



**HAL**  
open science

# Probabilistic backward McKean numerical methods for PDEs and one application to energy management.

Lucas Izydorzcyk

► **To cite this version:**

Lucas Izydorzcyk. Probabilistic backward McKean numerical methods for PDEs and one application to energy management.. Probability [math.PR]. Institut Polytechnique de Paris; Università degli studi di Milano - Bicocca, 2021. English. NNT : 2021IPPAE008 . tel-04415024

**HAL Id: tel-04415024**

**<https://theses.hal.science/tel-04415024v1>**

Submitted on 24 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT  
POLYTECHNIQUE  
DE PARIS



NNT : 2021IPPAAE008

Thèse de doctorat

# Probabilistic backward McKean numerical methods for PDEs and one application to energy management.

PhD of Institut Polytechnique de Paris and Università di Milano Bicocca  
prepared at ENSTA Paris

École doctorale n° 574 Ecole Doctorale de Mathématiques  
Hadamard (EDMH)

Joint PhD Program in Mathematics Milano Bicocca - Pavia - INdAM

Specialty: Mathématiques appliquées, Dottorato in Matematica

PhD presented and discussed at Palaiseau, on July 6th 2021.

**LUCAS IZYDORCZYK**

## Members of the Committee.

Stefano BONACCORSI Università di Trento	Reviewer
Emmanuel GOBET Ecole Polytechnique, Institut Polytechnique de Paris	President
Christian LEONARD Université Paris Nanterre	Reviewer
Federica MASIERO Università di Milano Bicocca	Examiner
Stéphane MENOZZI Université d'Evry	Examiner
Nadia OUDJANE EDF R & D	Thesis co-supervisor
Francesco RUSSO ENSTA Paris, Institut Polytechnique de Paris	Thesis supervisor
Agnès SULEM INRIA Paris	Examiner
Gianmario TESSITORE Università di Milano Bicocca	Thesis co-supervisor

NNT: 2021IPPAE008

## THÈSE DE DOCTORAT (PhD)

École doctorale de mathématiques Hadamard (EDMH, ED 574)

and

## JOINT PhD PROGRAM IN MATHEMATICS

MILANO BICOCCA - PAVIA - INDAM

ENSTA Paris - Università di Milano Bicocca

*Specialty:* Mathématiques appliquées and Dottorato in Matematica

**Lucas IZYDORCZYK**

Probabilistic backward McKean numerical methods for PDEs and one application to energy management.

# Contents

<b>Introduction</b>	<b>1</b>
<b>1 Fokker-Planck equations with terminal condition and related McKean probabilistic representation</b>	<b>9</b>
1.1 Introduction . . . . .	9
1.2 Notations and preliminaries.....	12
1.3 A Fokker-Planck PDE with terminal condition.....	13
1.3.1 Preliminary results on uniqueness.....	13
1.3.2 Uniqueness: the case of Dirac initial conditions .....	16
1.3.3 Uniqueness: the case of bounded, non-degenerate coefficients.....	20
1.3.4 Uniqueness: the case of Ornstein-Uhlenbeck semigroup.....	23
1.4 McKean SDEs related to time-reversal of diffusions.....	28
1.4.1 Preliminary considerations .....	28
1.4.2 PDE with terminal condition and existence for the McKean SDE .....	30
1.4.3 PDE with terminal condition and uniqueness for the McKean SDE.....	32
1.4.4 Well-posedness for the McKean SDE: the bounded coefficients case.....	33
1.4.5 Well-posedness for the McKean SDE: the Ornstein-Uhlenbeck semigroup.....	37
1.4.6 Proof of Lemma 1.4.11 .....	40
<b>2 A fully backward representation of semilinear PDEs applied to the control of thermostatic loads in power systems</b>	<b>43</b>
2.1 Introduction .....	43
2.2 Notations .....	46
2.3 Representation of semilinear PDEs.....	47
2.3.1 Around two backward ODEs .....	47
2.3.2 The representation formula for a general semilinear PDE.....	49
2.4 Representation of stochastic control problems.....	57
2.5 A heuristic algorithm .....	61
2.6 Stochastic control of thermostatically controlled loads.....	65
2.6.1 Model description .....	65

2.6.2	Simulation results.....	67
2.7	Appendix.....	69
2.7.1	A sufficient condition to obtain an equivalent probability.....	69
2.7.2	Proof of the local Lipschitz property of the cost functional $J$ .....	70
2.7.3	A simplified version of the envelope theorem.....	71
<b>3</b>	<b>McKean Feynman-Kac probabilistic representations of non-linear partial differential equations</b>	<b>73</b>
3.1	Introduction and motivations.....	73
3.1.1	General considerations.....	73
3.1.2	Some motivating examples.....	75
3.1.3	Structure of the paper.....	76
3.2	McKean representations of non-linear Fokker-Planck equations.....	77
3.2.1	Probabilistic representation of linear Fokker-Planck equations.....	77
3.2.2	McKean probabilistic representation of non-linear Fokker-Planck equation.....	77
3.3	McKean Feynman-Kac representations for non-conservative and non-linear PDEs.....	79
3.4	McKean representation of a Fokker-Planck equation with terminal condition.....	83
3.5	Probabilistic representation with jumps for non-conservative PDEs.....	85
3.6	McKean SDEs in random environment.....	88
3.6.1	The (S)PDE and the basic idea.....	88
3.6.2	Well-posedness of the SPDE.....	88
3.6.3	McKean equation in random environment.....	89
3.7	McKean representation of stochastic control problems.....	90
3.7.1	Stochastic control problems and non-linear Partial Differential Equations.....	90
3.7.2	McKean type representation in a toy control problem example.....	92

# Introduction

One of the purposes of stochastic analysis is to establish connections between deterministic possibly non-linear evolution problems and stochastic processes. The advantage one can take of these connections lies in the possibility to treat well-posedness and regularity problems for Partial Differential Equation (PDE) with probabilistic methods. Furthermore, such a correspondence can shed a new light on the numerical approximation of solutions, in particular when dealing with equations on possibly high dimensional spaces.

In the specific case of linear second order parabolic Cauchy problems set on  $[0, T] \times \mathbb{R}^d$  for a given time horizon  $T > 0$  and a given state space dimension  $d \in \mathbb{N}^*$ , the probabilistic objects that arise are solutions of Markovian Stochastic Differential Equations (SDE for short). Given two functions  $b, \sigma$  defined on  $[0, T] \times \mathbb{R}^d$  being respectively  $\mathbb{R}^d$ -valued and matrix-valued, we focus on two types of PDEs: the *forward Kolmogorov* (or *Fokker-Planck*) equation

$$\begin{cases} \partial_t \mathbf{u}(t) = \frac{1}{2} \sum_{i,j=1}^d \partial_{ij} \sigma \sigma^T(t, \cdot)_{ij} \mathbf{u}(t) - \sum_{i=1}^d \partial_i (b(t, \cdot)_i \mathbf{u}(t)), \quad t \in ]0, T[ \\ \mathbf{u}(0) = \nu, \end{cases} \quad (0.1)$$

where  $\nu$  is a prescribed measure and the *backward Kolmogorov* equation

$$\begin{cases} \partial_t v(t, \cdot) + \frac{1}{2} \sum_{i,j=1}^d \sigma \sigma^T(t, \cdot)_{ij} \partial_{ij} v(t, \cdot) + \sum_{i=1}^d b(t, \cdot)_i \partial_i v(t, \cdot) = 0, \quad t \in [0, T[ \\ v(T, \cdot) = g, \end{cases} \quad (0.2)$$

where  $g$  is a prescribed function from  $\mathbb{R}^d$  to  $\mathbb{R}$ . It is known that PDEs (0.1) and (0.2) can be represented using solutions  $X^{s,x}$ , whenever they exist, of the SDE

$$X_t^{s,x} = x + \int_s^t b(r, X_r^{s,x}) ds + \int_s^t \sigma(r, X_r^{s,x}) dW_r, \quad t \in [0, T], \quad (0.3)$$

for all  $(s, x) \in [0, T] \times \mathbb{R}^d$ , where  $W$  is a Brownian motion.

More precisely, the function  $\mathbf{u} : t \mapsto \int_{\mathbb{R}^d} \mathbb{L}aw_{X_t^{0,x}} \nu(dx)$  provides a measure-valued solution of

(0.1). The simplest case  $b \equiv 0, \sigma \equiv I_d$  expresses the classical link between the Brownian motion and the *heat* equation. Equations of type (0.1) are *conservative* PDEs, in the sense that the *mass* function  $t \mapsto \int_{\mathbb{R}^d} \mathbf{u}(t)(dx)$  is constant.

In addition, PDEs of type (0.2) are generally represented by the so called *Feynman-Kac* formula. In particular, if  $v$  is a smooth solution of (0.2) then

$$v(s, x) = \mathbf{E} \ g \ X_T^{s,x} \ , (s, x) \in [0, T] \times \mathbb{R}^d.$$

It is possible to get probabilistic representations of non-linear versions of previous PDEs for which other probabilistic objects arise. Among many others, two classes of representation processes are of particular importance.

1. Solutions of *McKean* SDEs, namely SDEs in which the coefficients depend not only on time and position of the *particle*, but also on the (marginal) laws of the solution. Such objects are primarily tailored to provide solutions of non-linear versions of PDE (0.1), see [85]. In Chapter 3, we will survey how this representation works, with significant extensions to a large class of non-conservative PDEs.
2. Solutions of *Forward Backward Stochastic Differential Equation* (FBSDE for short). An FBSDE is a system of equations, including a forward SDE and a stochastic differential equation with prescribed terminal condition, i.e. a backward SDE (BSDE), in the sense of Pardoux and Peng, see e.g. [91]. A solution of a BSDE is in general a couple of progressively measurable processes. In this introduction we will only mention the case when the (Markovian) forward SDE is of type (0.3), i.e. the case when the SDE and BSDE are decoupled. The data concerning the BSDE (with terminal condition) are in general the *driver*  $F : [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}$  and the *terminal* function  $g$ . More precisely, solving the (Markovian) BSDE with driver  $F$  and terminal function  $g$  consists in finding a pair of progressively measurable processes  $(Y^{s,x}, Z^{s,x})$  such that

$$Y_t^{s,x} = g \ X_T^{s,x} + \int_t^T F(r, X_r^{s,x}, Y_r^{s,x}, Z_r^{s,x}) dr - \int_t^T Z_r^{s,x} dW_r, t \in [s, T], \quad (0.4)$$

for each  $(s, x) \in [0, T] \times \mathbb{R}^d$ .

These equations are probabilistic tools adapted to represent *viscosity* solutions of semi-linear generalizations of PDE (0.2), i.e.

$$\begin{cases} \square \\ \partial_t v(t, \cdot) + \frac{1}{2} \sum_{i,j=1}^d \sigma \sigma^T(t, \cdot)_{ij} \partial_{ij} v(t, \cdot) + \sum_{i=1}^d b(t, \cdot)_i \partial_i v(t, \cdot) + F(t, \cdot, v(t, \cdot), \sigma(t, \cdot)^T \nabla_x v(t, \cdot)) = 0 \\ \square \\ v(T, \cdot) = g \end{cases} \quad (0.5)$$

for each  $t \in [0, T]$ . Under general assumptions, the function  $v := (s, x) \mapsto Y_s^{s,x}$  provides a viscosity solution of (0.5), see for example [92].

If  $\sigma$  is suitably non-degenerate, an important feature of FBSDEs of type (0.3)-(0.4) is the possibility one has to change the forward process  $X$  (as the solution of an SDE with different drift  $b$ ) without modifying the function  $v := (s, x) \mapsto Y_s^{s,x}$ , at the cost of substituting the driver

$$F := (t, x, v, z) \mapsto F(t, x, v, z) - \sigma^{-1}(t, x) \cdot (b - b(t, x), z) \ .$$

What precedes means that there are infinitely many FBSDE representations for a given semi-linear PDE. At the theoretical level, these representations are purely equivalent. At the numerical level, the choice of an ad hoc drift is crucial. Indeed, usual numerical methods to solve FBSDEs lie on *forward backward* schemes. Those schemes consist in simulating paths of the chosen forward process and using these paths to approximate, backward in time, conditional expectations by least squares Monte Carlo, see e.g. [50]. The FBSDE solution is provided by those conditional expectations which are estimated by Monte-Carlo regressions. The forward process paths constitute an *estimation grid*. When the dimension  $d$  of the state space increases, one needs to simulate and to store possibly a huge number of trajectories of the forward process to obtain a good approximation of the solution. Besides, to circumvent the curse of dimension, there is here a need of guiding this process in order to explore efficiently the state space.

A PDE for which the choice of the drift guiding the estimation grid is at first sight natural is the *Hamilton-Jacobi-Bellman* (HJB for short) equation, when it is semi-linear. In its generality, the HJB equation is a possibly fully non-linear PDE, which constitutes an important tool to study *stochastic control problems*. This equation reads

$$\begin{aligned} \square \partial_t v(t, \cdot) + H(t, \cdot, v(t, \cdot), \nabla_x v(t, \cdot), \nabla_x^2 v(t, \cdot)) &= 0 \\ \square v(T, \cdot) &= g, \end{aligned} \tag{0.6}$$

where the so called *Hamiltonian*  $H$  has the form

$$H : (t, x, \delta, \gamma) \mapsto \inf_{a \in A} \left[ f(t, x, a) + \sum_{i=1}^d b(t, x, a)_i \delta_i + \frac{1}{2} \sum_{i,j=1}^d \sigma \sigma^T(t, x, a)_{ij} \gamma_{ij} \right], \tag{0.7}$$

with  $A \subseteq \mathbb{R}^k$  for a given  $k \in \mathbb{N}^*$ . Equation (0.6) is in fact the dynamic programming equation of the stochastic control problem whose value function is

$$v : (t, x) \rightarrow \inf_{\alpha} \mathbb{E} \left[ \int_t^T f(s, X_s^{t,x,\alpha}, \alpha_s) ds + g(X_T^{t,x,\alpha}) \right], \tag{0.8}$$

where  $X^{t,x,\alpha}$  is a solution, whenever it exists, of the following *controlled SDE*

$$dX_t = b(t, X_t, \alpha_t) dt + \sigma(t, X_t, \alpha_t) dW_t, \quad t \in [0, T], \tag{0.9}$$

starting at time  $t \in [0, T]$  with value  $x \in \mathbb{R}^d$ , given an  $A$ -valued progressively measurable control process  $\alpha$ .

Under general conditions,  $v$  is the unique viscosity solution of (0.6), see for example [108, 42, 94].

Equation (0.6) becomes a semi-linear PDE when the control does not intervene in the function  $\sigma$ . By previous considerations, the viscosity solutions can be represented using FBSDEs. In that context, the natural choice of the drift mentioned earlier can be guided by the stochastic control problem.

