left|\int u[\mathsf{P}^\alpha, a : \nabla^2](\mathsf{P}^\alpha u) d\mathscr{L}^d\right| \to 0, \quad \text{in } L^1(0,T), \text{ as } \alpha \downarrow 0. \tag{3.12}$$

*Proof.* To establish (3.11), the underlying idea is to formally rewrite $a : \nabla^2 f = a : (\nabla^2\Delta^{-1})\Delta f$ and exploit the boundedness of the Riesz transform $\nabla^2\Delta^{-1}$ in $L^s_x$, together with a second order interpolation along the heat semigroup. To make computations more transparent, we argue on coordinates, i.e., we fix $i$, $j \in \{1, \ldots, d\}$ and consider the commutator

$$[\mathsf{P}^\alpha, a^{i,j}\partial^2_{i,j}]f = \mathsf{P}^\alpha(a^{i,j}\partial^2_{i,j} f) - a^{i,j}\partial^2_{i,j}(\mathsf{P}^\alpha f).$$

As in the proof of the previous lemma, we may also let $u \in \mathcal{A}_c$, $f$ and $a^{i,j}$ be sufficiently regular, e.g. $f$, $a^{i,j} \in C^4_b((0,T)\times\mathbb{R}^d)$, and argue at fixed $t \in (0,T)$. We consider the curve

$$[0,\alpha] \ni s \mapsto F(s) := \int u^s a^{i,j}\partial^2_{i,j} f^{\alpha-s} d\mathscr{L}^d,$$

which is $C^1_b(0,\alpha)$, with

$$F'(s) = \int u^s [\Delta, a^{i,j}\partial^2_{i,j}] f^{\alpha-s} d\mathscr{L}^d = \int u^s [\Delta, a^{i,j}]\partial^2_{i,j} f^{\alpha-s} d\mathscr{L}^d,$$

since the Laplacian and partial derivatives commute. We write $h^{\alpha-s} := \partial^2_{i,j} f^{\alpha-s} = (\partial^2_{i,j} f)^{\alpha-s}$ (since derivatives and heat semigroup commute), let $b := \nabla a^{i,j}$ and integrate by parts, obtaining

$$F'(s) = 2\int u^s b\cdot\nabla h^{\alpha-s} d\mathscr{L}^d + \int u^s (\Delta a^{i,j}) h^{\alpha-s} d\mathscr{L}^d.$$

Differentiating once more, since $F \in C^2_b(0,\alpha)$, we obtain

$$F''(s) = 2\int u^s [\Delta, b\cdot\nabla] h^{\alpha-s} d\mathscr{L}^d + \int u^s [\Delta, (\Delta a^{i,j})] h^{\alpha-s} d\mathscr{L}^d.$$

We introduce a second order interpolation based on the Taylor expansion

$$F(\alpha) - F(0) - \alpha F'(0) = \int_0^\alpha F''(\sigma)(\alpha-\sigma) d\sigma,$$

and we notice that the left hand side gives, up to integration on $(0,T)$, the left hand side of (3.11).

Let us notice first how we would conclude in case $\nabla^* b = \Delta a^{i,j} = 0$, and then address the general case. As in the previous lemma, we obtain the identity

$$\int u^s [\Delta, b\cdot\nabla] h^{\alpha-s} d\mathscr{L}^d = -2\int \left( (\nabla^2 a^{i,j})\nabla u^s, \nabla h^{\alpha-s} \right) d\mathscr{L}^d$$

and we estimate

$$\begin{aligned} |F''(s)| &\le 4\left\|\nabla^2 a^{i,j}\right\|_{L^q_x} \|\nabla u^s\|_{L^r_x} \|\nabla h^{\alpha-s}\|_{L^s_x} \\ &\le \frac{c}{\sqrt{s(\alpha-s)^3}} \left\|\nabla^2 a^{i,j}\right\|_{L^q_x} \|u\|_{L^r_x} \|f\|_{L^s_x}, \end{aligned}$$

where $c$ is some constant. Integrating with respect to $s \in (0,\alpha)$ and exploiting the factor $(\alpha-\sigma)$ to compensate the bound the norm of $h^{\alpha-s}$, we deduce (3.11).

To address the general case, we bound separately the terms

$$\int_0^\alpha \int u^s [\Delta, b\cdot\nabla] h^{\alpha-s} d\mathscr{L}^d\, (\alpha-s) ds \quad \text{and} \quad \int_0^\alpha \int u^s [\Delta, (\Delta a^{i,j})] h^{\alpha-s} d\mathscr{L}^d\, (\alpha-s) ds. \tag{3.13}$$

To deal with former one, we isolate a "leading term" which involves $\nabla^2 a^{i,j}$ and we bound the remaining terms it by adding and subtracting suitable quantities, with the only difficulty that we must take into account the second order expansion. Precisely, after arguing as in the case $\Delta a^{i,j} = 0$, we are left with estimating

$$\int_0^\alpha \int u^s (\Delta a^{i,j}) \Delta h^{\alpha-s} (\alpha-s) d\mathscr{L}^d ds, \tag{3.14}$$

and to this aim we add and subtract

$$\int_0^\alpha \int u^\alpha (\Delta a^{i,j}) \Delta h^{\alpha-s}(\alpha-s) d\mathscr{L}^d ds = \int_0^\alpha \int u^\alpha (\Delta a^{i,j}) R_{i,j} \Delta^2 f^{\alpha-s}(\alpha-s) d\mathscr{L}^d ds, \quad (3.15)$$

where we let $R_{i,j} f := \partial^2_{i,j} \Delta^{-1} f$ be the second-order Riesz transform along the directions $i$, $j$. The difference between the (3.14) and (3.15) is easily bounded and, to conclude, we exploit the identity

$$\int_0^\alpha \Delta^2 f^{\alpha-s}(\alpha-s) ds = \int_0^\alpha (\alpha-s) \partial^2_s f^{\alpha-s} ds = -f^\alpha + f + \alpha \Delta f^\alpha.$$

and use the fact that the latter quantity is uniformly bounded (as well as the fact that $R_{i,j}$ is a bounded operator).

To estimate the second expression in (3.13), we notice that

$$\int u^s [\Delta, (\Delta a^{i,j})] h^{\alpha-s} d\mathscr{L}^d = \frac{d}{ds} \int u^s (\Delta a^{i,j}) \partial^2_{i,j} f^{\alpha-s} d\mathscr{L}^d,$$

thus we integrate by parts with respect to $s \in (0,\alpha)$,

$$\int_0^\alpha \frac{d}{ds} u^s (\Delta a^{i,j}) \partial^2_{i,j} f^{\alpha-s} (\alpha-s) ds = -\alpha u (\Delta a^{i,j}) \partial^2_{i,j} f^\alpha + \int_0^\alpha u^s (\Delta a^{i,j}) \partial^2_{i,j} f^{\alpha-s}.$$

The first term in the right hand side above is bounded by $c \left\| \Delta a^{i,j} \right\|_{L^q_x} \|u\|_{L^r_x} \|f\|_{L^s_x}$. We write

$$\int_0^\alpha \int u^s (\Delta a^{i,j}) \partial^2_{i,j} f^{\alpha-s} d\mathscr{L}^d ds = \int_0^\alpha \int u^s (\Delta a^{i,j}) R_{i,j} \Delta f^{\alpha-s} d\mathscr{L}^d ds.$$

To conclude, we argue once more by adding and subtracting

$$\int_0^\alpha \int u^\alpha (\Delta a^{i,j}) R_{i,j} \Delta f^{\alpha-\sigma} d\mathscr{L}^d ds = \int u^\alpha (\Delta a^{i,j}) R_{i,j} (f^\alpha - f) d\mathscr{L}^d,$$

and estimating the differences involved. This settles the validity of (3.11), for smooth functions and at fixed $t \in (0,T)$. By integration and a density argument, the general case is deduced at once.

Next, we prove (3.12), which follows from the fact that $\alpha \int u[\Delta, a : \nabla^2] u^{2\alpha} d\mathscr{L}^d$ is infinitesimal, as $\alpha \downarrow 0$: indeed, a standard uniform boundedness and density argument gives that the left hand side in (3.11) is infinitesimal as $\alpha \downarrow 0$. To show it, we initially argue in the case of smooth functions $u$, $f$, and for fixed $i$, $j \in \{1, \ldots, d\}$, we let $b = \nabla a^{i,j}$ and integrate by parts

$$\begin{aligned}
\int u[\Delta, a^{i,j} \partial^2_{i,j}] f^\alpha d\mathscr{L}^d &= -2 \int (b \cdot \nabla u) \partial^2_{i,j} f^\alpha d\mathscr{L}^d - \int u (\Delta a^{i,j}) \partial^2_{i,j} f^\alpha d\mathscr{L}^d \\
&= -2 \int (\partial^2_{i,j} f) \mathsf{P}^\alpha (b \cdot \nabla u) d\mathscr{L}^d - \int u (\Delta a^{i,j}) \partial^2_{i,j} f^\alpha d\mathscr{L}^d \\
&= -2 \int (\partial^2_{i,j} f) \left\{ [\mathsf{P}^\alpha, b \cdot \nabla] u + (b \cdot \nabla u^\alpha) \right\} d\mathscr{L}^d - \int u (\Delta a^{i,j}) \Delta f^\alpha d\mathscr{L}^d.
\end{aligned}$$

Although the intermediate steps require some regularity for $u$, by the commutator estimate for Sobolev derivations established in the previous lemma, the resulting identity extends by continuity to $u \in L^\infty_t(L^r_x)$, $f \in L^\infty_t(W^{2,s}_x)$. Next, we specialize to the case $f := u^\alpha$. By the strong convergence provided by Lemma 3.5 and uniform boundedness of $\alpha \partial^2_{i,j} u^\alpha$ in $L^\infty_1(L^r_x)$, we have

$$\alpha \left| \int (\partial^2_{i,j} u^\alpha_t) [\mathsf{P}^\alpha, b_t \nabla] u_t d\mathscr{L}^d \right| \to 0, \quad \text{in } L^1(0,T) \text{ as } \alpha \downarrow 0.$$

Similarly, it holds (recall that the left hand side in (3.8) is infinitesimal)

$$\alpha\left|\int u(\Delta a^{i,j})\Delta f^\alpha d\mathscr{L}^d\right| \leq \left\|\Delta a^{i,j}\right\|_{L^q_x}\|u\|_{L^r_x}\|\alpha\partial^2_{i,j}u^{2\alpha}\|_{L^s_x}\to 0.$$

Finally, in order to handle the term $\alpha\int(\partial^2_{i,j}u^\alpha)(b\nabla u^\alpha)d\mathscr{L}^d$, the choice $f=u^\alpha$ and the symmetry of $a$ are crucial: we integrate by parts once, and since $b=\nabla a^{i,j}$, we obtain

$$\int(\partial^2_{i,j}u^\alpha)b\cdot\nabla u^\alpha d\mathscr{L}^d = -\sum_{k=1}^d\int\left[\partial_i u^\alpha(\partial^2_{j,k}a^{i,j})\partial_k u^\alpha + (\partial_i u^\alpha)(\partial_k a^{i,j})\partial^2_{k,j}u^\alpha\right]d\mathscr{L}^d.$$

The first term, when multiplied by $\alpha$, is clearly bounded and infinitesimal as $\alpha\downarrow 0$, so we focus on the last one. To show that it is bounded, we recall that $a$ is symmetric and we are actually interested in bounds for the whole sum on $i,\, j\in\{1,\dots d\}$; thus, by coupling the symmetric terms, it is sufficient to prove that

$$\alpha\int\partial_i u^\alpha(\partial_k a^{i,j})\partial^2_{k,j}u^\alpha+\partial_j u^\alpha(\partial_k a^{i,j})\partial^2_{k,i}u^\alpha d\mathscr{L}^d$$

is infinitesimal. This symmetric expression can be explicitly rewritten as

$$\frac{\alpha}{2}\int(\partial_k a^{i,j})\partial_k\left[(\partial_i u^\alpha+\partial_j u^\alpha)^2-(\partial_i u^\alpha)^2-(\partial_j u^\alpha)^2\right]d\mathscr{L}^d,$$

and at this stage we integrate by parts once more, obtaining a bound in terms of

$$\alpha\left\|\nabla^2 a\right\|_{L^q_x}\|\nabla u^\alpha\|_{L^r}\|\nabla u^\alpha\|_{L^s},$$

which is sufficient to conclude (as the left hand side in (3.7) is infinitesimal). $\square$