Our idea is to select a drift, which localizes the forward process in regions visited by optimally controlled trajectories. Hence, our candidate estimation grid in terms of computational costs is the



optimally controlled process. Nonetheless, this process cannot be simulated forwardly since optimal controls are discovered backwardly in time.

This observation suggests the following strategy to use effectively the optimally controlled process as an estimation grid. If a priori informations are available on its terminal law, one proceeds to simulate it backwardly in time. This will allow to bypass memory problems and dimensionality issues. Doing this, we expect that the number of paths to approximate the value function will drastically decrease.

Coming back to the case of a general semilinear PDE of type (0.5), once an ad hoc drift has been chosen to guide the forward process of an FBSDE, a natural way to take into account the terminal condition  $g$ , consists in simulating backwardly in time the resulting forward process. At the theoretical level, this comes down to derive *fully backward* representations of semi-linear PDEs, namely FBSDEs (Markovian) representations in which the chosen forward process  $\xi$  evolves backwardly in time. Even though the drift is fixed, the law of the forward process  $X$  is not yet determined, since no initial condition  $X_0$  is fixed. According to previous objective,  $\xi$  will be the time-reversal of some Markov process  $X$ . Time-reversal of Markov processes has been explored by several authors, see for instance [55] for the diffusion case in finite dimension, [45] for the diffusion case in infinite dimension and [64] for the jump case. We also mention the two very interesting recent preprints [31, 32] in relation with entropy.

Given a diffusion process  $X$  with coefficients  $b, \sigma$ , [55] provides sufficient conditions on the coefficients and on the marginal densities (whenever they exist)  $(q_t)_{t \in [0, T]}$  of  $X$  to ensure that  $X := X_{T-\cdot}$  is also a diffusion process. Moreover, it is shown that  $X$  is a solution (in law) of the SDE

$$X_t = X_{T-\cdot} - \int_0^t b(T-r, X_r) dr + \int_0^t \frac{\text{div}_y \sigma \sigma_i^{-T}(T-r, X_r) q_{T-r}(X_r)}{q_{T-r}(X_r)} \mathbf{1}_{i \in [1, d]} dr + \int_0^t \sigma(T-r, X_r) d\beta_r, \quad (0.10)$$

where  $\beta$  is a Brownian motion.

To simulate the process  $X$ , (0.10) is not usable in practice. Indeed, even choosing properly the law of  $X_T$  compatibly with  $g$ , (0.10) involves the unknown marginal laws of  $X$ .

An idea to avoid this limitation consists in substituting Equation (0.10) into a McKean SDE, where the marginal laws appearing in the time-reversal dynamics are no longer exogenous parameters but a part of the equation solution, i.e.

$$\begin{aligned} \xi_t &= \xi_0 - \int_0^t b(T-r, \xi_r) dr + \int_0^t \frac{\text{div}_y \sigma \sigma_i^{-T}(T-r, \xi_r) p_r(\xi_r)}{p_r(\xi_r)} \mathbf{1}_{i \in [1, d]} dr + \int_0^t \sigma(T-r, \xi_r) d\beta_r, \\ \xi_0 &\sim \nu \\ p_t &\text{ density of } \xi_t, t \in ]0, T[ \end{aligned} \quad (0.11)$$

where  $\nu$  is a prescribed probability measure, whose solution is the couple  $(\xi, \mathbf{p})$  where  $\mathbf{p}$  is the probability-valued function such that for all  $t \in ]0, T[$ ,  $\mathbf{p}_t$  has density  $p_t$ . By [55], the couple  $X, \mathbf{q}_{T-\cdot}$

is a solution of (0.11) when  $\nu$  is the law of  $X_T$ . Provided that the McKean SDE (0.11) is well-posed (in particular it admits a unique solution), the unique solution  $\xi$  will be the researched process  $X$ . In this case (0.11) governs intrinsically the dynamics of  $X$ .

Beyond the scope of time-reversal of diffusions and stochastic control problems, the study of well-posedness for Equation (0.11) has its own interest. Indeed, it turns out that it can be seen as a probabilistic representation of the following *Fokker-Planck* type PDE with terminal condition

$$\begin{aligned} \square \quad \partial_t \mathbf{u}(t) &= \frac{1}{2} \sum_{i,j=1}^d \partial_{ij} \sigma \sigma^T(t, \cdot)_{ij} \mathbf{u}(t) - \sum_{i=1}^d \partial_i (b(t, \cdot)_i \mathbf{u}(t)), \quad t \in [0, T[ \\ \square \quad \mathbf{u}(T) &= \nu. \end{aligned} \tag{0.12}$$

Final value problems as (0.12) arise in many applications involving *inverse* problems, especially in physical sciences, as heat conduction [14], material science [95] or hydrology [5]. In those applications modeled by a diffusion, the terminal configuration of a physical system is observed, so the uniqueness (and stability) issue is more important than the existence problem: the goal is to determine admissible initial configurations. The possible lack of uniqueness and stability for PDEs with given terminal condition, has been handled in the literature especially by regularization techniques, see e.g [107, 74]. To the best of our knowledge, a very few uniqueness and approximation results are available for the general equation (0.12) and implemented methods are purely analytic. On the other hand, a probabilistic representation provides a new light to the study of (0.12) and its approximation.

In Chapter 1, uniqueness problem for PDE (0.12) is addressed with probabilistic and analytical methods. Furthermore, we investigate connections between PDE (0.12) and McKean SDE (0.11) and use these connections to derive well-posedness results for the McKean SDE. It turns out that uniqueness for (0.12) is a key tool to study the McKean SDE as evidenced by Proposition 1.3.2 which states that if  $(\xi, \mathbf{p})$  solves (0.11) then  $\mathbf{p}_{T-}$  solves PDE (0.12). More precisely, the contributions of the Chapter 1 are twofold.

1. Uniqueness for PDE (0.12) is investigated in various classes of coefficients  $b, \sigma$  and different classes of measure-valued solutions. We produce essentially three kinds of results.

Concerning the first one, uniqueness for (0.12) is proved among solutions with initial values in a given class  $\mathbf{C}$  of finite measures, supposing the forward Fokker-Planck PDE (0.1) admits at most one solution for initial values in  $\mathbf{C}$ . In particular, we cover uniqueness among non-negative finite measure-valued solutions with initial value in the class of multiples of a Dirac mass. We obtain results in the case of Lipschitz possibly degenerate coefficients and in the case of bounded, non degenerate possibly irregular coefficients. We use here purely probabilistic methods. This is the object of Theorems 1.3.9 and 1.3.10.

Concerning the second one, uniqueness is proved in the wider class of finite measure-valued solutions without supposing uniqueness for the forward PDE. We treat the general case of time-homogeneous, bounded, Hölder continuous and non degenerate coefficients with an extension

to the piecewise time-homogeneous case. We use here an analytical method lying on a crucial property of the semigroup generated by the diffusion operator associated to coefficients  $b, \sigma$ . This is the object of Theorem 1.3.13 and Corollary 1.3.16.

Finally, we deal with the situation of a PDE associated to a time-inhomogeneous *Ornstein-Uhlenbeck* semigroup. Loosely speaking, the method consists here in performing a change of variable in the initial PDE that reduces to the case of the *backward heat* equation for which uniqueness is easily derived by the use of Fourier transform as in the forward case, see Theorem 1.3.19.

2. We derive well-posedness results for the McKean SDE (0.11). We begin by linking separately existence and uniqueness for PDE (0.12) and SDE (0.11). Then, we discuss existence and uniqueness in law for (0.11) in the case of bounded regular coefficients, see Theorem 1.4.15. We finally establish strong existence and pathwise uniqueness result in the Ornstein-Uhlenbeck case, see Theorem 1.4.19.

Back to stochastic control and to the considerations after (0.9), well-posedness results obtained in Chapter 1 give a theoretical starting point to derive a fully backward FBSDE representation of the HJB equation involving the time-reversal of a suitable approximation of the optimally controlled process. Nonetheless major challenges arise both at the theoretical and numerical levels.

1. At the theoretical level, we would need to find a suitable class of drifts approaching reasonably the optimally controlled drift and allowing to implement the McKean time-reversal techniques of Chapter 1.
2. At the numerical level, we would need to approach the optimally controlled process. Following our McKean time reversal methods, the appearance of marginal densities in the McKean problem is an important obstacle. The cost coming from their (non-parametric) estimation annihilates the gain obtained in terms of memory and dimension, see e.g. [103]. Secondly one would need to have an information about the terminal law. We expect that the extrapolation of such an information highly depends on the nature of the optimization problem, especially on the terminal cost.

A suitable class of drifts allowing us to perform previous tasks is the one of affine functions. In Chapter 2, we investigate the choice of a time-dependent affine drift  $b : (s, x) \rightarrow a(s)x + c(s)$  where  $a$  (resp.  $c$ ) is a piecewise continuous function on  $[0, T]$  with values in  $M_d(\mathbb{R})$  (resp.  $\mathbb{R}^d$ ).

As a first step, we derive a fully backward representation for general semi-linear PDEs. of the form

$$\begin{aligned} \square & \quad \partial_t v(t, \cdot) + \frac{1}{2} \sum_{i,j=1}^d \sigma \sigma^T(t)_{ij} \partial_{ij} v(t, \cdot) + H(t, \cdot, v(t, \cdot), \nabla_x v(t, \cdot)) = 0, \quad t \in [0, T[ \\ \square & \quad v(T, \cdot) = g. \end{aligned} \tag{0.13}$$

where  $\sigma$  is a continuous deterministic matrix-valued function.

Our concept of *fully backward* representation mentioned before, is based on a process  $\xi$  running backwardly with respect to the time  $t$  of the PDE (0.13).

To define  $\xi$ , we fix a Gaussian Borel measure  $\nu$  which will be the distribution of  $\xi_0$ . Let  $\bar{m}^\nu$  (resp.  $\bar{Q}^\nu$ ) denote the mean (resp. the covariance matrix) of the measure  $\nu$ . The process  $\xi$  will be the unique strong solution of

$$\xi_t = \xi_0 - \int_0^t b(T-r, \xi_r) + \sigma\sigma^T(T-r)Q(T-r)^{-1}(\xi_r - m(T-r))dr + \int_0^t \sigma(T-r)d\beta_r, t \in [0, T], \tag{0.14}$$

where  $\beta$  is a Brownian motion,  $t \mapsto m(T-t) = E(\xi_t)$  and  $t \mapsto Q(T-t)$ , the covariance matrix of  $\xi_t$ . Indeed, it is not difficult to show previous well-posedness and the fact that  $m : [0, T] \mapsto R^d$ ,  $Q : [0, T] \mapsto S_d(R)$  are respectively solutions of the following backward ODEs

$$\begin{aligned} \square \frac{d}{dt}m(t) &= a(t)m(t) + c(t), t \in [0, T] \\ \square m(T) &= \bar{m}^\nu, \end{aligned} \tag{0.15}$$

$$\begin{aligned} \square \frac{d}{dt}Q(t) &= Q(t)a(t)^T + a(t)Q(t) + \sigma\sigma^T(t), t \in [0, T] \\ \square Q & \end{aligned} \tag{0.16}$$

provided that  $Q(0)$  is positive semi-definite.

Under general assumptions on  $H, g$ , Theorem 2.3.10 states that a function  $v : [0, T] \times R^d \mapsto R$  being continuous in time and continuously differentiable in space with polynomial growth gradient is a viscosity solution of (0.13) if and only if for all  $t \in [0, T]$

$$v(t, \xi_t) = E \int_t^T H(s, \xi_s, v(s, \xi_s), \nabla_x v(s, \xi_s) - b(s, \xi_s), \nabla_x v(s, \xi_s)) ds + g(\xi_T, \xi_t), \tag{0.17}$$

where  $\xi := \xi_{T \dots}$ .

The fully backward representation of (0.13) consists in the coupling (0.14)-(0.17). We remark that Lemma 2.3.6 says that  $\xi$  is in fact the time reversal of a solution of an SDE of the type

$$dX_t = b(t, X_t)dt + \sigma(t)dW_t, \tag{0.18}$$

without knowing a priori its initial law.

In Corollary 2.4.6, this result is applied to the case when the semi-linear PDE (0.13) is an HJB equation associated with a control problem when  $\sigma$  is a time-dependent continuous function. In that case, the coupling (0.14)-(0.17). becomes a representation formula for the value function of the aforementioned control problem. This leads us to suggest a heuristic algorithm to solve this control problem, i.e. to approximate its value function and also determine an approximated optimal strategy.

Our approach consists in determining a process  $\xi$  solving (0.14), for which the affine drift coefficient  $\tilde{b}$  (and therefore  $a$  and  $c$ ) is discovered simultaneously as the value function, in such a way that  $\tilde{b}$  is determined via a linear regression, at each time step, to be as close as possible to the drift of the optimally controlled process. The resulting process  $\xi$  will be our estimation grid, which is generated backwardly and is supposed to explore efficiently the regions of interest of our problem. Indeed, that grid is *adaptive* in the sense that it exploits as it is generated, informations on the value function estimations that were already performed. Therefore, the method meets empirically our initial objective of parsimonious state space exploration. It also has the advantage to spare memory since there is no need to store the grid. Finally, the accuracy of the procedure is illustrated in the context of a stochastic control problem arising in energy management.

In Chapter 3, we survey the use of McKean SDEs to represent solutions of non conservative (and non linear) extensions of Equation (0.1). More precisely, we consider *perturbed* versions of (0.1) of the form

$$\begin{aligned} \square \quad \partial_t u(t, \cdot) &= \frac{1}{2} \sum_{i,j=1}^d \partial_{ij} \sigma \sigma^T(t, \cdot, u(t, \cdot))_{ij} u(t, \cdot) - \sum_{i=1}^d \partial_i (b(t, \cdot, u(t, \cdot))_i u(t, \cdot)) \\ &+ \Lambda(t, \cdot, u(t, \cdot), \nabla_x u(t, \cdot)) u(t, \cdot), \quad t \in ]0, T], \\ \square \quad u(0, \cdot) &= v \end{aligned} \tag{0.19}$$

where  $\Lambda : [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}$ . The probabilistic object that arises to represent (0.19) are solutions of the following McKean type equation

$$\begin{aligned} \square \quad Y_t &= Y_0 + \int_0^t \sigma(s, Y_s, u(s, Y_s)) dW_s + \int_0^t b(s, Y_s, u(s, Y_s)) ds \\ Y_0 &\sim \nu \\ \square \quad \int \phi(x) u(t, x) dx &= \mathbf{E} \left\{ \phi(Y_t) \exp \left( \int_0^t \Lambda(s, Y_s, u(s, Y_s), \nabla u(s, Y_s)) ds \right) \right\}, \quad t \in ]0, T], \end{aligned} \tag{0.20}$$

for any continuous bounded real valued test function  $\phi$ . The second line appearing in (0.20), is denominated the *linking equation* and can be seen as a *path-dependent* extension of the dependence of the law encountered when dealing with classical McKean SDEs (namely when  $\Lambda = 0$ ). This opens the way to new numerical methods based on *interacting particle systems*. We expose those methods and give some classes of nonconservative PDEs for which previous representation is of interest.