Finally, we deal with the bounded elliptic case: if $a$ is bounded and elliptic, then the form $L^2_t(W^{1,2}_x)\ni f\mapsto\int a(\nabla f,\nabla f)$ is a Dirichlet form, with associated Markov semigroup $\mathsf{P}^\alpha_a$ and (self-adjoint) generator $\Delta_a f=\operatorname{div}(a\nabla f)$, on its “abstract” domain $D(\Delta_a)$ (as given by the general theory of Dirichlet forms). When we choose $\mathsf{P}^\alpha_a$ as “smoothing operator”, the main difficulty is to prove that it preserves regularity with respect to $t\in(0,T)$, thus we need some estimate for the commutator $[\mathsf{P}^\alpha_a,\partial_t]$, which we initially define in following the weak sense, for $u\in\mathcal{A}_c$, $f\in\mathcal{A}$:

$$\int_{(0,T)\times\mathbb{R}^d}u[\mathsf{P}^\alpha_a,\partial_t]fd\mathscr{L}^{1+d}:=\int_{(0,T)\times\mathbb{R}^d}\left[(\mathsf{P}^\alpha_a u)\partial_t f+(\partial_t u)\mathsf{P}^\alpha_a f\right]d\mathscr{L}^{1+d}.$$

**Lemma 3.7.** *Let $a$ be bounded and elliptic, with $\partial_t a\in L^\infty_t(L^\infty_x)$. Then, for every $\alpha\in(0,1)$, $u\in\mathcal{A}_c$, $f\in\mathcal{A}$, it holds*

$$\left|\int u[\mathsf{P}^\alpha_a,\partial_t]fd\mathscr{L}^{1+d}\right|\leq c\|\partial_t a\|_{L^\infty_t L^\infty_x}\|u\|_{L^2_t L^2_x}\|f\|_{L^2_t L^2_x}, \tag{3.16}$$

*where $c$ is some constant (depending only on the ellipticity constant $\lambda$).*

Thanks to this lemma and a density argument, for $f\in W^{1,2}_t(L^2_x)$, we deduce that $\mathsf{P}^\alpha_a f\in W^{1,2}_t(L^2_x)$, and the “strong” commutator $[\mathsf{P}^\alpha_a,\partial_t]f:=\mathsf{P}^\alpha_a\partial_t f-\partial_t\mathsf{P}^\alpha_a f$ is well defined and it belongs to $L^2_t(L^2_x)$. Moreover, the usual uniform boundedness arguments shows that, for $u\in L^2_t(L^2_x)$ and any family $(f_\alpha)_{\alpha\geq 0}\subseteq L^2_t(W^{1,2}_x)$ converging in $L^2_{t,x}$, it holds

$$\int_{(0,T)\times\mathbb{R}^d}u[\mathsf{P}^\alpha_a,\partial_t]f_\alpha d\mathscr{L}^{1+d}\to 0,\quad\text{as }\alpha\downarrow 0.$$

*Proof.* We provide the following analogue of (3.16), where $\partial_t$ is replaced by $\sigma^{-1}(\mathsf{T}^\sigma - \mathsf{I})$, where $\mathsf{T}^\sigma f(t,x) = f(t+\sigma, x)$, and $\mathsf{I}$ is the identity operator (we also choose $\sigma \neq 0$ small enough, to avoid boundary terms, thanks to the assumption $u \in \mathcal{A}_c$):

$$\left| \int_{(0,T)\times\mathbb{R}^d} u[\mathsf{P}^\alpha_a, \sigma^{-1}(\mathsf{T}^\sigma - \mathsf{I})] f d\mathscr{L}^{1+d} \right| \le c \left\|\partial_t a\right\|_{L^\infty_t L^\infty_x} \left\|u\right\|_{L^2_t L^2_x} \left\|f\right\|_{L^2_t L^2_x},$$

(which $c$ depending on $\lambda$ only). Once this is is settled, we may let $\sigma \to 0$ and pass to the limit in the weak formulation. Let us notice that the identity operator plays no role above, and everything reduces to estimate $\sigma^{-1}\int u[\mathsf{P}^\alpha_a, \mathsf{T}^\sigma] f d\mathscr{L}^{1+d}$. By first-order interpolation along the semigroup $\mathsf{P}^s_a$, for $s \in (0,\alpha)$, it is sufficient to bound the infinitesimal commutator

$$\int u^s [\Delta_a, \mathsf{T}^\sigma] f^{\alpha-s} d\mathscr{L}^{1+d} = \int \left( (\mathsf{T}^\sigma a) \nabla u^s, \nabla \mathsf{T}^\sigma f^{\alpha-s} ) \right) - \left( a \nabla u^s, \nabla \mathsf{T}^\sigma f^{\alpha-s} \right) d\mathscr{L}^{1+d},$$

where we performed integration by parts with respect to the variable $x \in \mathbb{R}^d$ and the change of variables $t \mapsto t+\sigma$ in the first integral. We have therefore the bound

$$\left| \int u^s [\Delta_a, \mathsf{T}^\sigma] f^{\alpha-s} d\mathscr{L}^{1+d} \right| \le \int_0^\sigma \left| \partial_r \int (\mathsf{T}^r a)(\nabla u^s, \nabla \mathsf{T}^\sigma f^{\alpha-s}) d\mathscr{L}^{1+d} \right| dr$$

which by $\partial_r \mathsf{T}^r a = \mathsf{T}^r \partial_t a$ gives the thesis, after an application of Hölder inequality and using the smoothing effect in $L^2_t(L^2_x)$ of $\mathsf{P}_a$, i.e. $\left\|\nabla u^s\right\|_{L^2_t L^2_x} \le (s\lambda)^{-1/2} \left\|u\right\|_{L^2_t L^2_x}$ (a property which holds in general for semigroups associated to Dirichlet forms, see [13][Proposition 1.4.1]). □

It would be natural to extend the argument above for more general exponents beyond the case above; the main issue being that a smoothing effect for $\mathsf{P}_a$ akin to (3.7) is not ensured by Dirichlet form theory, when the exponent involved is different from 2. It seems plausible however to replace $L^2_t(L^2_x)$ with $L^\infty_t(L^2_x)$ and require only $\partial_t a \in L^1_t(L^\infty_x)$ (as the semigroup acts only fiberwise).

**Remark 3.8** (trace semigroup at $t=0$)**.** *Another consequence of Sobolev regularity of the lemma above is existence of a "trace" semigroup, e.g. at $t = 0$, defined as follows: for $f \in L^2_x$, consider a constant extension $f(t,x) = f(x)$ for $(t,x) \in (0,T)\times\mathbb{R}^d$, and let $\mathsf{P}^\alpha_0 f$ be the trace of the Sobolev function $\mathsf{P}^\alpha_a f$ at $t=0$. Alternatively, this can be obtained as the semigroup generated by the bilinear form given by the trace at 0 of $a$.*

### 3.5 Proof of well-posedness results

In this section, we address the proof of Theorem 3.1 and Theorem 3.3. As already remarked, existence is easily settled by approximations, so we focus on uniqueness.

*Proof of Theorem 3.1.* Let $u$ be the difference between any two narrowly continuous solutions in $L^\infty_t(L^r_x)$ and let $\mathsf{P}^\alpha$ be the heat semigroup on $\mathbb{R}^d$, extended on $(0,T)\times\mathbb{R}^d$ by acting on each fiber $\{t\}\times\mathbb{R}^d$, $t \in [0,T]$. For $\alpha > 0$, $\mathsf{P}^\alpha$ maps $W^{1,1}_t(L^{r'}_x) \cap D^{r'}(\mathcal{L})$ into itself, as $f^\alpha := \mathsf{P}^\alpha f$ is $C^2_b$ with respect to the variable $x \in \mathbb{R}^d$, for $\mathscr{L}^1$-a.e. $t \in (0,T)$ (to approximate $\mathsf{P}^\alpha f$ with functions in $\mathcal{A}$, we argue by convolution with a smooth kernel with respect to $t \in (0,T)$), thus (3.9) holds true for $f$ in such a space:

$$\int_0^T \int \left[(\partial_t + \mathcal{L}_t) f\right] u^\alpha_t d\mathscr{L}^d dt = \int f^\alpha_T u_T d\mathscr{L}^d - \int f^\alpha_0 u_0 d\mathscr{L}^d + \int_0^T \int u \left[\mathsf{P}^\alpha, (\partial_t + \mathcal{L}_t)\right] f d\mathscr{L}^d dt. \tag{3.17}$$

For $\alpha > 0$, we also have $u^\alpha \in D^{r'}(\mathcal{L})$: to use $u^\alpha$ as a test function, we deduce that $u^\alpha \in W^{1,1}_t(L^{r'}_x)$, which follows directly from the equation satisfied by $u^\alpha$. Indeed, (3.17) for $f \in \mathcal{A}_c$ entails that the distributional derivative $\partial_t u^\alpha$ coincides with the distribution $\mathcal{L}^* u^\alpha$, which is represented by a function, namely

$$\mathcal{L}^* u^\alpha = \nabla^*(b u^\alpha) + \frac{1}{2}(\nabla^*)^2(a u^\alpha) = -(\operatorname{div} \mathcal{L}) u^\alpha + \mathcal{L} u^\alpha + \frac{1}{2}(\nabla^* a)\cdot(\nabla u^\alpha) \in L^1_t L^{r'}_x.$$

Therefore, $\partial_t u^\alpha \in L^1_t(L^{r'}_x)$, and $u^\alpha$ admits an absolutely continuous continuous representative, which must coincide with the one that we would obtain by acting directly to the narrowly continuous representative $u_t$ with the heat semigroup $\mathsf{P}^\alpha$, at every $t \in [0,T]$: it holds in particular $u^\alpha_0 = 0$, since $u_0 = 0$. Moreover, the curve $t \mapsto \int_{\mathbb{R}^d} (u^\alpha_t)^2 d\mathscr{L}^d$ is absolutely continuous, with distributional and $\mathscr{L}^1$-a.e. derivative $\frac{d}{dt}\int (u^\alpha)^2_t d\mathscr{L}^d = 2\int (\partial_t u^\alpha) u^\alpha d\mathscr{L}^d$.

We are in a position to let $u^\alpha$ in the weak formulation (3.17), to obtain

$$\int_0^T \int \left[(\partial_t + \mathcal{L}_t) u^\alpha\right] u^\alpha_t d\mathscr{L}^d dt = \int (u^\alpha_T)^2 d\mathscr{L}^d + \int_{(0,T)\times\mathbb{R}^d} u \left[\mathsf{P}^\alpha, \mathcal{L}_t\right] u^\alpha d\mathscr{L}^{1+d}.$$

If we choose instead a test function $t \mapsto f(t) u^\alpha_t$, with $f \in C^1_c[0,T)$ and we apply (3.6), we eventually deduce the inequality

$$\frac{d}{dt}\int (u^\alpha)^2_t d\mathscr{L}^d \le \left\|\operatorname{div} \mathcal{L}^-_t\right\|_{L^\infty_x} \int_{\mathbb{R}^d} (u^\alpha_t)^2 d\mathscr{L}^d + \int_{\mathbb{R}^d} u_t \left[\mathsf{P}^\alpha, (\partial_t + \mathcal{L}_t)\right] u^\alpha_t d\mathscr{L}^d,$$

$\mathscr{L}^1$-a.e. $t \in (0,T)$ and in the sense of distributions on $(0,T)$. Gronwall lemma gives

$$\|u^\alpha\|^2_{L^\infty_t L^2_x} \le \exp\left\{\left\|\operatorname{div} \mathcal{L}^-\right\|_{L^1_t L^\infty_x}\right\} \int_0^T \left|\int_{\mathbb{R}^d} u_t \left[\mathsf{P}^\alpha, (\partial_t + \mathcal{L}_t)\right] u^\alpha_t d\mathscr{L}^d\right| dt.$$

As a consequence of Lemma 3.6, we deduce $\|u\|_{L^\infty_t L^2_x} \le \liminf_{\alpha \downarrow 0} \|u^\alpha\|_{L^\infty_t L^2_x} = 0$. □

*Proof of Theorem 3.3.* In our smoothing scheme, we choose $\mathsf{P}^\alpha = \mathsf{P}^\alpha_a$ be the semigroup associated to the Dirichlet form $f \mapsto \int a(\nabla f, \nabla f) d\mathscr{L}^{1+d}$, as introduced in the previous section. A first step consists in showing that (3.17) holds true, and we see it as a consequence of the fact that $\mathsf{P}^\alpha_a$ maps $W^{1,2}_t(L^2_x) \cap D^2(\mathcal{L})$ into itself: if $f \in W^{1,2}_t(L^2_x)$, then Lemma 3.7 shows that $f^\alpha \in W^{1,2}_t(L^2_x)$ as well; to show $f^\alpha \in D^2(\mathcal{L})$, we rely on the assumption on $a \in L^1_t(W^{1,p}_x)$, and show that the smooth approximations obtained by means of the standard heat semigroup $\mathsf{P}^s(f^\alpha)$ converge towards $f^\alpha$ in $D^2(\mathcal{L})$, i.e. $\mathcal{L}\mathsf{P}^s(f^\alpha) \to \mathcal{L}(f^\alpha)$ in $L^1_t(L^2_x)$ (this is the only point where we use the first order regularity assumption on $a$). Such convergence can be seen by the commutator lemma for Sobolev vector fields, Lemma 3.5, noticing that the claim of convergence amounts to show

$$[\mathcal{L}, \mathsf{P}^s] f^\alpha \to 0 \text{ in } L^1_t(L^2_x),$$

but since derivatives and the standard heat semigroup commute, it holds

$$[\mathcal{L}, \mathsf{P}^s] f^\alpha = \sum_{i,j=1}^d [a_{i,j}\partial_i, \mathsf{P}^s]\partial_j f^\alpha + \sum_{i=1}^d [b_i, P^s]\partial_i f^\alpha \to 0$$

since $\partial_j f^\alpha \in L^\infty_t L^2_x$ and Lemma 3.5 shows convergence towards $0$ in $L^1_t L^2_x$, as $s \downarrow 0$.

As a second step, we notice that we may let $u^\alpha$ be a test function in (3.17). Indeed, it holds $u^\alpha \in H^{1,2}(\mathcal{L}, a(\nabla \otimes \nabla))$ by what we just proved, while the fact that $\partial_t u^\alpha$ is represented by some function in $L^1_t L^2_x$ follows from a duality argument: for a.e. $t \in (0,T)$ the linear functional $f \mapsto \int_{\mathbb{R}^d} u_t \mathcal{L}_t f^\alpha$ is bounded in $L^2_x$. From (3.17), we have

$$\partial_t \int (u^\alpha)^2_t d\mathscr{L}^d + 2\lambda \int |\nabla u^\alpha|^2 d\mathscr{L}^d \le \int \left[(u^\alpha_t)^2 (\nabla^*)^2 a_t + u_t b_t \nabla u^\alpha_t + u_t \left[\mathsf{P}^\alpha_a, \partial_t\right] u^\alpha_t\right] d\mathscr{L}^d,$$

where we applied (3.6) only for the diffusion part $a : \nabla^2$, as we deal with the drift term separately, using the inequality

$$|u_t b_t \cdot \nabla u^\alpha| \le \lambda \left|\nabla u^\alpha\right|^2 + 4\lambda^{-1} \left|u_t\right|^2 \left|b_t\right|^2,$$

to bound the contribution of the drift part. To conclude, we apply Gronwall inequality and finally let $\alpha \downarrow 0$, using (3.12) to deduce that the commutator term gives no contribution in the limit and uniqueness holds. □

## A The superposition principle for multidimensional diffusions

To prove Theorem 2.5, we follow a general scheme, whose structure is shared by many proofs of superposition principles appearing in the literature, see e.g. [5, Theorem 8.2.1], [2, Theorem 12], [4, Theorem 4.5], [19, Theorem 2.6], [6, Theorem 7.1], that we summarize below. The derivation is rather elementary, although the "right" underlying framework would that of Young (or random) measures. For simplicity, we let $T = 1$ in this section. Let $\nu = (\nu_t)_{t\in[0,1]} \subseteq \mathscr{P}(\mathbb{R}^d)$ be a narrowly continuous weak solution of the FPE (2.2). To deduce existence of a superposition solution for $\nu$, we perform the following steps.

**Step 1** (approximation)**.** We build from $\nu$ a sequence of solutions $(\nu^n)_n$ of FPE's associated to diffusion operators $(\mathcal{L}^n)_n$, for which the superposition principle is already known to hold, thus obtaining a sequence of superposition solutions $(\boldsymbol{\eta}^n)_n$ of MP's. Here, the difficulty is to exhibit a sufficiently good approximation, so that $\nu^n$ converge towards $\nu$, e.g., narrowly, and $\mathcal{L}^n$ towards $\mathcal{L}$, in a sense to be made precise, as $n \to \infty$.

**Step 2** (tightness)**.** We prove that $(\boldsymbol{\eta}^n)_n \subseteq \mathscr{P}(C([0,1];\mathbb{R}^d))$ is tight, yielding a narrow limit point $\boldsymbol{\eta}$. By Ascoli-Arzelà criterion, this step reduces to show uniform bounds on the modulus of continuity of the canonical process $(e_t)_{t\in[0,1]}$ with respect to $\boldsymbol{\eta}^n$.

**Step 3** (limit)**.** From convergence $\nu^n \to \nu$, $\mathcal{L}^n \to \mathcal{L}$, as $n \to \infty$, we conclude that $\boldsymbol{\eta}$ is a superposition solution for $\nu$. Here, the problem is to deal with convergence for possibly non-continuous functions, as they involve the coefficients $a$, $b$.

### A.1 Approximation

We approximate the limit solution by means of mollification by convolutions or push-forwards via smooth maps (in probabilistic jargon, by conditioning with respect to some observables).

**Push forward via smooth maps.** This technique is inspired by the approach in [6, Theorem 7.1]. Let $\pi \in C^2(\mathbb{R}^d;\mathbb{R}^d)$ $\pi = (\pi^1, \ldots, \pi^d)$, with uniformly bounded first and second derivatives. Then, it is possible to define a diffusion operator $\pi(\mathcal{L})$ on $\mathbb{R}^d$ such that $\pi_\sharp \nu := (\pi_\sharp \nu_t)_{t\in[0,1]}$ is a solution to the associated FPE (in duality with $\mathcal{A} = C_b^{1,2}((0,T) \times \mathbb{R}^d)$). Indeed, the composition $f \circ \pi(t,x) := f(t, \pi(x))$ belongs to $\mathcal{A}$, and if we let $f \circ \pi$ in the weak formulation (2.4), the chain rule gives

$$\mathcal{L}(f \circ \pi) = \sum_{i=1}^d \mathcal{L}(\pi^i) \left[(\partial_i f) \circ \pi\right] + \frac{1}{2} \sum_{i,j=1}^d a(\nabla \pi^i, \nabla \pi^j) \left[(\partial_{i,j} f) \circ \pi\right].$$

We define, for $(t,x) \in (0,T) \times \mathbb{R}^d$,

$$\pi(a)_t^{i,j}(x) := \mathbb{E}_{\nu_t}\left[a(\nabla \pi^i, \nabla \pi^j) \,|\, \pi = x\right] = \frac{d\pi_\sharp \left[a(\nabla \pi^i, \nabla \pi^j)\nu_t\right]}{d\pi_\sharp \nu_t}(x), \quad \text{for } i,j \in \{1,\ldots,d\},$$

$$\pi(b)_t^i(x) := \mathbb{E}_{\nu_t}\left[\mathcal{L}(\pi^i) \,|\, \pi = x\right] = \frac{d\pi_\sharp \left[\mathcal{L}(\pi^i)\nu_t\right]}{d\pi_\sharp \nu_t}(x), \quad \text{for } i \in \{1,\ldots,d\}.$$

Then, $\pi(\mathcal{L}) := \mathcal{L}(\pi(a), \pi(b))$ is a diffusion operator on $\mathbb{R}^d$ and the (narrowly continuous) curve of measures $\pi_\sharp \nu$ is a weak solution of the FPE $\partial_t \pi(\nu) = \pi(\mathcal{L})^* \pi(\nu)$, in $(0,T)\times\mathbb{R}^d$. Let us remark that $\pi(\mathcal{L})$ depends upon $\nu$, although it is not evident in the notation.

Since conditional expectations reduce norms and the derivatives of $\pi^i$ are uniformly bounded, integral bounds on $a$, $b$ are naturally transferred on $\pi(a)$, $\pi(b)$: precisely, we use the fact that, for every convex, lower semicontinuous function $\Theta : \mathbb{R} \to [0,\infty]$ and non-negative Borel $f$ it holds

$$\int \Theta\left(\mathbb{E}_{\nu_t}\left[f \mid \pi\right]\right) d\pi_\sharp \nu_t \leq \int \Theta\left(f\right) d\nu. \tag{A.1}$$

In particular, uniform bounds on the coefficients are preserved; however, local integrability conditions could be lost.

**Mollification by convolutions.** This is a more standard technique, already employed e.g. in [5, Theorem 8.2.1] and [19, Theorem 2.6]. Let $\rho \geq 0$ be a smooth probability density (with respect to$\mathscr{L}^d$), with full support. Then, the family of measures $\nu * \rho := (\nu_t * \rho)_{t\in[0,1]}$, solve a FPE associated to a suitably defined diffusion operator. Indeed, for $f \in \mathcal{A}$, it holds $f * \rho \in \mathcal{A}$ with

$$\mathcal{L}(f*\rho) = \sum_{i=1}^d b^i \partial_i (f*\rho) + \frac{1}{2}\sum_{i,j=1}^d a^{i,j}\partial_{i,j}(f*\rho) = \sum_{i=1}^d b^i (\partial_i f)*\rho + \frac{1}{2}\sum_{i,j=1}^d a^{i,j}(\partial_{i,j} f)*\rho,$$

since derivatives and convolution commute. We define

$$(a^\rho)^{i,j}_t := \frac{d(a^{i,j}\nu_t)*\rho}{d(\nu_t*\rho)}, \quad (b^\rho)^i_t := \frac{d(b^i\nu_t)*\rho}{d(\nu_t*\rho)}, \quad \text{for } i,j \in \{1,\ldots,d\}.$$

so $(\nu_t * \rho)_{t\in[0,1]}$ is a weak solution of the FPE associated to $\mathcal{L}^\rho := \mathcal{L}(a^\rho, b^\rho)$, as

$$\partial_t \int f d(\nu*\rho) = \int (\partial_t f)*\rho \, d\nu = \int \partial_t (f*\rho)\, d\nu = -\int \mathcal{L}(f*\rho)\, d\nu = -\int \mathcal{L}^\rho f \, d(\nu*\rho).$$

Integrability and regularity properties of $a^\rho$ and $b^\rho$ are collected by the following lemma, see [5, Lemma 8.1.10] for a detailed proof.