# Chapter 1

## Fokker-Planck equations with terminal condition and related McKean probabilistic representation

This chapter is the object of the paper [62].

### 1.1 Introduction

The main objective of the paper consists in studying well-posedness and probabilistic representation of the Fokker-Planck PDE with terminal condition

$$\begin{aligned} \square \quad \partial_t \mathbf{u} &= \frac{1}{2} \sum_{i,j=1}^d \partial_{ij}^2 (\sigma \sigma^T)_{ij}(t, x) \mathbf{u} - \operatorname{div} (b(t, x) \mathbf{u}) \\ \square \quad \mathbf{u}(T) &= \mu, \end{aligned} \tag{1.1}$$

where  $\sigma : [0, T] \times \mathbb{R}^d \rightarrow M_{d,m}(\mathbb{R})$ ,  $b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\mu$  is a prescribed finite Borel measure on  $\mathbb{R}^d$ . When  $\mathbf{u}(t)$  admits a density for some  $t \in [0, T]$  we write  $\mathbf{u}(t) = u(t, x)dx$ . This equation is motivated by applications in various domains of physical sciences and engineering, as heat conduction [14], material science [95] or hydrology [5]. In particular, *hydraulic inversion* is interested in inverting a diffusion process representing the concentration of a pollutant to identify the pollution source location when the final concentration profile is observed. Those models are often formulated by PDE problems which are in general ill-posed because, either the solution is not unique or the solution is not stable. For this issue, the existence is ensured by the fact that the observed contaminant is necessarily originated from some place at a given time (as soon as the model is correct). Several authors have handled the lack of uniqueness problem by introducing regularization methods approaching the problem by well-posed PDEs, see typically [107] and [74]. A second issue, when the problem is well-approximated by a regularized problem, consists in providing a numerical approx-

imating scheme to the backward diffusion process. In particular for (1.1) there are very few results even concerning existence and uniqueness.

Our point of view is that a probabilistic representation of (1.1) can bring new insights to the treatment of the two mentioned issues: well-posedness and numerical approximation. To realize this objective we consider the renormalized PDE

$$\begin{aligned} \square \quad \partial_t \bar{\mathbf{u}} &= \frac{1}{2} \sum_{i,j=1}^d (\sigma \sigma^T)_{i,j}(t, x) \bar{\mathbf{u}} - \operatorname{div}(b(t, x) \bar{\mathbf{u}}) \\ \square \quad \bar{\mathbf{u}}(T) &= \bar{\mu}, \end{aligned} \tag{1.2}$$

where  $\bar{\mu} = \frac{\mu}{\mu(\mathbb{R}^d)}$  is a probability measure. We remark that the PDEs (1.2) and (1.1) are equivalent in the sense that a solution (1.2) (resp. (1.1)) provides a solution to the other one. The program consists in considering the McKean type stochastic differential equation (SDE)

$$\begin{aligned} \square \quad Y_t &= Y_0 - \int_0^t b(T-r, Y_r) dr + \int_0^t \frac{\operatorname{div}_y(\sum_i (T-r, Y_r) p_r(Y_r))}{p_r(Y_r)} dr + \int_0^t \sigma(T-r, Y_r) d\beta_r, \\ \square \quad p_t &\text{ density law of } \mathbf{p}_t = \text{law of } Y_t, t \in ]0, T[, \\ \square \quad Y_0 &\sim \bar{\mu}, \end{aligned} \tag{1.3}$$

where  $\beta$  is a  $m$ -dimensional Brownian motion and  $\Sigma = \sigma \sigma^T$ , whose solution is the couple  $(Y, \mathbf{p})$ . Indeed an application of Itô formula (see Proposition 1.4.3) shows that whenever  $(Y, \mathbf{p})$  is a solution of (1.3) then  $t \mapsto \mathbf{p}_{T-t}$  is a solution of (1.2).

The idea of considering (1.3) comes from the SDE verified by time-reversal of a diffusion. Time-reversal of Markov processes was explored by several authors: see for instance [55] for the diffusion case in finite dimension, [45] for the diffusion case in infinite dimension and [64] for the jump case.

Consider a *forward* diffusion process  $X$  solution of

$$X_t = X_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s, \quad t \in [0, T], \tag{1.4}$$

where  $\sigma$  and  $b$  are Lipschitz coefficients with linear growth and  $W$  is a standard Brownian motion on  $\mathbb{R}^m$ .  $\hat{X}_t := X_{T-t}, t \in [0, T]$  will denote the time-reversal process. In [55] the authors gave sufficient general conditions on  $\sigma, b$  and the marginal laws  $p_t$  of  $X_t$  so that  $Y := \hat{X}$  is a solution (in law) of the SDE

$$Y_t = X_T - \int_0^t b(T-r, Y_r) dr + \int_0^t \frac{\operatorname{div}_y(\sum_i (T-r, Y_r) p_{T-r}(Y_r))}{p_{T-r}(Y_r)} dr + \int_0^t \sigma(T-r, Y_r) d\beta_r. \tag{1.5}$$

The key idea to show well-posedness of the McKean SDE (1.3), is the study of uniqueness of the PDE (1.2) (or (1.1)). For instance, the trivial case of the heat equation with terminal condition produces uniqueness. Suppose indeed that  $u : [0, T] \mapsto \mathcal{S}(\mathbb{R}^d)$  solves

$$\begin{aligned} \square \quad \partial_t \mathbf{u} &= \Delta \mathbf{u} \\ \square \quad \mathbf{u}(T) &= \mu. \end{aligned} \tag{1.6}$$

Then, the Fourier transform of  $u, v(t, \cdot) := \mathbf{F}u(t, \cdot), t \in [0, T]$  solves the ODE (for fixed  $\xi \in \mathbb{R}^d$ )

$$\begin{aligned} \square \frac{d}{dt}v(t, \xi) &= -|\xi|^2 v(t, \xi), (t, \xi) \in [0, T] \times \mathbb{R}^d \\ \square v(T, \cdot) &= \mathbf{F}\mu. \end{aligned} \tag{1.7}$$

This admits at most one solution, since setting  $\mathbf{F}\mu = 0$  the unique solution of (1.7) is the null function.

Another relatively simple situation is described below to study uniqueness among the solutions of (1.2) starting in the class of Dirac measures. Suppose for a moment that the PDE in the first line of (1.2), but with initial condition (see (3.2)) is well-posed. Sufficient conditions for this will be provided in Remark 1.3.3. Let  $x \in \mathbb{R}^d$  and  $u$  be a solution of (1.2) such that  $u(0, \cdot) = \delta_x$ . If  $X^x$  is the solution of (1.4) with initial condition  $x$ , it is well-known that the family of laws of  $X^x_t, t \in [0, T]$ , is a solution of (1.2). So this coincides with  $u(t, \cdot)$  and in particular  $\bar{\mu}$  is the law of  $X^x_T$ . To conclude we only need to determine  $x$ .

Consider the example when  $\sigma$  is continuous bounded non-degenerate and the drift  $b$  is affine i.e.  $b(s, y) = b_0(s) + b_1(s)y, (s, y) \in [0, T] \times \mathbb{R}^d, b_0$  (resp.  $b_1$ ) being mappings from  $[0, T]$  to  $\mathbb{R}^d$  (resp. to  $M_d(\mathbb{R})$ ). Taking the expectation in the SDE fulfilled by  $X^x$ , we show that the function  $t \mapsto E^x(t) := \mathbf{E}(X^x_t)$  is solution of

$$E^x(t) = \int_{\mathbb{R}^d} y \mu(dy) - \int_t^T (b_0(s) + b_1(s)E^x(s)) ds.$$

Previous linear ODE has clearly a unique solution. At this point  $x = E(0)$  is determined.

Those examples give a flavor of how to tackle the well-posedness issue. However, generalizing those approaches is far more complicated and constitutes the first part of the present work. The contributions of the paper are twofold.

1. We investigate uniqueness for the Fokker-Planck PDE with terminal condition (1.2). This is done in Section 1.3 in two different situations: the case when the coefficients are bounded and the situation of a PDE associated with an inhomogeneous Ornstein-Uhlenbeck semigroup. In Section 1.3.3 we show uniqueness when the coefficients are stepwise time-homogeneous. In Theorem 1.3.13 the coefficients are time-homogeneous, bounded and Hölder, with non-degenerate diffusion. Corollary 1.3.16 extends previous results to the case of stepwise time-inhomogeneous coefficients. In Section 1.3.4, Theorem 1.3.19 treats the Ornstein-Uhlenbeck case. In Section 1.3.2 we show uniqueness for bounded continuous coefficients for solutions starting in the class  $\mathcal{C}$  of multiples of Dirac measures. In Proposition 1.3.9 we discuss the framework of dimension  $d = 1$ . Theorem 1.3.10 is devoted to the case  $d \geq 2$ . We distinguish the non-degenerate case from the possibly degenerate case but with smooth coefficients: we prove uniqueness for small time horizon  $T$ .
2. We study existence and uniqueness in law for the McKean SDE (1.3), with some specific remarks concerning strong existence and pathwise uniqueness. We differentiate specifically between existence and uniqueness. After some preliminary considerations in Section 1.4.1, Sec-



tions 1.4.2 and 1.4.3 link the well-posedness of the PDE (1.2) to the well-posedness of the McKean SDE (1.3). In particular Proposition 1.4.6 (resp. Corollary 1.4.9) links the existence (resp. uniqueness) of (1.2) with (1.3). In Section 1.4.4, Proposition 1.4.14 and Theorem 1.4.16 discuss the case of bounded coefficients. Theorem 1.4.19 in Section 1.4.5 is devoted to the case of Ornstein-Uhlenbeck (with not necessarily Gaussian terminal condition), where strong existence and pathwise uniqueness are established.

## 1.2 Notations and preliminaries

Let us fix  $d, m \in \mathbb{N}^*, T > 0$ .  $C_c^\infty(\mathbb{R}^d)$  is the linear space of smooth functions with compact support. For a given  $p \in \mathbb{N}^*$ ,  $[[1, p]]$  denotes the set of all integers between 1 and  $p$  included.  $M_{d,m}(\mathbb{R})$  stands for the set of  $d \times m$  matrices. If  $d = m$ , we simply use the notation  $M_d(\mathbb{R})$ . For a given  $A \in M_d(\mathbb{R})$ ,  $Tr(A)$  (resp.  $A^T$ ) symbolizes the trace (resp. the transpose) of the matrix  $A$ .  $\|A\|$  denotes the usual Frobenius norm.  $\cdot, \cdot$  denotes the usual scalar product on  $\mathbb{R}^d$ , with associated norm  $|\cdot|$ . For a given  $f : \mathbb{R}^p \rightarrow \mathbb{R}^l, p, l \in \mathbb{N}^*$ ,  $\partial_j f^i, (i, j) \in [[1, l]] \times [[1, p]]$  denote the partial derivatives of  $f$  being defined in the sense of distributions on  $\mathbb{R}^p$  whenever they exist. We also introduce the mapping  $Jf$  from  $\mathbb{R}^p$  to  $M_{l,p}(\mathbb{R})$  such that  $Jf : z \mapsto (\partial_j f^i(z))_{(i,j) \in [[1, l]] \times [[1, p]]}$ .

Let  $\alpha \in ]0, 1[, n \in \mathbb{N}$ .  $C_b(\mathbb{R}^d)$  (resp.  $\mathbb{C}^n(\mathbb{R}^d)$ ) indicates the space of bounded continuous functions (resp. bounded functions of class  $C^n$  such that all the derivatives are bounded).  $C^\alpha(\mathbb{R}^d)$  is the Banach space of bounded  $\alpha$ -Hölder functions  $\mathbb{R}^d \rightarrow \mathbb{R}$  equipped with the norm  $\|\cdot\|_\alpha := \|\cdot\|_\infty + [\cdot]_\alpha$ , where

$$[f]_\alpha := \sup_{x, y \in \mathbb{R}^d, x \neq y} \frac{|f(x) - f(y)|}{|x - y|^\alpha} < \infty.$$

If  $n$  is some integer  $C^{\alpha+n}(\mathbb{R}^d)$  is the Banach space of bounded functions  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  such that all its derivatives up to order  $n$  are bounded and such that the derivatives of order  $n$  are  $\alpha$ -Hölder continuous. This is equipped with the norm obtained as the sum of the  $C_b^n(\mathbb{R}^d)$ -norm plus the sum of the quantities  $[g]_\alpha$  where  $g$  is an  $n$ -order derivative of  $f$ . For more details, see Section 0.2 of [82]. If  $E$  is a linear Banach space, we denote by  $\|\cdot\|_E$  the associated operator norm and by  $L(E)$  the space of linear bounded operators  $E \rightarrow E$ . Often in the sequel we will have  $E = C^{2\alpha}(\mathbb{R}^d)$ .

$\mathbb{M}^d$  (resp.  $\mathbb{M}_+^d, \mathbb{M}_f^d, \mathbb{M}_b^d$ ) denotes the set of probability measures (resp. non-negative finite valued measures, finite signed measures) on  $\mathbb{R}^d, \mathbb{R}^d$ . We also denote by  $\mathbb{S}^d$  the space of Schwartz functions and by  $\mathbb{S}'^d$  the space of tempered distributions. For all  $\varphi \in \mathbb{S}^d$  and  $\mu \in \mathbb{M}_f^d$ , we set the notations

$$F\varphi : \xi \mapsto \int_{\mathbb{R}^d} e^{-i \xi \cdot x} \varphi(x) dx, F\mu : \xi \mapsto \int_{\mathbb{R}^d} e^{-i \xi \cdot x} \mu(dx).$$

Given a mapping  $\mathbf{u} : [0, T] \rightarrow \mathbb{M}_f^d$ , we convene that when for  $t \in [0, T]$ ,  $\mathbf{u}(t)$  has a density, this is denoted by  $u(t, \cdot)$ . We also introduce, for a given  $t$  in  $[0, T]$ , the differential operator,

$$L_t f := \frac{1}{2} \sum_{i,j=1}^d \Sigma_{ij}(t, \cdot) \partial_{ij} f + \sum_{i=1}^d b_i(t, \cdot) \partial_i f, \tag{2.1}$$

$f \in C^2(\mathbb{R}^d)$  and denote by  $L_t^*$  its formal adjoint, which means that for a given signed measure  $\eta$

$$L_t^* \eta := \frac{1}{2} \sum_{i,j=1}^d \partial_{ij}^2 (\Sigma_{ij}(t, x) \eta) - \text{div} (b(t, x) \eta) . \tag{2.2}$$

With this notation, equation (1.1) rewrites

$$\begin{aligned} \square \partial_t \mathbf{u} &= L_t^* \mathbf{u} \\ \square \mathbf{u}(T) &= \mu. \end{aligned} \tag{2.3}$$

In the sequel we will often make use of the following assumptions.