**Lemma A.1.** *Let $\rho$ be a smooth probability kernel on $\mathbb{R}^d$ with $\rho > 0$ everywhere and $\left|\nabla^i \rho\right| \leq C\rho$, for $i \in \{1,\ldots k\}$, for some constant $C \geq 0$. Let $\mu$, $\nu \in \mathscr{M}^+(\mathbb{R}^d)$, with $\mu \ll \nu$. Then, it holds $\mu*\rho \ll \nu*\rho$, and the function*

$$\frac{d(\mu*\rho)}{d(\nu*\rho)}(x) = \frac{\int \rho(x-y)\, d\mu(y)}{\int \rho(x-y)\, d\nu(y)}, \quad \text{for } x \in \mathbb{R}^d$$

*provides a $C^k(\mathbb{R}^d)$ version of the density $d(\mu*\rho)/d(\nu*\rho)$. Moreover, for every convex, lower semicontinuous function $\Theta : \mathbb{R} \to [0,\infty]$, it holds*

$$\int \Theta\left(\frac{d(\mu*\rho)}{d(\nu*\rho)}\right) d(\nu*\rho) \leq \int \Theta\left(\frac{d\mu}{d\nu}\right) d\nu. \tag{A.2}$$

*Similar conclusions hold when $\mu = (\mu_t)_{t\in[0,1]} \subseteq \mathscr{M}^+(\mathbb{R}^d)$ is a Borel curve and $\nu = (\nu_t)_{t\in[0,1]} \subseteq \mathscr{P}(\mathbb{R}^d)$ is narrowly continuous, with $\mu_t \ll \nu_t$ for every $t \in [0,1]$. In addition, it holds*

$$\sup_{t\in[0,1]} \left\| \frac{d(\mu_t*\rho)}{d(\nu_t*\rho)} \right\|_{C^k_b(B)} < \infty,$$

*for every open bounded set $B \subseteq \mathbb{R}^d$.*

When applied to a solution $\nu = (\nu_t)_{t\in[0,1]}$ of the FPE (2.2) we deduce that, if $a$, $b \in L^p(\nu)$, then $a^\rho$, $b^\rho \in L^p(\nu*\rho)$ (for $p \in [1,\infty]$) and $a^\rho_t$, $b^\rho_t$ are $C^k(\mathbb{R}^d)$, uniformly in $t \in [0,1]$, with uniformly bounded first and second (spatial) derivatives on compact sets of $[0,T]\times\mathbb{R}^d$.

### A.2 Tightness

We prove a compactness criterion for solutions of martingale problems, under minimal integrability conditions on the coefficients. In the deterministic case, tightness is achieved by estimating the metric velocity of absolutely continuous curves which solve the ODE; in the stochastic case, we rely on analogous results for martingales, using Burkholder-Davis-Gundy inequalities and an argument reminiscent of Lévy's modulus of continuity for the Brownian motion, to estimate the modulus of continuity of the canonical process (yielding in some cases Hölder regularity).

**Theorem A.2.** *Let $\theta$, $\Theta_1$, $\Theta_2 : [0,+\infty) \to [0,+\infty)$ be functions with $\Theta_1$, $\Theta_2$ convex, l.s.c.,*

$$\lim_{x\to\infty} \theta(x) = \lim_{x\to+\infty} \frac{\Theta_1(x)}{x} = \lim_{x\to+\infty} \frac{\Theta_2(x)}{x} = \infty$$

*and, for some constant $C \geq 0$, $\Theta_2(2x) \leq C\Theta_2(x)$, for $x \geq 0$. Then, there exists some coercive function $\Psi : C([0,1];\mathbb{R}) \to [0,+\infty]$ (i.e. $\{\Psi \leq c\}$ is compact for every $c \geq 0$) such that, for every filtered probability space $(\Omega, (\mathcal{F}_t)_{t\in[0,1]}, \mathbb{P})$ and progressively measurable processes $\varphi = (\varphi_t)$, $\beta = (\beta_t)_t$, $\alpha = (\alpha_t)_t$ with*

$$[0,1] \ni t \mapsto M_t := \varphi_t - \int_0^t \beta_s \, ds, \quad \text{and} \quad [0,1] \ni t \mapsto M_t^2 - \int_0^t \alpha_s \, ds$$

*$\mathbb{P}$-a.s. continuous local martingales, and $\alpha \geq 0$ $\mathbb{P}$-a.s. it holds*

$$\mathbb{E}\left[\Psi(\varphi)\right] \leq \mathbb{E}\left[\theta(\varphi_0) + \int_0^1 \left[\Theta_1\left(|\beta_t|\right) + \Theta_2\left(\alpha_t\right)\right] dt\right]. \tag{A.3}$$

**Remark A.3.** *We introduce the "moderate growth" assumption $\Theta_2(2x) \leq C\Theta_2(x)$, for $x \geq 0$, to apply suitable versions of Burkholder-Gundy inequalities. In the proof of Theorem 2.5, we argue with some $\Theta_2$ provided by de la Vallée Poussin criterion: despite the fact that such growth assumption may fail for $\Theta_2$, we may always introduce some $\tilde{\Theta}_2$ such that all the assumptions of the theorem above hold, and $\tilde{\Theta}_2 \leq c\Theta_2$, for some constant $c > 0$. Hence, if the right hand side in* (A.3) *is finite, it is finite as well with $\tilde{\Theta}_2$ in place of $\Theta_2$.*

*A way to build such a $\tilde{\Theta}_2$ is then via Legendre duality. Indeed, without any loss of generality, we may assume that the given convex, l.s.c. function $\Theta_2$ satisfies $\Theta_2(0) = 0$ and $(\Theta_2)^*(x) = 0$, for $x \in [0,1]$, where $\Theta_2^*(y) = \sup_{x\geq 0}\{xy - \Theta_2(x)\}$, for $y \in [0,\infty)$, is the Legendre transform of $\Theta_2$. Indeed, if this is not the case, we may introduce the function $x \mapsto \Theta_2(x) - \Theta_2(0) + x$, for which one has, for some constant $c > 0$, $(\Theta_2(x) - \Theta_2(0) + x) < c\Theta_2(x)$, for every $x \in [0,\infty)$, because $\Theta_2$ has "faster than linear growth" at $\infty$.*

*To provide $\tilde{\Theta}_2$, we first define a convex, l.s.c. function $G : [0,\infty) \to [0,\infty]$ with $G(0) = 0$, and then we let $\tilde{\Theta}_2(x) := \sup_{y\geq 0}\{xy - G(y)\}$. In particular, the "faster than linear growth" at $\infty$ for $\tilde{\Theta}_2$ turns out to be equivalent to $G(y) < \infty$ for every $y \in [0,\infty)$, and the moderate grow assumptions reads as follows: for some (fixed) $c > 0$,*

$$cG\left(\frac{2}{c}y\right) \leq cG\left(y\right), \quad \text{for every } y \in [0,\infty). \tag{A.4}$$

*Finally, $\tilde{\Theta}_2 \leq \Theta_2$, is equivalent to $G \geq \Theta_2^*$. Motivated by these requirements, we define e.g.*

$$G(y) := \sup_{k\in\mathbb{N}} 3^k \Theta_2^*\left(\left(\frac{2}{3}\right)^k y\right) \text{ for } y \in [0,\infty).$$

*Being the supremum of convex, l.s.c. functions null at* $0$*,* $G$ *is convex, l.s.c. with* $G(0)=0$*. Since* $\left(\frac{2}{3}\right)^k y<1$ *for* $k$ *sufficiently large, one has* $G(y)<\infty$ *and* $G(y)\geq \Theta_2^*(y)$ *is obvious. Finally, one has*

$$3G\left(\frac{2}{3}y\right)=\sup_{k\in\mathbb{N}} 3^{k+1}\Theta_2^*\left(\left(\frac{2}{3}\right)^{k+1}y\right)\leq G(y)\,,\quad \textit{for every } y\in[0,\infty).$$

*hence* (A.4) *holds with* $c=3$*.*

*Proof.* Clearly, if the right hand side above is not finite, the conclusion trivially holds. We prove separately the existence of functions $\Psi_1$ and $\Psi_2$, depending respectively on $\Theta_1$, $\Theta_2$ only, taking integer and non-negative values, such that, if we let

$$\Psi(\gamma):=\theta(|\gamma_0|)+\inf_{\gamma^1+\gamma^2=\gamma}\left\{\Psi_1(\gamma^1)+\Psi_2(\gamma^2)\right\},\tag{A.5}$$

then $\Psi$ is coercive and (A.3) holds. For every $\varepsilon>0$, $i\in\{1,2\}$ we let $\delta_i=\delta_{i,\varepsilon}$ be the largest number in the form $1/n$, (with $n\geq 1$ natural) such that $\Theta_i(\varepsilon^i/\delta)\delta\geq\varepsilon^{-1}$: such a choice is possible because $\lim_{x\to+\infty}\Theta_i(x)/x=\infty$. Then, we introduce the closed sets

$$A_i(\varepsilon):=\Big\{\gamma\in C([0,1],\mathbb{R})\,:\,\sup_{k=1,\ldots,\delta_i^{-1}}\ \sup_{s\in[(k-1)\delta_i,k\delta_i]}|\gamma_s-\gamma_{(k-1)\delta_i}|\leq\varepsilon\Big\},\tag{A.6}$$

and we let $\Psi_i(\gamma):=\sum_{m\geq 0}(m+1)\chi_{A_i(2^{-m})^c}$. By construction, $\Psi(\gamma)\leq m$ entails $\gamma\in A_i(2^{-k})$, for every $k\geq m$.

To show that $\Psi$ defined by (A.5) is coercive, it is sufficient to apply Ascoli-Arzelà criterion, noticing that $\gamma\in\{\Psi\leq m\}$ can be decomposed as the sum of two curves $\gamma_1+\gamma_2$, and $\gamma_i$ ($i\in\{1,2\}$) admits the following modulus of continuity

$$\omega_{i,m}(x):=\begin{cases}2^{1-k} & \text{if } x\in[\delta_{i,2^{-(k+1)}},\delta_{i,2^{-k}}) \text{ with } k\geq m,\\ 2^{1-m}/\delta_{i,2^{-m}} & \text{if } x\in[\delta_{i,2^{-m}},+\infty).\end{cases}\tag{A.7}$$

To show that (A.3) holds, we assume that the right hand side therein is finite. The assumptions entail therefore that $(M_t)_t$ is a $\mathbb{P}$-a.s. continuous local martingale, whose quadratic variation process is $t\mapsto\int_0^t\alpha_s ds$. If we let $\gamma_t^1:=\int_0^t\beta_s ds$ and $\gamma_t^2=M_t$, for $t\in[0,1]$, then the left hand side in (A.3) is smaller than

$$\mathbb{E}\left[\theta(\varphi_0)+\Psi_1(\gamma^1)+\Psi_2(\gamma^2)\right]\leq\mathbb{E}\left[\theta(\varphi_0)\right]+\sum_{m\geq 0}(m+1)\mathbb{E}\left[\chi_{A_1(2^{-m})}\circ\gamma^1+\chi_{A_2(2^{-m})}\circ\gamma^2\right].$$

Next, we focus on the addends in the series above, writing for brevity $\varepsilon$ in place of $2^{-m}$. For $i\in\{1,2\}$, using (A.6), we have

$$\mathbb{E}\left[\chi_{A_i(\varepsilon)^c}\circ\gamma^i\right]=P\Big(\sup_{k=1,\ldots,\delta_i^{-1}}(\gamma^i)_k^*>\varepsilon\Big)\leq\sum_{k=1}^{\delta_i^{-1}}P\left((\gamma^i)_k^*>\varepsilon\right),$$

where we write, $(\gamma^i)_k^*:=\sup_{s\in[(k-1)\delta_i,k\delta_i]}|\gamma_s^i-\gamma_{(k-1)\delta_i}^i|$.

Let us focus on the case $i=1$ (thus we write $\delta=\delta_1$, $\Theta=\Theta_1$). Since $|\gamma_s-\gamma_t|\leq\int_s^t|\beta_r|\,dr$, we estimate

$$P\left((\gamma^1)_k^*>\varepsilon\right)\leq\frac{\mathbb{E}\left[\Theta\left(\frac{1}{\delta}\int_{(k-1)\delta}^{k\delta}|\beta_s|\,ds\right)\right]}{\Theta\left(\varepsilon/\delta\right)}\leq\varepsilon\mathbb{E}\left[\int_{(k-1)\delta}^{k\delta}\Theta\left(|\beta_s|\right)ds\right],$$

where the last inequality is a consequence of Jensen's inequality and our preliminary choice for $\delta$. Summing upon $k\in\left\{1,\ldots,\delta^{-1}\right\}$, we conclude that $\mathbb{E}\left[\Psi\circ\gamma\right]\leq$

$c\int_0^1 \mathbb{E}\left[\Theta\left(|\beta_s|\right)\right] ds$, for some constant $c \geq 0$ (in this case, the constant does not even depend upon $\Theta$).