**Assumption 1.**  $b, \sigma$  are Lipschitz in space uniformly in time, with linear growth.

**Assumption 2.**  $b$  and  $\sigma$  are bounded and  $\Sigma$  is continuous.

**Assumption 3.** There exists  $E > 0$  such that for all  $t \in [0, T], \xi \in \mathbb{R}^d, x \in \mathbb{R}^d$

$$\Sigma(t, x) \xi, \xi \geq E |\xi|^2. \tag{2.4}$$

For a given random variable  $X$  on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ ,  $L_{\mathbb{P}}(X)$  denotes its law under  $\mathbb{P}$  and  $E_{\mathbb{P}}(X)$  its expectation under  $\mathbb{P}$ . When self-explanatory, the subscript will be omitted in the sequel.

### 1.3 A Fokker-Planck PDE with terminal condition

#### 1.3.1 Preliminary results on uniqueness

In this section, we consider a Fokker-Planck type PDE with terminal condition for which the notion of solution is clarified in the following definition.

**Definition 1.3.1.** Fix  $\mu \in \mathcal{M}_+(\mathbb{R}^d)$ . We say that a mapping  $\mathbf{u}$  from  $[0, T]$  to  $\mathcal{M}_+(\mathbb{R}^d)$  solves the PDE (1.1), if for all  $\varphi \in C_c^\infty(\mathbb{R}^d)$  and  $t \in [0, T]$

$$\int_{\mathbb{R}^d} \varphi(y) \mathbf{u}(t)(dy) = \int_{\mathbb{R}^d} \varphi(y) \mu(dy) - \int_t^T \int_{\mathbb{R}^d} L_s \varphi(y) \mathbf{u}(s)(dy) ds. \tag{3.1}$$

We consider the following assumption related to a given class  $\mathcal{C} \subseteq \mathcal{M}_+(\mathbb{R}^d)$ . Later we will establish uniqueness results for (1.1) provided that the solution starts in  $\mathcal{C}$ .

**Assumption 4.** For all  $\nu \in \mathcal{C}$ , the PDE

$$\begin{aligned} \square \partial_t \mathbf{u} &= L_t^* \mathbf{u} \\ \square \mathbf{u}(0) &= \nu \end{aligned} \tag{3.2}$$

admits at most one solution  $\mathbf{u} : [0, T] \rightarrow \mathcal{M}_+(\mathbb{R}^d)$ .

We recall that, for a given  $\nu \in \mathcal{M}_+(\mathbb{R}^d)$ ,  $\mathbf{u} : [0, T] \times \mathbb{R}^d \rightarrow \mathcal{M}_+(\mathbb{R}^d)$  is a solution of the PDE (3.2) if for all  $\varphi \in C_c^\infty(\mathbb{R}^d)$  and all  $t \in [0, T]$ ,  $\mathbf{M}$

$$\int_{\mathbb{R}^d} \varphi(y) \mathbf{u}(t)(dy) = \int_{\mathbb{R}^d} \varphi(y) \nu(dy) + \int_0^t \int_{\mathbb{R}^d} L_s \varphi(y) \mathbf{u}(s)(dy) ds. \tag{3.3}$$

Suppose there is an  $\mathbf{M}_+$ -valued solution of (3.2)  $\mathbf{u}$  and Assumption 4 with respect to some class  $\mathcal{C}$  holds and such that  $\mathbf{u}(0) \in \mathcal{C}$ . Then this unique solution will be denoted by  $\mathbf{u}^\nu$  in the sequel. We remark that, whenever Assumption 4 holds with respect to a given  $\mathcal{C} \subset \mathcal{M}_+(\mathbb{R}^d)$  then (3.2) admits at most one  $\mathbf{M}_+$ -valued solution with any initial value belonging to  $\mathcal{C} := (\alpha\nu)_{\alpha>0, \nu \in \mathcal{C}}$ .

We start with a simple but fundamental observation.

**Proposition 1.3.2.** *Let us suppose  $\sigma, b$  to be locally bounded,  $\nu$  be a Borel probability on  $\mathbb{R}^d$ ,  $\alpha > 0$ ,  $\xi$  be a r.v. distributed according to  $\nu$ . Suppose that there is a solution  $X$  of SDE*

$$X_t = \xi + \int_0^t b(r, X_r) dr + \int_0^t \sigma(r, X_r) dW_r, \quad t \in [0, T], \quad \mathbf{P}\text{-a.s.}, \tag{3.4}$$

where  $W$  is an  $m$ -dimensional standard Brownian motion. Then the  $\mathbf{M}_+$ -valued function  $t \mapsto \alpha \mathbf{L}(X_t)$  is a solution of (3.2) with initial value  $\alpha\nu$ .

*Proof.* One first applies Itô formula to  $\phi(X_t)$ , where  $\phi$  is a smooth function with compact support and then one takes the expectation. □

**Remark 1.3.3.** 1. Suppose that the coefficients  $b, \Sigma$  are bounded. Assumption 4 holds with respect to  $\mathcal{C} := \mathbf{M}_+(\mathbb{R}^d)$  as soon as the martingale problem associated with  $b, \Sigma$  admits uniqueness for all initial condition of the type  $\delta_x, x \in \mathbb{R}^d$ . Indeed, this is a consequence of Lemma 2.3 in [44].

2. Suppose  $b$  and  $\sigma$  with linear growth. Let  $\nu \in \mathcal{M}_+(\mathbb{R}^d)$  not vanishing (resp.  $\nu \in \mathcal{M}_+(\mathbb{R}^d)$ ). The existence of a  $\mathbf{M}_+$ -valued (resp.  $\mathbf{P}$ -valued) solution for (3.2) (even on  $t \geq 0$ ) is ensured when the martingale problem associated to  $b, \Sigma$  admits existence (and consequently when the SDE (3.4) admits weak existence) with initial condition  $\nu$  (resp.  $\frac{\nu}{|\nu|}$ ). This follows by Proposition 1.3.2. We remark that, for example, this happens when the coefficients  $b, \sigma$  are continuous with linear growth: see Theorem 12.2.3 in [104] for the case of bounded coefficients, the general case can be easily obtained by truncation.

3. The martingale problem associated to  $b, \Sigma$  is well-posed for all deterministic initial condition, for instance in the following cases.

- When  $\Sigma, b$  have linear growth and  $\Sigma$  is continuous and non-degenerate, i.e. Assumption 3, see [104] Corollary 7.1.7 and Theorem 10.2.2.
- Suppose  $d = 1$  and  $\sigma$  is bounded. When  $\sigma$  is lower bounded by a positive constant on each compact set, see [104], Exercise 7.3.3.
- When  $d = 2$ ,  $\Sigma$  is non-degenerate and  $\sigma$  and  $b$  are time-homogeneous and bounded, see [104], Exercise 7.3.4.

- When  $\sigma, b$  are Lipschitz with linear growth (with respect to the space variable), in which case we have even strong solutions of the corresponding stochastic differential equation.

**Lemma 1.3.4.** *Let  $T > 0$  be arbitrary and  $\nu \in \mathcal{M}_+(\mathbb{R}^d)$ . We suppose the validity of Assumptions 2 and 3. Then there is a unique  $\mathcal{M}_+(\mathbb{R}^d)$ -valued solution  $\mathbf{u}^\nu$  to (3.2) with  $\mathbf{u}^\nu(0) = \nu$ . Moreover  $\mathbf{u}^\nu$  takes values in  $\mathcal{P}(\mathbb{R}^d)$ .*

*Proof.* Existence follows by items 2. and 3. of Remark 1.3.3. Uniqueness is a consequence of items 1. and 3. of the same Remark. □

Below we give two uniqueness results for the PDE (1.2).

**Proposition 1.3.5.** *Suppose Assumption 4 holds with respect to a given  $\mathcal{C} \subseteq \mathcal{M}_+(\mathbb{R}^d)$ . Suppose that for all  $\nu \in \mathcal{C}$  there exists an  $\mathcal{M}_+(\mathbb{R}^d)$ -valued solution of (3.2) with initial value  $\nu$ . Then, the following properties are equivalent.*

1. *The mapping from  $\mathcal{C}$  to  $\mathcal{M}_+(\mathbb{R}^d)$   $\nu \mapsto \mathbf{u}^\nu(T)$  is injective.*
2. *For all  $\mu \in \mathcal{M}_+(\mathbb{R}^d)$ , the PDE (2.3) with terminal value  $\mu$  admits at most a solution in the sense of Definition 1.3.1 among all  $\mathcal{M}_+(\mathbb{R}^d)$ -valued solutions starting in the class  $\mathcal{C}$ .*

*Proof.* Concerning the converse implication, consider  $(\nu, \nu^i) \in \mathcal{C}^2$  such that  $\mathbf{u}^\nu(T) = \mathbf{u}^{\nu^i}(T)$  and suppose that uniqueness holds for equation (2.3) for all terminal values in  $\mathcal{M}_+(\mathbb{R}^d)$  in the sense of Definition 1.3.1 among non-negative measure-valued solutions starting in the class  $\mathcal{C}$ . We remark that  $\mathbf{u}^\nu, \mathbf{u}^{\nu^i}$  are such solutions and are associated to the same terminal value. Uniqueness gives  $\mathbf{u}^\nu = \mathbf{u}^{\nu^i}$  and in particular  $\nu = \nu^i$ .

Concerning the direct implication, consider  $\mathbf{u}^1, \mathbf{u}^2$  two non-negative measure-valued solutions of equation (1.2) in the sense of Definition 1.3.1, with the same terminal value in  $\mathcal{M}_+(\mathbb{R}^d)$ , such that  $\mathbf{u}^i(0), i \in \{1, 2\}$ , belong to  $\mathcal{C}$  and suppose that  $\nu \mapsto \mathbf{u}^\nu(T)$  is injective from  $\mathcal{C}$  to  $\mathcal{M}_+(\mathbb{R}^d)$ . Setting  $\nu^i := \mathbf{u}^i(0)$ , we remark that for a given  $i \in \{1, 2\}$

$$\begin{aligned} \square \partial_t \mathbf{u}^i &= L \sharp \mathbf{u}^i \\ \square \mathbf{u}^i(0) &= \nu^i, \end{aligned} \tag{3.5}$$

in the sense of identity (3.3). Then, the fact  $\mathbf{u}^1(T) = \mathbf{u}^2(T)$  gives  $\mathbf{u}^{\nu^1}(T) = \mathbf{u}^{\nu^2}(T)$ . By injectivity  $\nu^1 = \nu^2$  and the result follows by Assumption 4. □

Proceeding in the same way as for the proof of Proposition 1.3.5 we obtain the following.

**Proposition 1.3.6.** *Suppose that for all  $\nu \in \mathcal{M}_f(\mathbb{R}^d)$ , there exists a unique solution  $\mathbf{u}^\nu : [0, T] \rightarrow \mathcal{M}_f(\mathbb{R}^d)$  of (3.2) with initial value  $\nu$ . Then, the following properties are equivalent.*

1. *The mapping  $\nu \mapsto \mathbf{u}^\nu(T)$  is injective.*
2. *For all  $\mu \in \mathcal{M}_f(\mathbb{R}^d)$ , the PDE (1.1) with terminal value  $\mu$  admits at most a solution in the sense of Definition 1.3.1.*

- Remark 1.3.7.** 1. Suppose that the coefficients  $\Sigma, b$  are bounded. Then, any measure-valued solution  $\mathbf{u} : [0, T] \rightarrow \mathbf{M}_+(\mathbb{R}^d)$  of (3.2) such that  $\mathbf{u}(0) \in \mathbf{P}(\mathbb{R}^d)$  takes values in  $\mathbf{P}(\mathbb{R}^d)$ . Indeed, this can be shown approaching the function  $\phi \equiv 1$  from below by smooth functions with compact support.
2. Replacing  $\mathbf{M}_+(\mathbb{R}^d)$  with  $\mathbf{P}(\mathbb{R}^d)$  in Assumption 4, item 2. in Proposition 1.3.5 can be stated also replacing  $\mathbf{M}_+(\mathbb{R}^d)$  with  $\mathbf{P}(\mathbb{R}^d)$ .

### 1.3.2 Uniqueness: the case of Dirac initial conditions

In this section we give examples of functions  $b, \sigma$  for which uniqueness of (2.3) among  $\mathbf{M}_+(\mathbb{R}^d)$ -valued solutions is ensured, supposing Assumption 4 is in force with respect to  $\mathbf{C} := (\alpha\delta_x)_{\alpha>0, x \in \mathbb{R}^d}$ .

**Remark 1.3.8.** Let  $\alpha \geq 0$ . Let  $x \in \mathbb{R}^d$ . Suppose that there is a solution  $X^x$  of SDE (3.4) with  $\xi = x$ .

- By Proposition 1.3.2, the  $\mathbf{M}_+(\mathbb{R}^d)$ -valued mapping  $t \mapsto \alpha \mathbf{L}(X_t^x)$  is a solution of (3.2) with initial value  $\alpha\delta_x$ .
- $t \mapsto \alpha \mathbf{L}(X_t^x)$  can be identified with  $\mathbf{u}^{\alpha\delta_x}$  and in particular  $\int_{\mathbb{R}^d} \mathbf{u}^{\alpha\delta_x}(t)(dy) = \alpha, \forall t \in [0, T]$ .

If Assumption 1 holds,  $X^x$  denotes the unique solution of equation (3.4) with initial value  $x \in \mathbb{R}^d$ . We start with the case of dimension  $d = m = 1$ .

**Proposition 1.3.9.** Suppose the validity of Assumption 4 with  $\mathbf{C} = (\alpha\delta_x)_{\alpha>0, x \in \mathbb{R}}$  and 1 with  $d = m = 1$ . Then, for all  $\mu \in \mathbf{M}_+(\mathbb{R})$ , equation (1.2) with terminal value  $\mu$  admits at most one solution in the sense of Definition 1.3.1 among the  $\mathbf{M}_+(\mathbb{R})$ -valued solutions starting in  $\mathbf{C}$ .