To deal with the case $i = 2$ (again, we omit to specify $i$ in what follows), i.e., the martingale part, for each $k \in \left\{1, \ldots, \delta^{-1}\right\}$, we estimate from above,

$$P\left(M_k^* > \varepsilon\right) \leq \frac{\mathbb{E}\left[\Theta\left((M_k^*)^2/\delta\right)\right]}{\Theta\left(\varepsilon^2/\delta\right)} \leq c_\Theta \frac{\mathbb{E}\left[\Theta\left(\frac{1}{\delta}\int_{(k-1)\delta}^{k\delta} \alpha_s ds\right)\right]}{\Theta\left(\varepsilon^2/\delta\right)},$$

where $c_\Theta$ is some constant depending on $\Theta$ only: indeed, it is sufficient to apply Burkholder-Davis-Gundy inequalities, e.g. in the form [25, Theorem 2.1], to the martingale $M_s := \delta^{-1/2} M_{s+(k-1)\delta}$, $s \in [0, \delta]$ and the convex function with "moderate growth" $x \mapsto \Theta(x^2)$. By Jensen's inequality and our definition of $\delta_\varepsilon$ we conclude that

$$\frac{\mathbb{E}\left[\Theta\left(\frac{1}{\delta}\int_{(k-1)\delta}^{k\delta} \alpha_s ds\right)\right]}{\Theta\left(\varepsilon^2/\delta\right)} \leq \varepsilon \mathbb{E}\left[\int_{(k-1)\delta}^{k\delta} \Theta\left(\alpha_s\right) ds\right].$$

As in the previous case, by summing upon $k \in \left\{0, \ldots, \delta^{-1}\right\}$, we deduce that

$$\mathbb{E}\left[\chi_{A(\varepsilon)^c} \circ M\right] \leq \varepsilon c_\Theta \mathbb{E}\left[\int_0^1 \Theta\left(\alpha_s\right) ds\right]$$

and so we deduce the desired bound for $\mathbb{E}\left[\Psi(M)\right]$. □

**Corollary A.4.** *In the situation of the theorem above, let* $\Theta_1(x) = |x|^{p_1}$ *and* $\Theta_2(x) = |x|^{p_2}$, *for* $p_1$, $p_2 \in (1, \infty)$ *and assume that the right hand side in* (A.3) *is finite. Then, for every* $r > 0$ *with* $r < r(p_1, p_2) := \min\left\{1 - \frac{1}{p_1}, \frac{1}{2}\left(1 - \frac{1}{p_2}\right)\right\}$, *it holds*

$$\mathbb{P}\left(\limsup_{h\downarrow 0} \sup_{|t-s|\leq h} \frac{|\varphi_t - \varphi_s|}{|t-s|^r} = 0\right) = 1.$$

*Proof.* It is sufficient to let $\delta_i := \varepsilon^{1/r}$, for $i \in \{1, 2\}$. Thanks to this choice, the probabilities of $A_i(2^{-m})^c$ decay sufficiently fast as $m \to \infty$ so that, by Borel-Cantelli lemma, there exists $P$-a.s. some $m \geq 1$ such that the curve $(\varphi_t)_{t\in[0,1]}$ can be written as a sum of two curves having $\omega_{i,m}$, defined in (A.7), as a modulus of continuity. This entails $r$-Hölder estimates for $\varphi$: since the condition on $r$ is open-ended, by arguing with a $\tilde{r}$ slightly larger than $r$, the thesis follows. □

It is not clear if $\varphi$ in the previous corollary is actually $\mathbb{P}$-a.s. Hölder continuous with exponent $r(p_1, p_2)$: one might exploit the existence of functions $\tilde{\Theta}_i$ ($i \in \{1, 2\}$) with $\tilde{\Theta}_i(x)/|x|^{p_i} \to \infty$ as $x \to \infty$, and the right hand side in (A.3) still finite, but it seems insufficient.

**Corollary A.5.** *Let* $a$, $b$ *be Borel maps as in* (2.1), *let* $\boldsymbol{\eta} \in \mathscr{P}(C([0,1]; \mathbb{R}^d))$ *be a solution of the martingale problem associated to* $\mathcal{L}(a, b)$. *For any* $\theta$, $\Theta_1$ *and* $\Theta_2$, *as in the theorem above, let* $\Psi$ *be the associated coercive functional. Then, for every* $f \in C_b^{1,2}((0,T) \times \mathbb{R}^d)$, *it holds*

$$\int \Psi(f_t \circ e_t) d\boldsymbol{\eta} \leq \int \theta(|f_0|) d\eta_0 + \int_0^1 \left[\Theta_1(|\partial_t f + \mathcal{L}_t f|) + \Theta_2(a_t(\nabla f_t, \nabla f_t))\right] d\eta_t dt.$$

*Proof.* We prove that $t \mapsto \varphi_t := f_t \circ e_t$ satisfies the assumptions of Theorem A.2, with $\beta_t := (\partial_t + \mathcal{L}_t) f_t \circ e_t$ and $\alpha_t := a_t(\nabla f_t, \nabla f_t) \circ e_t$. For simplicity of notation, we omit to write

$e_t$ below (since its appearance is quite natural). Since both $f$ and $f^2 \in C_b^{1,2}((0,T) \times \mathbb{R}^d)$, both

$$t \mapsto M_t^f := f_t - f_0 - \int_0^t (\partial_s + \mathcal{L}_s) f \, ds, \quad \text{and} \quad t \mapsto M_t^{f^2} := f_t^2 - f_0^2 - \int_0^t (\partial_s + \mathcal{L}_s) f^2 \, ds$$

are martingales. By developing $(M_t^f)^2$, we see that

$$t \mapsto (M_t^f)^2 - \int_0^t (\partial_s + \mathcal{L}_s) f^2 \, ds - 2 \int (\partial_s + \mathcal{L}_s) f \left[ \int_s^t (\partial_r + \mathcal{L}_r) f dr - f_t \right] ds$$

is also a martingale. We add and subtract the process $t \mapsto 2 \int_0^t f_s (\partial_t + \mathcal{L}) f_s ds$, thus

$$t \mapsto (M_t^f)^2 - \int_0^t \alpha_s \, ds + 2 \int_0^t (\partial_s + \mathcal{L}_s) f \left[ f_t - f_s - \int_s^t (\partial_r + \mathcal{L}_r) f dr \right] ds$$

is a martingale. To conclude, we notice that

$$t \mapsto \int_0^t (\partial_s + \mathcal{L}_s) f \left[ f_t - f_s - \int_s^t (\partial_r + \mathcal{L}_r) f \, dr \right] ds = \int_0^t \beta_s \left( M_t^f - M_s^f \right) ds$$

is a local martingale (see also [31, Theorem 1.2.8]). Indeed, by a stopping time argument, we are easily reduced to the case where $M^f$ is replaced by a martingale $M$ with $M_1 \in L^\infty(\mathbb{P})$, thus $\int_0^t \beta_s (M_t - M_s) \, ds \in L^1(\mathbb{P})$, for $t \in [0,1]$. To prove that increments are orthogonal, we let $t \in [0,1]$ and show that

$$\mathbb{E} \left[ \int_0^1 \beta_s (M_1 - M_s) \, ds \,|\, \mathcal{F}_t \right] = \int_0^t \beta_s (M_t - M_s) \, ds.$$

By the integrability assumptions, we exchange between conditional expectation and integration. The thesis follows by direct consideration of the cases, $s \in [0,t]$ and $s \in (t,1]$. □

### A.3 Limit

In the third step, we assume that the probability measures $(\boldsymbol{\eta}^n)_n$, obtained as superposition solutions for a suitable approximating sequence $(\nu^n)_n$ narrowly converge in $\mathscr{P}(C([0,T];\mathbb{R}^d))$ towards some limit $\boldsymbol{\eta}$. The fact that $\boldsymbol{\eta}$ provides a superposition solution for $\nu$ is not straightforward, since we must deal with a limit in the weak formulation, where terms involving the coefficients $a$, $b$ appear (in general, they are not continuous).

Indeed, $\boldsymbol{\eta} \in P(C([0,1];\mathbb{R}^d))$ is a solution of the martingale problem associated to $\mathcal{L}(a,b)$ if and only if the following property holds: for every $s$, $t \in [0,1]$ with $s \le t$, for every $f \in C_c^{1,2}([0,1] \times \mathbb{R}^d)$ (with $\|f\|_{C^{1,2}} \le 1$) and for every bounded continuous and $\mathcal{F}_s$-measurable function $g$ on $C([0,T];\mathbb{R}^d)$ (with $\|g\|_\infty \le 1$) it holds

$$\int g \left[ f_t \circ e_t - f_s \circ e_s - \int_s^t [(\partial_t + \mathcal{L}_r) f] \circ e_r \, dr \right] d\boldsymbol{\eta} = 0.$$

As the correspondent identity holds for $\boldsymbol{\eta}^n$ and $\mathcal{L}^n$, i.e.

$$\int g \left[ f_t \circ e_t - f_s \circ e_s - \int_s^t [(\partial_t + \mathcal{L}_r^n) f] \circ e_r \, dr \right] d\boldsymbol{\eta}^n = 0,$$

to deduce that $\boldsymbol{\eta}$ is a solution to the martingale problem associated to $\mathcal{L}$, since $f$ and $\partial_t f$ are bounded and continuous, the crucial limit is

$$\int g \left[ \int_s^t (\mathcal{L}_r^n f) \circ e_r \, dr \right] d\boldsymbol{\eta}^n - \int g \left[ \int_s^t (\mathcal{L}_r f) \circ e_r \, dr \right] d\boldsymbol{\eta} \to 0, \tag{A.8}$$

whose validity we now investigate, according to the approximations from Section A.1.

**Push forward via smooth maps.** For $n \ge 1$, let $\pi^n \in C^2_b(\mathbb{R}^d, \mathbb{R}^d)$ with $(\pi_n)_n$ converging to the identity map locally uniformly and assume that the sequence of first and second derivatives converge (towards the respective limits), pointwise and uniformly bounded, i.e., $\nabla \pi^n(x) \to Id$ for $x \in \mathbb{R}^d$, $\nabla^2 \pi^n(x) \to 0$, for every $x \in \mathbb{R}^d$, and there exists some constant $C \ge 0$ such that $\left|\nabla^i \pi^n(x)\right| \le C$, for $x \in \mathbb{R}^d$ and $i \in \{1,2\}$.