*Proof.* Fix  $(x, y) \in \mathbb{R}^2$  and  $\alpha, \beta \geq 0$  such that

$$\mathbf{u}^{\alpha\delta_x}(T) = \mathbf{u}^{\beta\delta_y}(T). \quad (3.6)$$

It suffices to show that  $\alpha = \beta$  and  $x = y$  to conclude, thanks to Proposition 1.3.5. By item 2. of Remark 1.3.8, we have  $\alpha = \beta$  and consequently  $\mathbf{L}_P(X_t^x) = \mathbf{L}_P(X_t^y)$ . In particular  $\mathbf{E}(X_t^x) = \mathbf{E}(X_t^y)$ . Since  $b, \sigma$  are Lipschitz in space, they have bounded derivatives in the sense of distributions that we denote by  $\partial_x b$  and  $\partial_x \sigma$ .

Set  $Z^{x,y} := X^y - X^x$ . We have

$$Z_t^{x,y} = (y - x) + \int_0^t b_s^{x,y} Z_s^{x,y} ds + \int_0^t \sigma_s^{x,y} Z_s^{x,y} dW_s, \forall t \in [0, T], \quad (3.7)$$

where for a given  $s \in [0, T]$

$$b_s^{x,y} = \int_0^1 \partial_x b(s, aX_s^y + (1-a)X_s^x) da, \quad \sigma_s^{x,y} = \int_0^1 \partial_x \sigma(s, aX_s^y + (1-a)X_s^x) da.$$

The unique solution of (3.7) is well-known to be

$$Z^{x,y} = \exp \left( \int_0^{\cdot} b_s^{x,y} ds \right) \mathbf{E} \left[ \int_0^{\cdot} \sigma_s^{x,y} dW_s \right] (y - x),$$

where  $E(\cdot)$  denotes the Doléans exponential. Finally, we have

$$E \exp \int_0^T b_s^{x,y} ds = E \exp \int_0^T \sigma_s^{x,y} dW_s \quad (y - x) = 0.$$

Since the quantity appearing in the expectation is strictly positive, we conclude  $x = y$ .  $\square$

We continue now with a discussion concerning the multidimensional case  $d \geq 2$ . The uniqueness result below only holds when the time-horizon is small enough. Later, in Section 1.3.3 we will present in a framework of piecewise time-homogeneous coefficients results which are valid for any time-horizon. Theorem 1.3.10 distinguishes two cases: the first one with regular possibly degenerate coefficients, the second one with non-degenerate possibly irregular coefficients.

**Theorem 1.3.10.** *We suppose Assumption 4 with  $C = (\alpha \delta_x)_{\alpha > 0, x \in \mathbb{R}^d}$  and the validity of either item (a) or (b) below.*

(a) Assumption 1.

(b) Assumptions 2 and 3.

There is  $T > 0$  small enough such that the following holds. For all  $\mu \in \mathbb{R}^d$ , equation (1.2) admits at most one solution in the sense of Definition 1.3.1 among the  $\mathbf{M}_+ \mathbb{R}^d$ -valued solutions starting in  $C$ .

The proof of item (a) of Theorem 1.3.10 relies on a basic lemma of moments estimation.

**Lemma 1.3.11.** *We suppose Assumption 1. Let  $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ . Then,  $\sup_{t \in [0, T]} E |X_t^x - X_t^y|^2 \leq |y - x|^2 e^{KT}$ , with  $K := 2K^b + \sum_{j=1}^m K^{\sigma_j}$ , where*

$$K^b := \sup_{s \in [0, T]} \| \|Jb(s, \cdot)\| \|_\infty$$

and for all  $j \in \llbracket 1, m \rrbracket$

$$K^{\sigma_j} := \sup_{s \in [0, T]} \| \|J\sigma_j(s, \cdot)\| \|_\infty.$$

**Proof** (of Lemma 1.3.11).

For a given  $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$  we set

$$Z_t^{x,y} := X_t^y - X_t^x, t \in [0, T].$$

We have

$$Z_t^{x,y} = y - x + \int_0^t B_r^{x,y} Z_r^{x,y} dr + \int_0^t \sum_{j=1}^m C_r^{x,y,j} Z_r^{x,y} dW^j, \quad t \in [0, T], \quad (3.8)$$

with, for all  $r \in [0, T]$

$$B_r^{x,y} := \int_0^1 Jb(r, aX_r^y + (1-a)X_r^x) da, \quad C_r^{x,y,j} := \int_0^1 J\sigma_j(r, aX_r^y + (1-a)X_r^x) da, \quad \forall j \in \llbracket 1, m \rrbracket.$$

By the classical existence and uniqueness theorem for SDEs with Lipschitz coefficients we know that

$$\mathbf{E}(\sup_{s \leq T} |X_s^z|^2) < \infty, \tag{3.9}$$

for all  $z \in \mathbf{R}^d$ . This implies

$$\mathbf{E}(\sup_{t \in [0, T]} |Z_t^{x,y}|^2) < \infty. \tag{3.10}$$

Now, Itô's formula gives, for all  $t \in [0, T]$

$$|Z_t^{x,y}|^2 = |y - x|^2 + 2 \int_0^t B_r^{x,y} Z_r^{x,y} dr + \sum_{j=1}^m \int_0^t C_r^{x,y,j} Z_r^{x,y}{}^2 dr + 2 \sum_{i=1}^d M_t^{x,y,i}, \tag{3.11}$$

where, for a given  $i \in \llbracket 1, d \rrbracket$ ,  $M^{x,y,i}$  denotes the local martingale  $\int_0^\cdot Z_s^{x,y,i} \sum_{j=1}^m C_s^{x,y,j} Z_s^{x,y} dW_s^j$ .

Consequently, for all  $i \in \llbracket 1, d \rrbracket$ , we have

$$\begin{aligned} [M^{x,y,i}]_T &= \sum_{j=1}^m \int_0^T Z_r^{x,y,i} C_r^{x,y,j} Z_r^{x,y}{}^2 dr, \\ &\leq \sum_{j=1}^m \int_0^T C_r^{x,y,j} Z_r^{x,y}{}^2 |Z_r^{x,y}|^2 dr, \\ &\leq T \sum_{j=1}^m (K^{\sigma_j})^2 \sup_{r \in [0, T]} |Z_r^{x,y}|^2. \end{aligned} \tag{3.12}$$

By the latter inequality and (3.10), we know that  $\mathbf{E} [M^{x,y,i}]_T^{\frac{1}{2}} < \infty$ , so for all  $i \in \llbracket 1, d \rrbracket$ ,  $M^{x,y,i}$  is a true martingale. Taking expectation in identity (3.11), we obtain

$$\mathbf{E} |Z_t^{x,y}|^2 = |y - x|^2 + \int_0^t \mathbf{E} \sum_{k=1}^m 2 B_r^{x,y} Z_r^{x,y} + \sum_{k=1}^m C_r^{x,y,k} Z_r^{x,y}{}^2 dr.$$

Hence, thanks to Cauchy-Schwarz inequality and to the definition of  $K^b$  and  $K^{\sigma_j}$  for all  $j \in \llbracket 1, m \rrbracket$

$$\mathbf{E} |Z_t^{x,y}|^2 \leq |y - x|^2 + K \int_0^t \mathbf{E} |Z_r^{x,y}|^2 dr$$

and we conclude via Gronwall's Lemma. □

**Proof** (of Theorem 1.3.10).

Fix  $(x_1, x_2) \in \mathbf{R}^d \times \mathbf{R}^d$ ,  $\alpha, \beta \geq 0$  such that

$$\mathbf{u}^{\alpha \delta x_1}(T) = \mathbf{u}^{\beta \delta x_2}(T). \tag{3.13}$$

To conclude, it suffices to show  $\alpha = \beta$  and  $x_1 = x_2$  thanks to Proposition 1.3.5.

1. We suppose first Assumption 1. Once again, item 2. of Remark 1.3.8 gives  $\alpha = \beta$  and

$$\mathbb{E} X_T^{x_1} = \mathbb{E} X_T^{x_2} . \tag{3.14}$$

Adopting the same notations as in the proof of Lemma 1.3.11, a similar argument as in (3.12), together with (3.10) allow to show that the local martingale part of  $Z^{x_1, x_2} = X^{x_2} - X^{x_1}$  defined in (3.8) is a true martingale. So, taking the expectation in (3.12) with  $x = x_1, y = x_2$ , by Lemma 1.3.11 we obtain

$$\begin{aligned} \mathbb{E} X_T^{x_2} - X_T^{x_1} - (x_2 - x_1) &\leq K_b \int_0^T \mathbb{E} |X_r^{x_2} - X_r^{x_1}| dr \\ &\leq K_b \int_0^T \mathbb{E} (|X_r^{x_2} - X_r^{x_1}|) dr \\ &\leq \frac{K}{2} T e^{\frac{K}{2} T} |x_2 - x_1|. \end{aligned}$$

Remembering (3.14), this implies

$$1 - \frac{K}{2} T e^{\frac{K}{2} T} |x_2 - x_1| \leq 0.$$

Taking  $T$  such that  $\frac{K}{2} T < M$  with  $M e^{\frac{K}{2} T} < 1$ , we have  $1 - \frac{K}{2} T e^{\frac{K}{2} T} > 0$ , which implies  $|x_2 - x_1| = 0$ .

2. We suppose here Assumptions 2 and 3. Firstly, point 1. of Theorem 1. in [112] ensures the existence of probability spaces  $\Omega^i, \mathbb{F}^i, \mathbb{P}^i, i \in \{1, 2\}$  on which are defined respectively two  $m$ -dimensional Brownian motions  $W^1, W^2$  and two processes  $X^1, X^2$  such that

$$X_t^i = x_i + \int_0^t b_s X_s^i ds + \int_0^t \sigma_s X_s^i dW_s^i \mathbb{P}^i\text{-a.s.}, t \in [0, T].$$

Once again, item 2. of Remark 1.3.8 implies  $\alpha_1 = \alpha_2$  and

$$\mathbb{L}_{\mathbb{P}^1} X_T^1 = \mathbb{L}_{\mathbb{P}^2} X_T^2 . \tag{3.15}$$

Secondly, point b. of Theorem 3 in [112] shows that for every given bounded  $D \subset \mathbb{R}^d$ , for all  $\varphi : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  belonging to  $W_p^{1,2}([0, T] \times D)$  (see Definition of that space in [112]) for a given  $p > d + 2$ , we have for all  $t \in [0, T], i \in \{1, 2\}$ ,

$$\varphi(t, X_t^i) = \varphi(0, x_i) + \int_0^t (\partial_t + L_s) \varphi(s, X_s^i) ds + \int_0^t J \varphi(s, X_s^i) \sigma_s X_s^i dW_s^i \mathbb{P}^i\text{-a.s.} \tag{3.16}$$

where the application of  $\partial_t + L_t, t \in [0, T]$  has to be understood componentwise.

Thirdly, Theorem 2. in [112] shows that if  $T$  is sufficiently small, then the system of  $d$  PDEs

$$\forall (t, x) \in [0, T] \times \mathbb{R}^d, \begin{cases} \partial_t \varphi(t, x) + L_t \varphi(t, x) = 0, \\ \varphi(T, x) = x, \end{cases} \tag{3.17}$$



admits a solution  $\varphi$  in  $W_p^{1,2}([0, T] \times D)$  for all  $p > 1$  and all bounded  $D \subset \mathbb{R}^d$ . Moreover the partial derivatives in space of  $\varphi$  are bounded (in particular  $J\varphi$  is bounded) and  $\varphi(t, \cdot)$  is injective for all  $t \in [0, T]$ .

Combining now (3.17) with identity (3.16), we observe that  $\varphi(\cdot, X^i)$ ,  $i \in \{1, 2\}$ , are local martingales. Using additionally the fact that  $J\varphi$  and  $\sigma$  are bounded, it is easy to show that they are true martingales. Taking the expectation in (3.16) with respect to  $\mathbb{P}^i$ ,  $i = 1, 2$ , gives

$$\varphi(0, x_i) = \mathbf{E}_{\mathbb{P}^i}[\varphi(T, X_T^i)], \quad i \in \{1, 2\}.$$

In parallel, identity (3.15) gives

$$\mathbf{E}_{\mathbb{P}^1}[\varphi(T, X_T^1)] = \mathbf{E}_{\mathbb{P}^2}[\varphi(T, X_T^2)].$$

So,  $\varphi(0, x_1) = \varphi(0, x_2)$ . We conclude that  $x_1 = x_2$  since  $\varphi(0, \cdot)$  is injective. □

### 1.3.3 Uniqueness: the case of bounded, non-degenerate coefficients

In this section we consider the case of time-homogeneous, bounded and Hölder coefficients in dimension  $d \geq 1$ . We suppose that Assumption 3 holds and consider the following one.

**Assumption 5.** 1.  $b, \sigma$  are time-homogeneous and bounded.

2. For all  $(i, j) \in \llbracket 1, d \rrbracket^2$ ,  $b_i, \Sigma_{ij} \in C^{2\alpha}(\mathbb{R}^d)$ , for a given  $\alpha \in ]0, \frac{1}{2}[$ .

We refer to the differential operator (2.1)  $L_t$  and we simply set here  $L \equiv L_t$ .

**Remark 1.3.12.** Suppose the validity of Assumptions 3, 5.

1. Let  $T > 0$ . Proposition 4.2 in [44] implies that for every  $v \in \mathbf{M}_f \mathbb{R}^d$ , there exists a unique  $\mathbf{M}_f \mathbb{R}^d$ -valued solution of equation (3.2) with initial value  $v$ . This unique solution will be denoted by  $\mathbf{u}^v$ . In the sequel  $T$  will be omitted.
2. We remark that the uniqueness result mentioned in item 1. is unknown in the case of general bounded coefficients. In the general framework, only a uniqueness result for non-negative solutions is available, see Remark 1.3.3 point 1.
3. Since  $L$  is time-homogeneous, taking into account Assumptions 3, 5, operating a shift, uniqueness of (3.2) also holds replacing the initial time 0 by any other initial time, for every initial value in  $\mathbf{M}_f \mathbb{R}^d$ , with any other maturity  $T$ .

**Theorem 1.3.13.** Suppose the validity of Assumptions 3 and 5. Then, for all  $\mu \in \mathbf{M}_f \mathbb{R}^d$ , equation (1.2) with terminal value  $\mu$  admits at most one  $\mathbf{M}_f \mathbb{R}^d$ -valued solution in the sense of Definition 1.3.1.