Let $\nu^n = \pi^n_\sharp \nu$, $\pi^n(\mathcal{L})$, and let $\boldsymbol{\eta}^n$ be corresponding superposition solution. To prove that any narrow limit point $\boldsymbol{\eta}$ is indeed a superposition solution for $\nu$, with respect to the diffusion operator $\mathcal{L}$, we add and subtract the term

$$\int g \left[ \int_s^t (\overline{\mathcal{L}}_r f) \circ e_r \, dr \right] d\boldsymbol{\eta}^n - \int g \left[ \int_s^t (\overline{\mathcal{L}}_r f) \circ e_r \, dr \right] d\boldsymbol{\eta}$$

in (A.8), where $\overline{\mathcal{L}} = \mathcal{L}(\overline{a}, \overline{b})$ is any diffusion operator on $\mathbb{R}^d$, whose coefficients $\overline{a}$, $\overline{b}$ are continuous and compactly supported. The difference terms above are infinitesimal as $n \to \infty$, by narrow convergence of $\boldsymbol{\eta}^n$, thus we estimate (A.8), as $n \to \infty$, in terms of

$$\limsup_{n \to \infty} \int \left| \mathcal{L}^n f - \overline{\mathcal{L}} f \right| d\nu^n + \int \left| \mathcal{L} f - \overline{\mathcal{L}} f \right| d\pi(\nu). \tag{A.9}$$

Let us focus on first term above, at fixed $n \ge 1$ (for simplicity of notation, we drop the dependence upon $n$). Recalling the definition of $\pi(\mathcal{L})$,integration with respect to the push-forward measure gives

$$\int \left| \pi(\mathcal{L}) f - \overline{\mathcal{L}} f \right| d\pi_\sharp \nu = \int \left| \mathbb{E}_\nu \left[ \mathcal{L}(f \circ \pi) \,|\, \pi \right] - (\overline{\mathcal{L}} f) \circ \pi \right| d\nu.$$

Being $(\overline{\mathcal{L}} f) \circ \pi$ a function of $\pi$, up to $\nu$-negligible sets, we have

$$\begin{aligned} \int \left| \mathbb{E}_\nu \left[ \mathcal{L}(f \circ \pi) \,|\, \pi \right] - (\overline{\mathcal{L}} f) \circ \pi \right| d\nu &= \int \left| \mathbb{E}_\nu \left[ \mathcal{L}(f \circ \pi) - (\overline{\mathcal{L}} f) \circ \pi \,|\, \pi \right] \right| d\nu \\ &\le \int \left| \mathcal{L}(f \circ \pi) - (\overline{\mathcal{L}} f) \circ \pi \right| d\nu, \end{aligned}$$

since conditional expectation reduces $L^1(\nu)$-norms. Writing explicitly the difference

$$\mathcal{L}(f \circ \pi) - (\overline{\mathcal{L}} f) \circ \pi = \frac{1}{2} \sum_{i,j=1}^k \left[ a(\nabla \pi^i, \nabla \pi^j) - \overline{a}^{i,j} \circ \pi \right] (\partial_{i,j} f) \circ \pi + \sum_{i=1}^k \left[ \mathcal{L}(\pi^i) - \overline{b}^i \circ \pi \right] (\partial_i f) \circ \pi,$$

and recalling that $\|f\|_{C^{1,2}} \le 1$, we conclude that

$$\int \left| \pi(\mathcal{L}) f - \overline{\mathcal{L}} f \right| d\pi_\sharp \nu \le \int \frac{1}{2} \sum_{i,j=1}^k \left| a(\nabla \pi^i, \nabla \pi^j) - \overline{a}^{i,j} \circ \pi \right| d\nu + \int \sum_{i=1}^k \left| \mathcal{L}(\pi^i) - \overline{b}^i \circ \pi \right| d\nu.$$

Letting $n \to \infty$ (recall that $\pi = \pi^n$ above), using the assumption on the convergence of $\pi^n$ towards the identity map (in particular, we use Lebesgue dominated convergence w.r.t. the measure $\nu$), we deduce that (A.9) is bounded from above by twice the expression

$$\int \frac{1}{2} \sum_{i,j=1}^k \left| a^{i,j} - \overline{a}^{i,j} \right| d\nu + \int \sum_{i=1}^k \left| b^i - \overline{b}^i \right| d\nu.$$

To conclude, we choose $\overline{a}$, $\overline{b}$ that minimize the right hand side above: this can be made arbitrary small, by the density of continuous and compactly supported functions in $L^1(\nu)$).

**Mollification by convolution.** In this case, the argument is similar, and even more standard, see e.g. [5, Theorem 8.2.1], thus we only sketch it. Given a sequence $\rho^n$ of probability densities on $\mathbb{R}^d$ such that $\rho^n \mathscr{L}^d \to \delta_0$ narrowly as $n \to \infty$, let $\nu^n = \nu * \rho^n$ and $\mathcal{L}^n$ be the diffusion operator introduced in Section A.1. We add and subtract, in (A.8),

$$\int g \left[ \int_s^t \overline{\mathcal{L}}_r f \circ e_r \, dr \right] d\boldsymbol{\eta}^n - \int g \left[ \int_s^t \overline{\mathcal{L}}_r f \circ e_r \, dr \right] d\boldsymbol{\eta},$$

where $\overline{\mathcal{L}} = \mathcal{L}(\overline{a}, \overline{b})$ has continuous and compactly supported coefficients. Let $\overline{\omega}$ be a common (bounded and continuous) modulus of continuity for $\overline{a}$, $\overline{b}$.

As in the previous case, narrow convergence implies that the absolute value of (A.8) is bounded from above, as $n \to \infty$, by

$$\limsup_{n\to\infty} \int \left| \mathcal{L}^n f - \overline{\mathcal{L}} f \right| d\nu^n + \int \left| \mathcal{L} f - \overline{\mathcal{L}} f \right| d\nu.$$

First, we prove that $\lim_{n\to\infty} \int |\overline{\mathcal{L}}^n f - \overline{\mathcal{L}} f| d\nu^n = 0$, where $\overline{\mathcal{L}}^n$ has coefficients

$$\overline{a}^n := \frac{d(\overline{a}\nu * \rho_n)}{d(\nu * \rho_n)}, \qquad \overline{b}^n := \frac{d(\overline{b}\nu * \rho_n)}{d(\nu * \rho_n)}.$$

Indeed, recalling that $\|f\|_{C^{1,2}} \le 1$, we estimate

$$\begin{aligned}
\int \left| \overline{\mathcal{L}}^n f - \overline{\mathcal{L}} f \right| d\nu^n &\le \int |\overline{a}^n(x) - \overline{a}(x)| \, d\nu^n + \int |\overline{b}^n - \overline{b}| \, d\nu^n \\
&= \int |(\overline{a}\nu * \rho_n)(x) - \overline{a}(x)(\nu * \rho_n)(x)| + \left| (\overline{b}\nu * \rho_n)(x) - \overline{b}(x)(\nu * \rho_n)(x) \right| dx \\
&\le 2 \int \left[ \int \overline{\omega}(y-x) \rho_n(y-x) dx \right] \nu(dy) = 2 \int \overline{\omega}(z) \rho_n(z) dz \to 0.
\end{aligned}$$

Thanks to this fact, we write

$$\begin{aligned}
\limsup_{n\to\infty} \int \left| \mathcal{L}^n f - \overline{\mathcal{L}} f \right| d\nu^n &= \limsup_{n\to\infty} \int \left| \mathcal{L}^n f - \overline{\mathcal{L}}^n f \right| d\nu^n \\
&= \limsup_{n\to\infty} \int \left( |a^n - \overline{a}^n| + |b^n - \overline{b}^n| \right) d\nu^n \\
&\le \int \left( |a - \overline{a}| + |b - \overline{b}| \right) d\nu
\end{aligned}$$

where in the last step we apply (A.2). To conclude, it is sufficient to optimize upon $\overline{a}$, $\overline{b}$, by density of continuous and compactly supported functions in $L^1(\nu)$.

### A.4 Proof of Theorem 2.5

We argue by iterating the three-steps scheme, the base case being that of diffusion operators with smooth and uniformly bounded coefficients. First, we extend the validity to the case of uniformly bounded coefficients (without any regularity assumption), then to that of locally bounded coefficients, and finally integrable coefficients. Although everything could be obtain in a single iteration, we think the approach highlights the different roles played by different approximation procedures.

Indeed, our crucial improvement with respect to [19, Theorem 2.6] is to move from uniformly bounded to integrable coefficients, which is rather delicate: by comparison, in the deterministic case, one is able to deal directly with locally smooth coefficients (see e.g. [5, Proposition 8.1.8]), essentially because paths either go to infinity, i.e., the solution explodes in a finite time, or stay in a compact set. Roughly speaking, the source of difficulties in the stochastic case is that we have to deal with "averages" of such behaviours, and moreover the solution to a genuinely stochastic martingale problem is expected to instantaneously "diffuse" over compact sets (of course, with small probability as these sets become larger).

**Case of smooth and bounded coefficients.** Let $a$, $b$ be Borel maps as in (2.1), with

$$\int_0^T \left[ \|a_t\|_{C_b^2(\mathbb{R}^d)} + \|b_t\|_{C_b^2(\mathbb{R}^d)} \right] dt < \infty. \tag{A.10}$$

Then, the superposition principle holds for every solution $\nu = (\nu_t)_{t \in (0,T)} \subseteq \mathscr{P}(\mathbb{R}^d)$ of the FPE (2.2). This follows from two well-known facts: existence of Itô's stochastic differential equations and uniqueness for narrowly continuous solutions of FPE's.

The existence result is standard, with the possible exception of the integrable bounds with respect to the variable $t \in [0,T]$ (usually, one requires uniform bounds), but in fact such condition is sufficient for the various applications of Gronwall inequality. For the sole purpose of establishing a case base for the superposition principle, the usual stronger assumptions on the coefficients, e.g. $a$, $b \in C_b^\infty((0,T) \times \mathbb{R}^d)$ would even be sufficient, at the price of introducing an extra mollification step with respect to the variable $t \in [0,T]$.

**Theorem A.6.** *Let $a$, $b$ be Borel maps as in* (2.1)*, satisfying* (A.10)*. Then, for every $\bar\nu \in \mathscr{P}(\mathbb{R}^d)$, there exists a solution $\boldsymbol{\eta}$ of the MP associated to $\mathcal{L}(a,b)$, with $\eta_0 = \bar\nu$.*

*Proof.* The assumption $a \in L^1_t(C^2_b(\mathbb{R}^d))$ entails that the symmetric non-negative square-root of $a$, i.e. the (essentially unique) map

$$\sigma : [0,T] \times \mathbb{R}^d \to \mathrm{Sym}_+(\mathbb{R}^d) \quad \text{such that } \sigma_t^2 = a_t, \quad \mathscr{L}^1\text{-a.e. } t \in (0,T),$$

is bounded and Lipschitz with respect to $x \in \mathbb{R}^d$, with Lipschitz constant integrable w.r.t. $t \in (0,T)$, see e.g. [31, Lemma 3.2.3]. Then, it is sufficient to solve by Picard iteration the Itô stochastic differential equation

$$dX_t = b_t(X_t)dt + \sigma_t(X_t)dW_t, \quad X_0 = \overline{X},$$

where $\overline{X}$ is a r.v. independent of the $d$-dimensional Wiener process $W$. By Itô formula, the law of $X$, i.e. $X_\sharp \mathbb{P}$, is a solution of the martingale problem associated to $\mathcal{L}(a,b)$. ☐

Of course, the MP is also well-posed, but we need a stronger uniqueness result, for narrowly continuous solutions of FPE's, which is e.g. a consequence of results on backward Kolmogorov equations. We refer e.g. to the expository notes by [21] for more details; notice however that, also in this case, the standard literature studies equations of the form

$$\partial_t f = -\mathcal{L}_t f + g, \quad \text{in } (0,T) \times \mathbb{R}^d, \quad f_T = \bar f, \tag{A.11}$$

assuming $a$, $b$ smooth and $g \in C_c^\infty((0,T) \times \mathbb{R}^d)$. A solution to the equation (A.11) is defined as a function $f \in C_b^{1,2}((0,T) \times \mathbb{R}^d)$ such that

$$\partial_t f(s,x) = -\mathcal{L}_s f(s,x) + g(s,x), \quad \text{for } (s,x) \in (0,T) \times \mathbb{R}^d, \quad \text{with} \quad \lim_{s \uparrow T} f(s,x) = \bar f(x).$$

To our purposes, we need existence of a solution, together the following regularity results for the solution $f$ (which entails uniqueness):

$$\sup_{t \in [0,T]} \|f_t\|_{C^2_x} \le \left( \|\bar f\|_{C^2_x} + T \|g\|_{C^2_{t,x}} \right) C \left( \int_0^T \left[ \|a_t\|_{C^2_x} + \|b_t\|_{C^1_x} \right] dt + T \|g\|_{C^2_{t,x}} \right), \tag{A.12}$$

where $z \mapsto C(z)$ denotes some function depending on the dimension $d$ only (the proof gives that $C$ has an exponential behaviour). The proof follows by direct differentiation of the equation, see [31, Theorem 3.2.4] for a detailed derivation. Moreover, as a consequence of the maximum principle, if $\bar{f} \geq 0$, and $g \geq 0$, then the solution $f$ is non-negative as well.

We are in a position to prove the following result, akin to [5, Proposition 8.1.7]. Again, we provide a slightly stronger statement than what is needed for the superposition principle (e.g., we deduce uniqueness for possibly signed solutions of the FPE).