By Theorems 3.1.12, 3.1.14 and Corollary 3.1.16 in [82] the differential operator  $L$  suitably extends as a map  $(L) = D^{2\alpha+2}(\mathbb{R}^d) \subset C^{2\alpha}(\mathbb{R}^d) \subset C^{2\alpha} \mathbb{R}^d$  and that extension is sectorial, see Definition 2.0.1 in [82]. We set  $E := C^{2\alpha} \mathbb{R}^d$ . By the considerations below that Definition, in (2.0.2) and (2.0.3) therein, one defines  $P_t := e^{tL}$ ,  $P_t : E \rightarrow E$ ,  $t \geq 0$ . By Proposition 2.1.1 in [82],  $(P_t)_{t \geq 0}$  is a semigroup and  $t \mapsto P_t$  is analytical on  $]0, +\infty[$  with values in  $\mathcal{L}(E)$ , with respect to  $\|\cdot\|_E$ .

Before proving the theorem, we provide two lemmata.

**Lemma 1.3.14.** *Suppose the validity of Assumptions 3 and 5. Then, for all  $\varphi \in E$  and all  $\nu \in \mathcal{M}_f \mathbb{R}^d$ , the function from  $\mathbb{R}_+^*$  to  $\mathbb{R}$*

$$t \mapsto \int_{\mathbb{R}^d} P_t \varphi(x) \nu(dx)$$

is analytic.

*Proof.* The result can be easily established using the fact that  $t \mapsto P_t \varphi$  with values in  $\mathcal{L}(E)$  is analytic and the fact that the map  $\psi \mapsto \int_{\mathbb{R}^d} \psi(x) \nu(dx)$  is linear and bounded. □

**Lemma 1.3.15.** *Suppose the validity of Assumptions 3 and 5. Let  $T > 0$ . Then for all  $\nu \in \mathcal{M}_f \mathbb{R}^d$ ,  $t \in [0, T]$  and  $\varphi \in E$  we have the identity*

$$\int_{\mathbb{R}^d} P_t \varphi(x) \nu(dx) = \int_{\mathbb{R}^d} \varphi(x) \mathbf{u}^\nu(t)(dx), \tag{3.18}$$

where  $\mathbf{u}^\nu$  was defined in point 1. of Remark 1.3.12.

*Proof.* Let  $\nu \in \mathcal{M}_f \mathbb{R}^d$ . We denote by  $\mathbf{v}^\nu$  the mapping from  $[0, T]$  to  $\mathcal{M}_f \mathbb{R}^d$  such that  $\forall t \in [0, T]$ ,  $\forall \varphi \in E$

$$\int_{\mathbb{R}^d} \varphi(x) \mathbf{v}^\nu(t)(dx) = \int_{\mathbb{R}^d} P_t \varphi(x) \nu(dx). \tag{3.19}$$

Previous expression defines the measure  $\mathbf{v}^\nu(t, \cdot)$  since  $\varphi \mapsto \int_{\mathbb{R}^d} P_t \varphi(x) \nu(dx)$  is continuous with respect to the sup-norm, using  $\|P_t \varphi\|_\infty \leq \|\varphi\|_\infty$ , and Lebesgue dominated convergence theorem.

By approximating elements of  $E$  with elements of  $C_c^\infty \mathbb{R}^d$ , it will be enough to prove (3.18) for  $\varphi \in C_c^\infty \mathbb{R}^d$ .

Our goal is to show that  $\mathbf{v}^\nu$  is a  $\mathbb{R}^d$ -valued solution of (3.2) with initial value  $\nu$  to conclude  $\mathbf{v}^\nu = \mathbf{u}^\nu$  via point 1. of Remark 1.3.12 and so to prove (3.18) for  $\varphi \in C_c^\infty \mathbb{R}^d$ .

Let  $t \in [0, T]$  and  $\varphi \in C_c^\infty \mathbb{R}^d$ . On the one hand, point (i) of Proposition 2.1.1 in [82] gives

$$LP_t \varphi = P_t L \varphi, \tag{3.20}$$

since  $C_c^\infty \mathbb{R}^d \subset D(L) = C^{2\alpha+2} \mathbb{R}^d, \mathbb{R}$ . On the other hand, for all  $s \in [0, t]$ , we have

$$\begin{aligned} |LP_s \varphi|_E &= |P_s L \varphi|_{2\alpha} \\ &\leq \|P_s\|_E \|L \varphi\|_E \\ &\leq M_0 e^{-\omega s} \|L \varphi\|_E, \end{aligned}$$

with  $M_0, \omega$  the real parameters appearing in Definition 2.0.1 in [82] and using point (iii) of Proposition 2.1.1 in the same reference. Then the mapping  $s \mapsto LP_s\varphi$  belongs obviously to  $L^1([0, t]; E)$  and point (ii) of Proposition 2.1.4 in [82] combined with identity (3.20) gives

$$P_t\varphi = \varphi + \int_0^t P_s L\varphi ds.$$

Back to our main goal, using in particular Fubini's theorem, we have

$$\begin{aligned} \int_{\mathbb{R}^d} P_t\varphi(x) \nu(dx) &= \int_{\mathbb{R}^d} \varphi(x) \nu(dx) + \int_0^t \int_{\mathbb{R}^d} P_s L\varphi(x) ds \nu(dx) \\ &= \int_{\mathbb{R}^d} \varphi(x) \nu(dx) + \int_0^t \int_{\mathbb{R}^d} P_s L\varphi(x) \nu(dx) ds \\ &= \int_{\mathbb{R}^d} \varphi(x) \nu(dx) + \int_0^t \int_{\mathbb{R}^d} L\varphi(x) \mathbf{v}^\nu(s)(dx) ds. \end{aligned}$$

This shows that  $\mathbf{v}^\nu$  is a solution of equation (3.2).  $\square$

**Proof** (of Theorem 1.3.13).

Let  $\nu, \nu \in \mathbf{M}_f \mathbb{R}^d$  such that

$$\mu_T := \mathbf{u}^\nu(T) = \mathbf{u}^{\nu^j}(T).$$

Thanks to Proposition 1.3.6, it suffices to show that  $\nu = \nu$  i.e.

$$\forall \varphi \in C_c^\infty \mathbb{R}^d, \quad \int_{\mathbb{R}^d} \varphi(x) \nu(dx) = \int_{\mathbb{R}^d} \varphi(x) \nu(dx).$$

Since  $T > 0$  is arbitrary, by Remark 1.3.12 we can consider  $\mathbf{u}^{\nu, 2T}$  and  $\mathbf{u}^{\nu^j, 2T}$ , defined as the corresponding  $\mathbf{u}^\nu$  and  $\mathbf{u}^{\nu^j}$  functions obtained replacing the horizon  $T$  with  $2T$ . They are defined on  $[0, 2T]$  and by Remark 1.3.12 1. (uniqueness on  $[0, T]$ ), they constitute extensions of the initial  $\mathbf{u}^\nu$  and  $\mathbf{u}^{\nu^j}$ . By Remark 1.3.12 3., the uniqueness of an  $\mathbf{M}_f \mathbb{R}^d$ -valued solution of (3.2) (for  $t \in [T, 2T]$ , with  $T$  as initial time) holds for

$$\begin{aligned} \square \partial_t \mathbf{u}(\tau) &= L^* \mathbf{u}(\tau), \quad T \leq \tau \leq 2T \\ \square \mathbf{u}(T) &= \mu \end{aligned} \tag{3.21}$$

Now, the functions  $\mathbf{u}^{\nu, 2T}$  and  $\mathbf{u}^{\nu^j, 2T}$  solve (3.21) on  $[T, 2T]$ . This gives in particular

$$\forall \tau \geq T, \quad \forall \varphi \in C_c^\infty \mathbb{R}^d, \quad \int_{\mathbb{R}^d} \varphi(x) \mathbf{u}^{\nu, 2T}(\tau)(dx) = \int_{\mathbb{R}^d} \varphi(x) \mathbf{u}^{\nu^j, 2T}(\tau)(dx). \tag{3.22}$$

Fix  $\varphi \in C_c^\infty \mathbb{R}^d$ . Combining now the results of Lemmata 1.3.14 and 1.3.15, we obtain that the function

$$\tau \mapsto \int_{\mathbb{R}^d} \varphi(x) \mathbf{u}^{\nu, 2T}(\tau)(dx) - \int_{\mathbb{R}^d} \varphi(x) \mathbf{u}^{\nu^j, 2T}(\tau)(dx) \tag{3.23}$$

defined on  $[0, 2T]$ , is zero on  $[T, 2T]$  and analytic on  $]0, 2T[$ . Hence it is zero on  $]0, 2T[$ . By (3.18) we obtain

$$\int_{\mathbb{R}^d} P_\tau \varphi(x) \nu - \nu(dx) = 0, \quad \forall t \in ]0, 2T[. \tag{3.24}$$

Separating  $\nu$  and  $\nu$  in positive and negative components, we can finally apply dominated convergence theorem in (3.23) to send  $\tau$  to  $0^+$ . This is possible thanks to points (i) of Proposition 2.1.4 and (iii) of Proposition 2.1.1 in [82] together with the representation (3.18). Indeed  $P_\tau \varphi(x) \rightarrow \varphi(x)$  for every  $\varphi \in E, x \in \mathbb{R}^d$  when  $\tau \rightarrow 0^+$ . This shows  $\nu = \nu$  and ends the proof.  $\square$

For the sake of applications it is useful to formulate a piecewise time-homogeneous version of Theorem 1.3.13.

**Corollary 1.3.16.** *Let  $n \in \mathbb{N}^*$ . Let  $0 = t_0 < \dots < t_n = T$  be a partition. For  $k \in [[2, n]]$  (resp.  $k = 1$ ) we denote  $I_k = ]t_{k-1}, t_k]$  (resp.  $[t_0, t_1]$ ). Suppose that the following holds.*

1. *For all  $k \in [[1, n]]$ , the restriction of  $\sigma$  (resp.  $b$ ) to  $I_k \times \mathbb{R}^d$  is a time-homogeneous function  $\sigma^k : \mathbb{R}^d \rightarrow M_d(\mathbb{R})$  (resp.  $b^k : \mathbb{R}^d \rightarrow \mathbb{R}^d$ ).*
2. *Assumption 3.*
3. *Assumption 5 is verified for each  $\sigma^k, b^k$  and  $\Sigma^k$ , where we have set  $\Sigma^k := \sigma^k \sigma^{k-T}$ .*

*Then, for all  $\mu \in \mathbf{M}_f \mathbb{R}^d$ , equation (1.2) with terminal value  $\mu$  admits at most one  $\mathbf{M}_f \mathbb{R}^d$ -valued solution in the sense of Definition 1.3.1.*

*Proof.* For each given  $k \in [[1, n]]$ , we introduce the PDE operator  $L^k$  defined by

$$L^k := \frac{1}{2} \sum_{i,j=1}^d \Sigma_{ij}^k \partial_{ij} + \sum_{i=1}^d b_i^k \partial_i. \quad (3.25)$$

Let now  $\mathbf{u}^1, \mathbf{u}^2$  be two solutions of (1.2) with same terminal value  $\mu$ .

The measure-valued functions  $\mathbf{v}^i := \mathbf{u}^i(\cdot + t_{n-1}), i \in \{1, 2\}$  defined on  $[0, T - t_{n-1}]$  are solutions of

$$\begin{aligned} \square \partial_t \mathbf{v} &= (L^n) * \mathbf{v} \\ \square \mathbf{v}(T - t_{n-1}, \cdot) &= \mu \end{aligned} \quad (3.26)$$

in the sense of Definition 1.3.1 replacing  $T$  by  $T - t_{n-1}$  and  $L$  by  $L^n$ . Then, Theorem 1.3.13 gives  $\mathbf{v}^1 = \mathbf{v}^2$  and consequently  $\mathbf{u}^1 = \mathbf{u}^2$  on  $[t_{n-1}, T]$ . To conclude, we proceed by backward induction.  $\square$

### 1.3.4 Uniqueness: the case of Ornstein-Uhlenbeck semigroup

In this section, we consider the case  $b := (s, x) \mapsto C(s)x$  with  $C$  continuous from  $[0, T]$  to  $M_d(\mathbb{R})$  and  $\sigma$  continuous from  $[0, T]$  to  $M_{d,m}(\mathbb{R})$ . We set  $\Sigma := \sigma \sigma^T$ . We also denote by  $\mathbf{D}(t), t \in [0, T]$ , the unique solution of

$$\mathbf{D}(t) = I - \int_0^t C(s)^T \mathbf{D}(s) ds, \quad t \in [0, T]. \quad (3.27)$$

We recall that for every  $t \in [0, T]$ ,  $\mathbf{D}(t)$  is invertible and

$$\mathbf{D}^{-1}(t) = I + \int_0^t \mathbf{D}^{-1}(s)C(s)^{-T} ds, \quad t \in [0, T]. \tag{3.28}$$

For previous and similar properties, see Chapter 8 of [28].

In that setting, the classical Fokker-Planck PDE for finite measures reads

$$\begin{aligned} \square \partial_t \mathbf{u}(t) &= \sum_{i,j=1}^d \Sigma(t)_{ij} \partial_{ij} \mathbf{u}(t) - \sum_{i=1}^d \partial_i ((C(t)x)_i \mathbf{u}(t)) \\ \square \mathbf{u}(0) &= \nu \in \mathbf{M}_f \mathbb{R}^d. \end{aligned} \tag{3.29}$$

**Proposition 1.3.17.** *For all  $\nu \in \mathbf{M}_f \mathbb{R}^d$ , equation (3.29) with initial value  $\nu$  admits at most one  $\mathbf{M}_f \mathbb{R}^d$ -valued solution.*

*Proof.*

- Let  $\nu \in \mathbf{M}_f \mathbb{R}^d$  and  $\mathbf{u}$  be a solution of (3.2) with initial value  $\nu$ . Identity (3.3) can be extended to  $\mathbf{S} \mathbb{R}^d$  since for all  $t \in [0, T]$ ,  $\mathbf{u}(t)$  belongs to  $\mathbf{M}_f \mathbb{R}^d$ . Then,  $t \mapsto \mathbf{F}\mathbf{u}(t)$  verifies

$$\mathbf{F}\mathbf{u}(t)(\xi) = \mathbf{F}\nu(\xi) + \int_0^t C(s)^{-T} \xi, \nabla \mathbf{F}\mathbf{u}(s) ds - \frac{1}{2} \int_0^t \Sigma(s) \xi, \xi \mathbf{F}\mathbf{u}(s) ds, \quad (t, \xi) \in [0, T] \times \mathbb{R}^d. \tag{3.30}$$

In fact, the integrand inside the first integral has to be understood as a Schwartz distribution: in particular the symbol  $\nabla$  is understood in the sense of distributions and for each given  $s \in [0, T]$ ,  $C(s)^{-T} \xi, \nabla \mathbf{F}\mathbf{u}(s)$  denotes the tempered distribution

$$\phi \mapsto \sum_{i=1}^d \partial_i \mathbf{F}\mathbf{u}(s) \xi \mapsto C(s)^{-T} \xi \int \phi(\xi) d\xi.$$

Indeed, even though for any  $t$ ,  $\mathbf{F}\mathbf{u}(t)$  is a function, the equation (3.30) has to be understood in  $\mathbf{S} \mathbb{R}^d$ . Hence, for all  $\varphi \in \mathbf{S} \mathbb{R}^d$ , this gives

$$\begin{aligned} \int_{\mathbb{R}^d} \varphi(\xi) \mathbf{F}\mathbf{u}(t)(\xi) d\xi - \int_{\mathbb{R}^d} \varphi(\xi) \mathbf{F}\nu(\xi) d\xi & \tag{3.31} \\ = - \sum_{k,l=1}^d \int_0^t C(s)_{kl} \int_{\mathbb{R}^d} \xi_l \mathbf{F}\varphi_k(\xi) \mathbf{u}(s)(d\xi) ds - \frac{1}{2} \int_0^t \int_{\mathbb{R}^d} \Sigma(s) \xi, \xi \mathbf{F}\mathbf{u}(s)(\xi) \varphi(\xi) d\xi ds \\ = - \sum_{k,l=1}^d \int_0^t C(s)_{kl} \int_{\mathbb{R}^d} \mathbf{F}(\partial_l \varphi_k)(\xi) \mathbf{u}(s)(d\xi) ds - \frac{1}{2} \int_0^t \int_{\mathbb{R}^d} \Sigma(s) \xi, \xi \mathbf{F}\mathbf{u}(s)(\xi) d\xi ds \\ = - \int_0^t \int_{\mathbb{R}^d} \operatorname{div}_\xi C(s)^{-T} \xi \varphi(\xi) + \frac{1}{2} \Sigma(s) \xi, \xi \varphi(\xi) \mathbf{F}\mathbf{u}(s)(\xi) d\xi ds, \end{aligned}$$

where  $\varphi_k : \xi \mapsto \xi_k \varphi(\xi)$  for a given  $k \in \llbracket 1, d \rrbracket$ .

2. Let now  $\mathbf{v} : [0, T] \rightarrow \mathbf{M}_f \mathbb{R}^d$  defined by

$$\int_{\mathbb{R}^d} \varphi(x) \mathbf{v}(t)(dx) = \int_{\mathbb{R}^d} \varphi(\mathbf{D}(t)^{-T} x) \mathbf{u}(t)(dx), \quad (3.32)$$

$t \in [0, T]$ ,  $\varphi \in \mathcal{C}_b(\mathbb{R}^d)$ . For every  $\xi \in \mathbb{R}^d$ , we set  $\varphi(x) = \exp(-i \langle \xi, x \rangle)$  in (3.32) to obtain

$$\mathbf{Fv}(t)(\xi) = \mathbf{Fu}(t)(\mathbf{D}(t)\xi), \quad (3.33)$$

for all  $\xi \in \mathbb{R}^d$ , for all  $t \in [0, T]$ .

3. We want now to show that, for each  $\xi$ ,  $t \mapsto \mathbf{Fv}(t)$  fulfills an ODE. To achieve this, suppose for a moment that  $(t, \xi) \mapsto \mathbf{Fu}(t)(\xi)$  is differentiable with respect to the variable  $\xi$ . Then, on the one hand, we have for all  $(t, \xi) \in [0, T] \times \mathbb{R}^d$ ,

$$\mathbf{Fu}(t)(\xi) = \mathbf{Fv}(\xi) + \int_0^t C(s)^T \xi, \nabla_{\xi} \mathbf{Fu}(s)(\xi) ds - \frac{1}{2} \int_0^t \Sigma(s) \xi, \xi \mathbf{Fu}(s)(\xi) ds, \quad (3.34)$$

thanks to identity (3.30). This means in particular that, for each given  $\xi \in \mathbb{R}^d$ ,  $t \mapsto \mathbf{Fu}(t)(\xi)$  is differentiable almost everywhere on  $[0, T]$ .

On the other hand, for almost every  $t \in [0, T]$  and all  $\xi \in \mathbb{R}^d$ , we have

$$\begin{aligned} \partial_t \mathbf{Fv}(t)(\xi) &= \partial_t \mathbf{Fu}(t)(\mathbf{D}(t)\xi) + \sum_{i=1}^d \frac{d}{dt} (\mathbf{D}(t)\xi)_i \partial_i \mathbf{Fu}(t)(\mathbf{D}(t)\xi), \\ &= \partial_t \mathbf{Fu}(t)(\mathbf{D}(t)\xi) - \sum_{i=1}^d C(t)^T \mathbf{D}(t)\xi_i \partial_i \mathbf{Fu}(t)(\mathbf{D}(t)\xi), \\ &= -\frac{1}{2} \Sigma(t) \mathbf{D}(t)\xi, \mathbf{D}(t)\xi \mathbf{Fv}(t)(\xi), \end{aligned} \quad (3.35)$$

where from line 1 to line 2, we have used the fact  $\frac{d}{dt} (\mathbf{D}(t)\xi) = -C(t)^T \mathbf{D}(t)\xi$  for all  $(t, \xi) \in [0, T] \times \mathbb{R}^d$  and from line 2 to line 3, the identity (3.34). Since  $t \mapsto \mathbf{Fv}(t)(\xi)$  is absolutely continuous by (3.33), (3.35) implies

$$\mathbf{Fv}(t)(\xi) = \mathbf{Fv}(\xi) - \frac{1}{2} \int_0^t \Sigma(s) \mathbf{D}(s)\xi, \mathbf{D}(s)\xi \mathbf{Fv}(s)(\xi) ds, \quad \xi \in \mathbb{R}^d, \quad (3.36)$$

for all  $t \in [0, T]$ .

4. Now, if  $(t, \xi) \mapsto \mathbf{Fu}(t)(\xi)$  is not necessarily differentiable in the variable  $\xi$ , we will be able to prove (3.36) still holds by making use of calculus in the sense of distributions.

5. Suppose that (3.36) holds. This gives

$$\mathbf{Fu}(t)(\xi) = e^{-\int_0^t \frac{|\sigma(s)^T \xi|^2}{2} ds} \mathbf{Fv}(\mathbf{D}^{-1}(t)\xi). \quad (3.37)$$

6. The proof is now concluded after we have established the (3.36). Since both sides of it are continuous in  $(t, \xi)$ , it will be enough to show the equality as  $S(\mathbb{R}^d)$ -valued. This can be done differentiating (3.30), considered as an equality in  $S(\mathbb{R}^d)$ . For this we will apply Lemma 1.3.18 setting  $\Phi := \mathbf{F}\mathbf{u}(t)$  for every fixed  $t \in [0, T]$  and differentiating in time. We set  $\Phi_t(\xi) = \mathbf{F}\mathbf{v}(t)(\xi)$ ,  $\xi \in \mathbb{R}^d$  and  $\Phi_t(\phi) = \int_{\mathbb{R}^d} \phi(\xi)\Phi_t(\xi)d\xi$ ,  $\phi \in S(\mathbb{R}^d)$ . We remark that  $\Phi_t$  is compatible with the one defined in (3.38). (3.36) will directly follow from Lemma 1.3.18.

□

**Lemma 1.3.18.** Let  $\Phi \in S^1(\mathbb{R}^d)$ ,  $t \in [0, T]$ . We denote by  $\Phi_t$  the element of  $S^1(\mathbb{R}^d)$  such that for all

$$\phi \in S(\mathbb{R}^d) \quad \Phi_t(\phi) := \det \mathbf{D}^{-1}(t) \int_{\mathbb{R}^d} \phi(x) \mathbf{D}^{-1}(t) \cdot \dots \quad (3.38)$$

Then, for all  $t \in [0, T]$

$$\Phi_t(\phi) = \Phi(\phi) - \sum_{i=1}^d \int_0^t (\partial_i \Phi)_s(x) \mapsto C(s)^T \mathbf{D}(s)x \cdot \phi(x) ds. \quad (3.39)$$

*Proof.* We begin with the case  $\Phi \in S(\mathbb{R}^d)$  (or only  $C^\infty(\mathbb{R}^d)$ ). In this case,

$$\Phi_t(x) = \Phi(\mathbf{D}(t)x), \quad x \in \mathbb{R}^d, t \in [0, T].$$

Hence, for every  $t \in [0, T]$

$$\begin{aligned} \frac{d}{dt} \Phi_t(x) &= \left( \frac{d}{dt} \mathbf{D}(t)x, \nabla \Phi(\mathbf{D}(t)x) \right) \\ &= - C(t)^T \mathbf{D}(t)x, \nabla \Phi(\mathbf{D}(t)x) \\ &= - \sum_{i=1}^d C(t)^T \mathbf{D}(t)x \cdot (\partial_i \Phi)_t(x), \end{aligned}$$

Now, coming back to the general case, let  $\Phi \in S(\mathbb{R}^d)$  and  $(\phi_\varepsilon)_{\varepsilon > 0}$  a sequence of mollifiers in  $S(\mathbb{R}^d)$ , converging to the Dirac measure. Then for all  $\varepsilon > 0$ , the function  $\Phi * \phi_\varepsilon : x \mapsto \Phi(\phi_\varepsilon(x - \cdot))$  belongs to  $S(\mathbb{R}^d) \cap C^\infty(\mathbb{R}^d)$ . By the first part of the proof, (3.39) holds replacing  $\Phi = \Phi * \phi_\varepsilon$ . Now, this converges to  $\Phi$  in  $S(\mathbb{R}^d)$  when  $\varepsilon$  tends to  $0^+$ . (3.39) follows sending  $\varepsilon$  to  $0^+$ . Indeed, for all  $\phi \in S(\mathbb{R}^d)$ ,  $t \in [0, T]$ ,

setting  $\check{\varphi} : y \mapsto \varphi(-y)$ , we have

$$\begin{aligned} \Phi_t(\phi) &= \lim_{\rightarrow 0^+} \int_{\mathbb{R}^d} \phi(x) (\Phi * \varphi)_t(x) dx \\ &= \lim_{\rightarrow 0^+} \int_{\mathbb{R}^d} \phi(x) \Phi * \varphi(x) dx - \lim_{\rightarrow 0^+} \int_{i=1}^d \int_0^t \det \mathbf{D}^{-1}(s) \int_{\mathbb{R}^d} C(s)^T x \phi(\mathbf{D}^{-1}(s)x) \partial_i \Phi * \varphi(x) dx ds \\ &= \Phi(\phi) - \lim_{\rightarrow 0^+} \int_{i=1}^d \int_0^t \det \mathbf{D}^{-1}(s) \partial_i \Phi \int_{\mathbb{R}^d} C(s)^T \cdot \phi(\mathbf{D}^{-1}(s) \cdot) * \check{\varphi} ds \\ &= \Phi(\phi) - \int_{i=1}^d \int_0^t \det \mathbf{D}^{-1}(s) \partial_i \Phi \int_{\mathbb{R}^d} C(s)^T \cdot \phi(\mathbf{D}^{-1}(s) \cdot) ds \\ &= \Phi(\phi) - \int_{i=1}^d (\partial_i \Phi)_s \int_{\mathbb{R}^d} C(s)^T \mathbf{D}(s)x \phi(x) ds. \end{aligned}$$

To conclude, it remains to justify the commutation between the limit in  $E$  and the integral in time from line 3 to line 4 using Lebesgue dominated convergence theorem. On the one hand, for a given  $i \in \llbracket 1, d \rrbracket$ , the fact  $\partial_i \Phi$  belongs to  $\mathcal{R}^d$  implies that there exists  $C > 0, N \in \mathbb{N}$  such that for all

$$\phi \in \mathcal{S}(\mathbb{R}^d) \quad \left| \partial_i \Phi(\phi) \right| \leq C \sup_{|\alpha| \leq N} \sup_{x \in \mathbb{R}^d} (1 + |x|^2)^N |\partial^\alpha \phi(x)|,$$

see Chapter 1, Exercise 8 in [100]. On the other hand, the quantities

$$\sup_{x \in \mathbb{R}^d} (1 + |x|^2)^N \partial_x^\alpha x_j \phi(\mathbf{D}^{-1}(s) \cdot) * \check{\varphi}$$

are bounded uniformly in the couple  $(s, E)$ , for all  $j \in \llbracket 1, d \rrbracket, \alpha \in \mathbb{N}^d$ , taking also into account that the function  $s \mapsto \mathbf{D}^{-1}(s)$  is continuous and therefore bounded. Since  $C$  is also continuous on  $[0, T]$ , we are justified to use Lebesgue's dominated convergence theorem. □

**Theorem 1.3.19.** For all  $\mu \in \mathcal{M}_f(\mathbb{R}^d)$ , equation (1.2) with terminal value  $\mu$  admits at most one  $\mathcal{M}_f(\mathbb{R}^d)$ -valued solution in the sense of Definition 1.3.1.

*Proof.* Let  $\mu \in \mathcal{M}_f(\mathbb{R}^d)$  and  $\mathbf{u}$  a solution of (2.3) with terminal value  $\mu$ . Then,  $\mathbf{u}$  solves equation (3.2) with initial value  $\mathbf{u}(0)$ . As a consequence, by I (3.37) appearing at the end of the proof of Proposition 1.3.17, for all  $\xi \in \mathbb{R}^d$ ,

$$\mathbf{F}\mu(\xi) = e^{-\int_0^T \frac{|\sigma(s)^T \xi|^2}{2} ds} \mathbf{F}\mathbf{u}(0) \cdot \mathbf{D}^{-1}(T) \xi,$$

so that

$$\mathbf{F}\mathbf{u}(0)(\xi) = e^{\int_0^T \frac{|\sigma(s)^T \xi|^2}{2} ds} \mathbf{F}\mu(\mathbf{D}(T) \xi).$$

Hence,  $\mathbf{u}(0)$  is entirely determined by  $\mu$  and Proposition 1.3.17 gives the result. □



## 1.4 McKean SDEs related to time-reversal of diffusions

### 1.4.1 Preliminary considerations

In this last section we concentrate on the analysis of the well-posedness of the McKean SDE (1.3). Regarding  $b : [0, T] \times \mathbb{R}^d \mapsto \mathbb{R}^d$ ,  $\sigma : [0, T] \times \mathbb{R}^d \mapsto M_{d,m}(\mathbb{R})$ , we set  $b := b(T - \cdot, \cdot)$ ,  $\sigma := \sigma(T - \cdot, \cdot)$ ,  $\Sigma := \sigma^{-T}\sigma$ .

Given a probability-valued function  $\mathbf{p} : [0, T] \rightarrow \mathcal{P}(\mathbb{R}^d)$ , we denote by  $p_t$  the density of  $\mathbf{p}(t)$ , for  $t \in [0, T]$ , whenever it exists. In this section,  $\mu$  will denote the terminal condition of (1.1) supposed to be a probability. Remarking  $\mu = \bar{\mu}$ , we consider the following notion of solution.