**Theorem A.7.** *Let $a$, $b$ be Borel maps as in* (2.1) *with*

$$\int_0^T \|a_t\|_{C^2(B)} + \|b_t\|_{C^2(B)} \, dt < \infty, \quad \text{for every bounded open } B \subseteq \mathbb{R}^d,$$

*and $\nu = (\nu_t)_{t\in[0,T]} \subseteq \mathscr{M}(\mathbb{R}^d)$ be a narrowly continuous solution of the FPE associated* (2.2)*. If $\nu_0 \leq 0$, then $\nu_t \leq 0$, for every $t \in [0,T]$. Thus, for $\bar{\nu} \in \mathscr{M}(\mathbb{R}^d)$ there exists at most one narrowly continuous solution $\nu$ with $\nu_0 = \bar{\nu}$.*

*Proof.* Let $g \in C_c^\infty((0,T)\times\mathbb{R}^d)$, with $g \geq 0$: it is sufficient to show that $\int g \, d\nu \leq 0$. Fix $R \geq 1$ large enough so that the support of $g$ is contained in $(0,T)\times B_R(0)$ and let $\chi_R$ be a cut-off function, as below Remark 2.3. Notice that letting $a_R = a\chi_R$ and $b_R = b\chi_R$ in place of $a$, $b$, condition (A.10) holds and $\mathcal{L}_R f = \mathcal{L} f$ on $(0,T)\times B_R(0)$, for every $f \in C_b^2((0,T)\times\mathbb{R}^d)$.

For $\varepsilon > 0$, let $a_R^\varepsilon$, $b_R^\varepsilon$ be a double mollification with respect to the space and time variables, and define $\mathcal{L}_R^\varepsilon = \mathcal{L}(a_R^\varepsilon, b_R^\varepsilon)$, which is a diffusion operator with smooth and bounded coefficients, satisfying (A.10) uniformly in $\varepsilon > 0$. Let $f^\varepsilon$ be a solution to the backward Kolmogorov equation

$$\partial_t f^\varepsilon = -\mathcal{L}_R^\varepsilon f^\varepsilon + g, \quad f_T^\varepsilon = 0,$$

and choose $f^\varepsilon \chi_R$ in the weak formulation (2.4), which is admissible because $f^\varepsilon \in C_b^{1,2}((0,T)\times\mathbb{R}^d)$. Since $f^\varepsilon \leq 0$ and $\nu_0 \leq 0$, we have

$$\begin{aligned}
0 &\geq -\int f^\varepsilon \chi_R \, d\nu_0 = \int \left[\chi_R \, \partial_t f^\varepsilon + \mathcal{L}(f^\varepsilon \chi_R)\right] d\nu \\
&= \int \left[-\chi_R \, \mathcal{L}_R^\varepsilon f + \mathcal{L}(f^\varepsilon \chi_R)\right] d\nu \\
&= \int \left\{\chi_R \left[g + \mathcal{L}_R^\varepsilon f^\varepsilon - \mathcal{L} f^\varepsilon\right] + f^\varepsilon \mathcal{L}\chi_R + a(\nabla f^\varepsilon, \nabla \chi_R)\right\} d\nu \\
&\geq \int g \, d\nu - \sup_{t\in[0,T]} \|f_t^\varepsilon\|_{C_b^2(\mathbb{R}^d)} \int \left[\chi_R \left|a_R^\varepsilon - a\right| + \left|b_R^\varepsilon - b\right| + \left|\mathcal{L}\chi_R\right| + |a| \left|\nabla \chi_R\right|\right] d\,|\nu| \, .
\end{aligned}$$

As $\varepsilon \downarrow 0$, since $a_R = a$ and $b_R = b$ on $(0,T)\times B(0,R)$, the second integral converges to $\int \left[|\mathcal{L}\chi_R| + |a| \, |\nabla\chi_R|\right] d\,|\nu|$, and $\sup_{t\in[0,T]} \|f_t^\varepsilon\|_{C_b^2}$ is uniformly bounded in $\varepsilon > 0$, by (A.12). Finally, we let $R \to \infty$ and conclude, since $|\nabla\chi_R| + |\nabla\chi_R| \to 0$, pointwise and uniformly bounded. ☐

The superposition principle follows immediately from these two facts: since any weak solution $\nu = (\nu_t)_{t\in(0,T)}$ admits a narrowly continuous representative $\tilde{\nu}$, we let $\boldsymbol{\eta}$ be the solution of the MP associated to $\mathcal{L}(a,b)$, with $\bar{\nu} = \tilde{\nu}_0$ (Theorem A.6) and notice that the curve $\eta = (\eta_t)_{t\in[0,T]}$ is a narrowly continuous solution of the FPE associated to $\mathcal{L}$, with $\eta_0 = \tilde{\nu}_0$. By Theorem A.7, we conclude that $\eta_t = \tilde{\nu}_t$, for $t \in [0,T]$.

**Case of bounded coefficients.** We extend the validity of the superposition principle for diffusions with uniformly bounded coefficients: this already provides an extension of

[19, Theorem 2.6], as uniform bounds are imposed only with respect to $x \in \mathbb{R}^d$. Precisely, we assume that the coefficients $a$, $b$ satisfy

$$\int_0^T \sup_{x\in\mathbb{R}^d} |a_t(x)| + \sup_{x\in\mathbb{R}^d} |b_t(x)| dt < \infty. \tag{A.13}$$

*Step 1* (approximation). We argue by convolution with a kernel $\rho = a\exp(-\sqrt{1+|x|^2})$. For $\varepsilon \in (0,1)$, let $\rho^\varepsilon(x) = \varepsilon^{-n}\rho(x/\varepsilon)$ and notice that $\left|\nabla^i \rho^\varepsilon\right| \le C\varepsilon^{-2}\rho^\varepsilon$, for $i \in \{1,2\}$, where $C$ is some absolute constant. Then, $\nu^\varepsilon = \nu * \rho^\varepsilon$ solves a FPE with respect to a diffusion operator with coefficients $a^\varepsilon$, $b^\varepsilon$ satisfying (the correspondent of) (A.10), as a consequence of the last statement in Lemma A.1. Existence of superposition solutions $\boldsymbol{\eta}^\varepsilon \in \mathscr{P}(C([0,T];\mathbb{R}^d))$ for the associated martingale problems follows from the smooth case settled above.

*Step 2* (tightness). We notice first that, being $(\nu_0^\varepsilon)_{\varepsilon>0}$ a narrowly convergent sequence of probability measures (thus, it is also tight), there exists some increasing function $\theta : \mathbb{R} \to \mathbb{R}$ with $\lim_{z\to\infty}\theta(z) = \infty$ such that $\sup_{\varepsilon>0}\int \theta(|x|)d\nu_0^\varepsilon \le 1$. For $R \ge 1$, let then $\chi_R : \mathbb{R}^d \to [0,1]$ be the usual cut-off function and, for $i \in \{1,\ldots,d\}$, let $x_R^i(x) := x_i\chi_R \in \mathscr{A}$. We rely on the de la Vallée Poussin criterion, which improves the integral bound (A.13) to one of the form

$$\int_0^T \Theta\left(\sup_{x\in\mathbb{R}^d}|a_t(x)|\right) + \Theta\left(\sup_{x\in\mathbb{R}^d}|b_t(x)|\right) dt < \infty,$$

for some suitable convex non-decreasing function $\Theta : [0,\infty) \to [0,\infty)$ with faster than linear growth (but moderate, by Remark A.3). We apply Corollary A.5 to the solution $\eta^\varepsilon$ ($\varepsilon > 0$), with $f := x_R^i \circ e_t$ and $\theta$, $\Theta_1 := \Theta$, $\Theta_2 := \Theta$, to obtain some coercive functional $\Psi : C([0,T];\mathbb{R}) \to [0,\infty]$ (depending upon $\theta$ and $\Theta$ only) such that

$$\int \Psi(x_R^i \circ \gamma) d\boldsymbol{\eta}^\varepsilon(\gamma) \le \int \theta(|x_R^i|) d\eta_0^\varepsilon + \int_0^T \int \left(\Theta\left(\left|\mathcal{L}_t^\varepsilon x_R^i\right|\right) + \Theta\left(a_t^\varepsilon(\nabla x_R^i, \nabla x_R^i)\right)\right) d\eta_t^\varepsilon dt. \tag{A.14}$$

Since $\left\|\nabla x_R^i\right\|_\infty$ is uniformly bounded and $\left\|\nabla^2 x_R^i\right\|_\infty$ is infinitesimal as $R \to \infty$, we may let $R \to \infty$, and by lower-semicontinuity of $\Psi$, Fatou's lemma and Lebesgue dominated convergence theorem, we obtain a similar bound with the functions $x^i$ in place of $x_R^i$:

$$\int \Psi(\gamma^i) d\boldsymbol{\eta}^\varepsilon(\gamma) \le \int \theta(|x^i|) d\nu_0^\varepsilon(x) + \int_0^T \int \Theta\left(\left|(b^\varepsilon)_t^i\right|\right) + \Theta\left((a^\varepsilon)_t^{i,i}\right) d\nu_t^\varepsilon dt, \tag{A.15}$$

where we also make explicit the fact that $\eta_t^\varepsilon = \nu_t^\varepsilon$, for $t \in [0,T]$.

Inequality (A.2) and the assumptions on $\theta$ entails the uniform bounds for $\varepsilon > 0$

$$\begin{aligned}\int \Psi(\gamma^i) d\boldsymbol{\eta}^\varepsilon(\gamma) &\le 1 + \int_0^T \int \Theta\left(|b_t^i|\right) + \Theta\left(a_t^{i,i}\right) d\nu_t dt \\ &\le 1 + \int_0^T \Theta\left(\sup_{x\in\mathbb{R}^d}|b_t(x)|\right) + \Theta\left(\sup_{x\in\mathbb{R}^d}|a_t(x)|\right) dt\end{aligned}$$

Tightness follows since $\gamma \mapsto \sum_{i=1}^d \Psi(\gamma^i)$ is coercive in $C([0,T];\mathbb{R}^d)$.

*Step 3* (limit). This step is fully covered in Section A.3.

**Case of locally bounded coefficients.** Next, we assume that

$$\int_0^T \sup_{x\in B}\left[|a_t(x)| + |b_t(x)|\right] dt < \infty, \quad \text{for every bounded borel } B \subseteq \mathbb{R}^d. \tag{A.16}$$

and we prove the validity of the superposition principle for every weak solution $\nu =$

$(\nu_t)_{t\in(0,T)} \subseteq \mathscr{P}(\mathbb{R}^d)$ of the FPE (2.2) (recall that we also assume (2.3)).
*Step 1* (approximation). We approximate via push-forward by smooth maps. For $M \geq 1$, let $\chi_M$ be the usual cut-off function and let $\pi_M : \mathbb{R}^d \mapsto \mathbb{R}^d$ be the map

$$\pi_M(x) = x\chi_M(x), \quad \text{so that } \pi^i_M(x) = x^i\chi_M(x) \in C^2_c(\mathbb{R}^d).$$

By (A.16), it holds $\left|\mathcal{L}(\pi^i_M)\right| \leq \left\|\pi^i_M\right\|_{C^2} \sup_{|x|\leq 2M}\left[|a(x)| + |b(x)|\right]$ for $x \in \mathbb{R}^d$, $i \in \{1, \dots d\}$, and similarly $\left|a(\nabla\pi^i_M, \nabla\pi^j_M)\right| \leq \left\|\pi^i_M\right\|_{C^1} \sup_{|x|\leq 2M}|a(x)|$, for $x \in \mathbb{R}^d$, $i, j \in \{1, \dots d\}$.

Since conditional expectations reduce norms, we deduce that $\nu^M := \pi^M(\nu)$ solves a FPE associated to a diffusion on $\mathbb{R}^d$, whose coefficients $a^M$, $b^M$ satisfy (A.13): thus the previous argument gives superposition solutions $\boldsymbol{\eta}^M$.

*Step 2* (tightness). The argument is very similar to the previous case, but ultimately relies on inequality (A.1) instead of (A.2). Indeed, de la Vallée Poussin criterion improves the integral bound (2.3) to one of the form

$$\int_0^T \int \left[\Theta_1(|b|) + \Theta_2(|a|)\right] d\nu_t dt < \infty,$$

for suitable $\Theta_1$, $\Theta_2$ that fulfil the assumptions of Theorem A.2 (without loss of generality, $\Theta_2$ has moderate growth, by Remark A.3). With such a choice of $\Theta_1$, $\Theta_2$, and building $\theta$ as in the previous case, since $(\nu^M)_{M>0}$ is tight, we obtain some coercive functional $\Psi : C([0,T];\mathbb{R}) \to [0,\infty]$ (depending upon $\theta$, $\Theta_1$ and $\Theta_2$ only) such that the following inequality, completely analogous to (A.14), holds true:

$$\int \Psi(x^i_R \circ \gamma) d\boldsymbol{\eta}^M(\gamma) \leq \int \theta(\left|x^i_R\right|) d\eta^M_0 + \int_0^T \int \left(\Theta_1\left(\left|\mathcal{L}^M_t x^i_R\right|\right) + \Theta_2\left(a^M_t(\nabla x^i_R, \nabla x^i_R)\right)\right) d\eta^M_t dt,$$

for $i \in \{1, \dots, d\}$, and any $R > 0$.

Since $\left\|\nabla x^i_R\right\|_\infty$ is uniformly bounded and $\left\|\nabla^2 x^i_R\right\|_\infty$ is infinitesimal as $R \to \infty$, we may let $R \to \infty$, and by lower-semicontinuity of $\Psi$, Fatou's lemma and Lebesgue dominated convergence theorem, we obtain a similar bound with the functions $x^i$ in place of $x^i_R$:

$$\int \Psi(\gamma^i) d\boldsymbol{\eta}^M(\gamma) \leq \int \theta(\left|x^i\right|) d\nu^M_0(x) + \int_0^T \int \Theta_1\left(\left|(b^M)^i_t\right|\right) + \Theta_2\left((a^M)^{i,i}_t\right) d\nu^\varepsilon_t dt,$$

where we also make explicit the fact that $\eta^M_t = \nu^M_t$, for $t \in [0, T]$.

Jensen's inequality in the form (A.1) and the assumptions on $\theta$ entail the uniform bounds

$$\int \Psi(\gamma^i) d\boldsymbol{\eta}^M(\gamma) \leq 1 + \int_0^T \int \Theta_1\left(|b^i_t|\right) + \Theta_2\left(a^{i,i}_t\right) d\nu_t dt$$

Tightness follows once again because $\gamma \mapsto \sum_{i=1}^d \Psi(\gamma^i)$ is coercive on $C([0,T];\mathbb{R}^d)$.

*Step 3* (limit). This step is described in Section A.3.

**General case.** The final step consists in removing assumption (A.16).

*Step 1* (approximation). We perform once again an approximation via convolution, i.e., as in the case of uniformly bounded coefficients. To provide superposition solutions for the approximating $\nu^\varepsilon$, we use the fact that these are solutions to FPE's associated to diffusion operators whose coefficients are locally bounded.

*Step 2* (tightness). We argue exactly as in the previous cases, i.e. we exploit de la Vallée Poussin criterion to provide suitable $\Theta_1$, $\Theta_2$ (again with moderate growth without any loss of generality, by Remark A.3) and notice that bounds akin to (A.15), uniform in $\varepsilon > 0$, hold also in this case. By Corollary A.5, tightness follows also in this case.

*Step 3* (limit). This step is covered in Section A.3.

The proof of Theorem 2.5 is then completed. As already remarked at the beginning of this section, one could combine all the arguments above and prove it starting from the "base case" with a single combination of mollifications and push-forwards approximations. On a technical level, the main difficulty is to obtain the result for locally bounded coefficients, and this is done after we establish the result for uniformly bounded coefficients, regardless of their regularity, essentially because the push-forward approximation may not preserve it.

## References


[1] Luigi Ambrosio, *Transport equation and cauchy problem for $BV$ vector fields*, Invent. Math. **158** (2004), no. 2, 227–260. MR-2096794

[2] Luigi Ambrosio and Gianluca Crippa, *Existence, uniqueness, stability and differentiability properties of the flow associated to weakly differentiable vector fields*, Transport equations and multi-D hyperbolic conservation laws, Lect. Notes Unione Mat. Ital., vol. 5, Springer, Berlin, 2008, pp. 3–57. MR-2409676

[3] Luigi Ambrosio and Gianluca Crippa, *Continuity equations and ODE flows with non-smooth velocity*, Proc. Roy. Soc. Edinburgh Sect. A **144** (2014), no. 6, 1191–1244. MR-3283066

[4] Luigi Ambrosio and Alessio Figalli, *On flows associated to Sobolev vector fields in Wiener spaces: an approach à la DiPerna-Lions*, J. Funct. Anal. **256** (2009), no. 1, 179–214. MR-2475421

[5] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Gradient flows in metric spaces and in the space of probability measures*, second ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008. MR-2401600

[6] Luigi Ambrosio and Dario Trevisan, *Well-posedness of Lagrangian flows and continuity equations in metric measure spaces*, Anal. PDE **7** (2014), no. 5, 1179–1234. MR-3265963

[7] V. I. Bogachev, G. Da Prato, and M. Röckner, *On parabolic equations for measures*, Comm. Partial Differential Equations **33** (2008), no. 1, 397–418. MR-2398235

[8] V. I. Bogachev, G. Da Prato, M. Röckner, and W. Stannat, *Uniqueness of solutions to weak parabolic equations for measures*, Bull. Lond. Math. Soc. **39** (2007), no. 4, 631–640. MR-2346944

[9] V. I. Bogachev, M. Röckner, and S. V. Shaposhnikov, *On uniqueness problems related to the Fokker-Planck-Kolmogorov equation for measures*, J. Math. Sci. (N. Y.) **179** (2011), no. 1, 7–47, Problems in mathematical analysis. No. 61. MR-3014097

[10] V. I. Bogachev, M. Röckner, and S. V. Shaposhnikov, *Uniqueness problems for degenerate Fokker–Planck–Kolmogorov equations*, J. Math. Sci. **207** (2015-04-21), no. 2, 147–165.

[11] Vladimir I. Bogachev, Michael Röckner, and Stanislav V. Shaposhnikov, *On uniqueness of solutions to the cauchy problem for degenerate Fokker–Planck–Kolmogorov equations*, J. Evol. Equ. **13** (2013-05-07), no. 3, 577–593.

[12] François Bouchut and Gianluca Crippa, *Uniqueness, renormalization, and smooth approximations for linear transport equations*, SIAM J. Math. Anal. **38** (2006), no. 4, 1316–1328. MR-2274485

[13] Nicolas Bouleau and Francis Hirsch, *Dirichlet forms and analysis on Wiener space*, de Gruyter Studies in Mathematics, vol. 14, Walter de Gruyter & Co., Berlin, 1991. MR-1133391

[14] Gianluca Crippa and Camillo De Lellis, *Estimates and regularity results for the DiPerna-Lions flow*, J. Reine Angew. Math. **616** (2008), 15–46. MR-2369485

[15] G. Da Prato, F. Flandoli, E. Priola, and M. Röckner, *Strong uniqueness for stochastic evolution equations in hilbert spaces perturbed by a bounded measurable drift*, Ann. Probab. **41** (2013), no. 5, 3306–3344. MR-3127884

[16] R. J. DiPerna and P.-L. Lions, *Ordinary differential equations, transport theory and Sobolev spaces*, Invent. Math. **98** (1989), no. 3, 511–547. MR-1022305

[17] Stewart N. Ethier and Thomas G. Kurtz, *Markov processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1986, Characterization and convergence. MR-838085

[18] Shizan Fang, Dejun Luo, and Anton Thalmaier, *Stochastic differential equations with coefficients in Sobolev spaces*, J. Funct. Anal. **259** (2010), no. 5, 1129–1168. MR-2652184

[19] Alessio Figalli, *Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients*, J. Funct. Anal. **254** (2008), no. 1, 109–153. MR-2375067

[20] David Gilbarg and Neil S. Trudinger, *Elliptic partial differential equations of second order*, Classics in Mathematics, Springer-Verlag, Berlin, 2001, Reprint of the 1998 edition. MR-1814364

[21] N. V. Krylov, *On Kolmogorov's equations for finite-dimensional diffusions*, Stochastic PDE's and Kolmogorov equations in infinite dimensions (Cetraro, 1998), Lecture Notes in Math., vol. 1715, Springer, Berlin, 1999, pp. 1–63. MR-1731794

[22] N. V. Krylov and M. Röckner, *Strong solutions of stochastic equations with singular time dependent drift*, Probab. Theory Related Fields **131** (2005), no. 2, 154–196. MR-2117951

[23] Thomas G. Kurtz and Richard H. Stockbridge, *Existence of markov controls and characterization of optimal markov controls*, SIAM J. Control Optim. **36** (1998), no. 2, 609–653 (electronic). MR-1616514

[24] C. Le Bris and P.-L. Lions, *Existence and uniqueness of solutions to Fokker-Planck type equations with irregular coefficients*, Comm. Partial Differential Equations **33** (2008), no. 7, 1272–1317. MR-2450159

[25] E. Lenglart, D. Lépingle, and M. Pratelli, *Présentation unifiée de certaines inégalités de la théorie des martingales*, Seminar on Probability, XIV (Paris, 1978/1979) (French), Lecture Notes in Math., vol. 784, Springer, Berlin, 1980, With an appendix by Lenglart, pp. 26–52. MR-580107

[26] De Jun Luo, *Fokker-Planck type equations with Sobolev diffusion coefficients and BV drift coefficients*, Acta Math. Sin. (Engl. Ser.) **29** (2013), no. 2, 303–314. MR-3016531

[27] Emanuele Paolini and Eugene Stepanov, *Decomposition of acyclic normal currents in a metric space*, J. Funct. Anal. **263** (2012), no. 11, 3358–3390. MR-2984069

[28] Michael Röckner and Xicheng Zhang, *Weak uniqueness of Fokker-Planck equations with degenerate and bounded coefficients*, C. R. Math. Acad. Sci. Paris **348** (2010), no. 7, 435–438. MR-2607035

[29] R. E. Showalter, *Monotone operators in banach space and nonlinear partial differential equations*, Mathematical Surveys and Monographs, vol. 49, American Mathematical Society, Providence, RI, 1997. MR-1422252

[30] S. K. Smirnov, *Decomposition of solenoidal vector charges into elementary solenoids, and the structure of normal one-dimensional flows*, Algebra i Analiz **5** (1993), no. 4, 206–238. MR-1246427

[31] Daniel W. Stroock and S. R. Srinivasa Varadhan, *Multidimensional diffusion processes*, Classics in Mathematics, Springer-Verlag, Berlin, 2006, Reprint of the 1997 edition. MR-2190038

[32] A. Ju. Veretennikov, *Strong solutions and explicit formulas for solutions of stochastic integral equations*, Mat. Sb. (N.S.) **111(153)** (1980), no. 3, 434–452, 480. MR-568986

[33] M. Röckner V.I. Bogachev, N.V. Krylov and S.V. Shaposhnikov, *Fokker-Planck-Kolmogorov equations*, in preparation.

[34] Toshio Yamada and Shinzo Watanabe, *On the uniqueness of solutions of stochastic differential equations*, J. Math. Kyoto Univ. **11** (1971), 155–167. MR-0278420

[35] Xicheng Zhang, *Stochastic flows of SDEs with irregular coefficients and stochastic transport equations*, Bull. Sci. Math. **134** (2010), no. 4, 340–378. MR-2651896

[36] Xicheng Zhang, *Stochastic partial differential equations with unbounded and degenerate coefficients*, J. Differential Equations **250** (2011), no. 4, 1924–1966. MR-2763561

[37] Xicheng Zhang, *Degenerate irregular SDEs with jumps and application to integro-differential equations of Fokker-Planck type*, Electron. J. Probab. **18** (2013), no. 55, 25. MR-3065865