**Definition 1.4.1.** *On a given filtered probability space  $\Omega, \mathbf{F}, (\mathbf{F}_t)_{t \in [0, T]}$ ,  $\mathbf{P}$  equipped with an  $m$ -dimensional  $(\mathbf{F}_t)_{t \in [0, T]}$ -Brownian motion  $\beta$ , a **solution** of equation (1.3) is a couple  $(Y, \mathbf{p})$  fulfilling (1.3) with Brownian motion  $\beta$ , such that  $Y$  is  $(\mathbf{F}_t)_{t \in [0, T]}$ -adapted and such that for all  $i \in \llbracket 1, d \rrbracket$ , all compact  $K \subset \mathbb{R}^d$ , all  $\tau < T$*

$$\int_0^\tau \int_K \operatorname{div}_y \Sigma_i(r, y) p_r(y) \, dy dr < \infty. \tag{4.1}$$

**Remark 1.4.2.** *For a given solution  $(Y, \mathbf{p})$  of equation (1.3), identity (4.1) appearing in Definition 1.4.1 implies in particular that, for all  $i \in \llbracket 1, d \rrbracket$ , all  $\tau < T$*

$$\int_0^\tau \frac{\operatorname{div}_y \Sigma_i(r, Y_r) p_r(Y_r)}{p_r(Y_r)} \, dr < \infty, \mathbf{P}\text{-a.s.}$$

The terminology stating that (1.3) constitutes a probabilistic representation of (1.2) is justified by the result below.

**Proposition 1.4.3.** *Suppose  $b, \sigma$  locally bounded. If  $(Y, \mathbf{p})$  is a solution of (1.3) in the sense of Definition 1.4.1, then  $\mathbf{p}(T - \cdot)$  is a solution of (1.1), with  $\mu = \mathbf{p}(0)$  in the sense of Definition 1.3.1.*

*Proof.* Let  $(Y, \mathbf{p})$  be a solution of (1.3) in the sense of Definition 1.4.1 with a Brownian motion symbolized by  $\beta$ . Let  $\varphi \in C_c^\infty \mathbb{R}^d$  and  $t \in ]0, T[$ . Itô's formula gives

$$\varphi(Y_{T-t}) = \varphi(Y_0) + \int_0^{T-t} \tilde{b}(s, Y_s; p_s) \, ds + \frac{1}{2} \int_0^{T-t} \operatorname{Tr} \Sigma(s, Y_s) \nabla^2 \varphi(Y_s) \, ds + \int_0^{T-t} \nabla \varphi(Y_s)^T \sigma(s, Y_s) \, d\beta_s, \tag{4.2}$$

with

$$\tilde{b}(s, y; p_s) := \sum_{j \in \llbracket 1, d \rrbracket} \frac{\operatorname{div}_y \Sigma_j(s, y) p_s(y)}{p_s(y)} - b(s, y), \quad (s, y) \in ]0, T[ \times \mathbb{R}^d.$$

We now want to take the expectation in identity (4.2). On the one hand, Remark 1.4.2, implies that for all  $i \in \llbracket 1, d \rrbracket$  and  $s \in ]0, T[$

$$\int_0^T ds \mathbf{E} \frac{\operatorname{div}_y \Sigma_i(s, Y_s) p_s(Y_s)}{p_s(Y_s)} \partial_i \varphi(Y_s) < \infty.$$

On the other hand

$$\int_0^T \mathbb{E} \operatorname{Tr} \Sigma(s, Y_s) \nabla^2 \varphi(Y_s) ds = \int_0^T \int_{\mathbb{R}^d} \sum_{i,j=1}^d \Sigma_{ij}(s, y) \partial_{ij} \varphi(y) p_s(y) dy ds \text{ p.s.}$$

Previous expression is finite since  $\sigma$  is bounded on compact sets and the partial derivatives of  $\varphi$  have compact supports. With similar arguments we prove that  $\int_0^T ds \mathbb{E} b(s, Y_s), \nabla \varphi(Y_s) < \infty, s \in ]0, T[$ . Moreover, fixing  $s \in ]0, T[$ , integrating by parts we have

$$\begin{aligned} \mathbb{E} \left\{ \tilde{b}(s, Y_s; p_s), \nabla \varphi(Y_s) \right\} &= \int_{\mathbb{R}^d} \sum_{k,j=1}^d \partial_k \Sigma_{jk}(s, y) p_s(y) \partial_j \varphi(y) dy - \int_{\mathbb{R}^d} b(s, y), \nabla \varphi(y) p_s(y) dy \\ &= - \int_{\mathbb{R}^d} \operatorname{Tr} \Sigma(s, y) \nabla^2 \varphi(y) p_s(y) dy - \int_{\mathbb{R}^d} b(s, y), \nabla \varphi(y) p_s(y) dy. \end{aligned} \tag{4.3}$$

Now, the quadratic variation of the local martingale  $M^Y := \int_0^\cdot \nabla \varphi(Y_s)^T \sigma(s, Y_s) d\beta_s$  yields

$$\langle M^Y \rangle_0^\cdot = \int_0^\cdot \nabla \varphi(Y_s)^T \Sigma(s, Y_s) \nabla \varphi(Y_s) ds.$$

We remark in particular that  $\mathbb{E} \langle M^Y \rangle_T < \infty$  since  $\sigma$  is bounded on compact sets and  $\varphi$  has compact support. This shows  $M^Y$  is a true (even square integrable) martingale and all terms involved in (4.2) are integrable.

At this point we evaluate the expectation in (4.2) taking into account the considerations above together with (1.4.1) and (4.3). We obtain

$$\mathbb{E}(\varphi(Y_{T-t})) = \int_{\mathbb{R}^d} \varphi(y) \mu(dy) - \int_0^{T-t} \int_{\mathbb{R}^d} L_{T-s} \varphi(y) p_s(y) dy ds.$$

Applying the change of variable  $t \mapsto T - t$ , we finally obtain the identity

$$\int_{\mathbb{R}^d} \varphi(y) p_{T-t}(y) dy = \int_{\mathbb{R}^d} \varphi(y) \mu(dy) - \int_t^T \int_{\mathbb{R}^d} L_s \varphi(y) p_{T-s}(y) dy ds,$$

which means that  $\mathbf{p}(T - \cdot)$  solves (1.2) in the sense of Definition 1.3.1 with terminal value  $\mu$ . □

We also provide the different notions of existence and uniqueness for (1.3) we will use in the sequel.

**Definition 1.4.4.** Let  $\mathbf{A}$  be a class of measure-valued functions from  $[0, T]$  to  $\mathbf{P}(\mathbb{R}^d)$ .

1. We say that (1.3) admits **existence in law** in  $\mathbf{A}$ , if there exists a complete filtered probability space  $(\Omega, \mathbf{F}, (\mathbf{F}_t)_{t \in [0, T]}, \mathbf{P})$  equipped with an  $m$ -dimensional  $(\mathbf{F}_t)_{t \in [0, T]}$ -Brownian motion  $\beta$  and a couple  $(Y, \mathbf{p})$  solution of (1.3) in the sense of Definition 1.4.1 such that  $\mathbf{p}$  belongs to  $\mathbf{A}$ .
2. Let  $Y^1, \mathbf{p}^1, Y^2, \mathbf{p}^2$  be two solutions of (1.3) in the sense of Definition 1.4.1 associated to some complete filtered probability spaces  $(\Omega^1, \mathbf{F}^1, (\mathbf{F}_t^1)_{t \in [0, T]}, \mathbf{P}^1), (\Omega^2, \mathbf{F}^2, (\mathbf{F}_t^2)_{t \in [0, T]}, \mathbf{P}^2)$  respectively, equipped with Brownian motions  $\beta^1, \beta^2$  respectively and such that  $\mathbf{p}^1, \mathbf{p}^2$  belong to  $\mathbf{A}$ . We say that (1.3) admits **uniqueness in law** in  $\mathbf{A}$ , if  $Y_0^1, Y_0^2$  have the same law implies that  $Y^1, Y^2$  have the same law.

3. We say that (1.3) admits **strong existence** in  $\mathbf{A}$  if for any complete filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$  equipped with an  $m$ -dimensional  $(\mathcal{F}_t)_{t \in [0, T]}$ -Brownian motion  $\beta$ , there exists a solution  $(Y, \mathbf{p})$  of equation (1.3) in the sense of Definition 1.4.1 such that  $\mathbf{p}$  belongs to  $\mathbf{A}$ .
4. We say that (1.3) admits **pathwise uniqueness** in  $\mathbf{A}$  if for any complete filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$  equipped with an  $m$ -dimensional  $(\mathcal{F}_t)_{t \in [0, T]}$ -Brownian motion  $\beta$ , for any solutions  $Y^1, \mathbf{p}^1, Y^2, \mathbf{p}^2$  of (1.3) in the sense of Definition 1.4.1 such that  $Y_0^1 = Y_0^2$   $\mathbb{P}$ -a.s. and  $\mathbf{p}^1, \mathbf{p}^2$  belong to  $\mathbf{A}$ , we have  $Y^1 = Y^2$ ,  $\mathbb{P}$ -a.s.

We finally define the sets in which we will formulate existence and uniqueness results in the sequel.

**Notation 1.** 1. For a given  $\mathbb{R}^d \mathcal{C} \subseteq \mathcal{C}$  denotes the set of measure-valued functions  $\mathbf{p}$  from  $[0, T]$  to  $\mathcal{C}$  such that  $\mathbf{p}(T)$  belongs to  $\mathcal{C}$ . Furthermore, for a given measure-valued function  $\mathbf{p} : [0, T] \rightarrow \mathcal{C}$ , we will write

$$b(t, \cdot; \mathbf{p}_t) := \frac{\sum_{i \in [1, d]} \sigma_{ij} p_t^j}{p_t^i}, \quad (4.4)$$

for almost all  $t \in [0, T]$  whenever  $p_t$  exists and the right-hand side quantity is well-defined. The function  $(t, x) \mapsto b(t, x; \mathbf{p}_t)$  is defined on  $[0, T] \times \mathbb{R}^d$  with values in  $\mathbb{R}^d$ .

2. Let  $\mathbf{A}_1$  (resp.  $\mathbf{A}_2$ ) denote the set of measure-valued functions from  $[0, T]$  to  $\mathcal{C}$  such that, for all  $t \in [0, T[$ ,  $\mathbf{p}(t)$  admits a density  $p_t$  with respect to the Lebesgue measure on  $\mathbb{R}^d$  and such that  $(t, x) \mapsto b(t, x; \mathbf{p}_t)$  is locally bounded (resp. is locally Lipschitz in space with linear growth) on  $[0, T] \times \mathbb{R}^d$ .

We state now existence and uniqueness results for equation (1.3) in different settings.

### 1.4.2 PDE with terminal condition and existence for the McKean SDE

The existence result for equation (1.3) will be based on two pillars: the reachability condition constituted by the existence of a solution of the Fokker-Planck PDE with terminal condition and the time-reversal techniques of [55]. More precisely, we suppose that Assumption 4 is in force for a fixed  $\mathcal{C} \subseteq \mathcal{P}(\mathbb{R}^d)$  and consider the following extra assumptions, i.e. Assumptions 6, 7 and 8, still with respect to  $(\mathcal{C}, \mu)$ .

**Assumption 6.** The backward PDE (1.1) with terminal condition  $\mu$  admits at least an  $\mathbf{M}_+$ -valued solution  $\mathbf{u}$  in the sense of Definition 1.3.1 verifying the following.

1.  $\mathbf{u}(0)$  belongs to  $\mathcal{C}$ .
2.  $\forall t \in ]0, T[$ ,  $\mathbf{u}(t)$  admits a density with respect to the Lebesgue measure on  $\mathbb{R}^d$  (denoted by  $u(t, \cdot)$ ) and for all  $t_0 > 0$  and all compact  $K \subset \mathbb{R}^d$

$$\int_{t_0}^T \int_K |u(t, x)|^2 + \sum_{i=1}^d \sum_{j=1}^m |\sigma_{ij}(t, x) \partial_i u(t, x)|^2 dx dt < \infty. \quad (4.5)$$

**Remark 1.4.5.** Suppose Assumption 1 holds and let  $\mathbf{u}$  be the measure-valued function appearing in Assumption 6. Then (4.5) implies that the family of densities  $u(T - t, \cdot)$ ,  $t \in ]0, T[$  verifies condition (4.1) appearing in Definition 1.4.1. To show this, it suffices to check that for all  $t_0 > 0$ , all compact  $K \subset \mathbf{R}^d$  and all  $(i, j, k) \in [[1, d]]^2 \times [[1, m]]$

$$\int_{t_0}^T \int_K |\partial_j (\sigma_{ik}(s, y) \sigma_{jk}(s, y) u(s, y))| dy ds < \infty. \quad (4.6)$$

The integrand appearing in (4.6) is well-defined. Indeed, in the sense of distributions we have

$$\partial_j (\sigma_{ik} \sigma_{jk} u) = \sigma_{ik} \sigma_{jk} \partial_j u + u (\sigma_{ik} \partial_j \sigma_{jk} + \sigma_{jk} \partial_j \sigma_{ik}); \quad (4.7)$$

moreover the components of  $\sigma$  are Lipschitz, so they are (together with their space derivatives) locally bounded. Also  $u$  and  $\sigma_{jk} \partial_j$  are square integrable by (4.5). This implies (4.6).

We introduce two new assumptions.

**Assumption 7.** Let  $\mathbf{u}$  be the measure-valued mapping appearing in Assumption 6. We suppose that  $\mu$  admits a density and  $\mathbf{u}(T - \cdot) \lfloor_{]0, T[ \times \mathbf{R}^d}$  belongs to  $\mathbf{A}_1$ .

**Assumption 8.** Let  $\mathbf{u}$  be the measure-valued mapping appearing in Assumption 6. We suppose that  $\mu$  admits a density and  $\mathbf{u}(T - \cdot) \lfloor_{]0, T[ \times \mathbf{R}^d}$  belongs to  $\mathbf{A}_2$ .

We remark that Assumption 8 implies 7.

**Proposition 1.4.6.** Suppose the validity of Assumptions 1, Assumption 4 with respect to  $\mathbf{C}$  and Assumption 6 with respect to  $(\mathbf{C}, \mu)$ . Then (1.3) admits existence in law in  $\mathbf{A}_C$ .

In particular if, moreover, Assumption 7 (resp. 8) holds, then (1.3) admits existence in law in  $\mathbf{A}_C \cap \mathbf{A}_1$  (resp. strong existence in  $\mathbf{A}_C \cap \mathbf{A}_2$ ).

*Proof.* By Assumption 6, there is an  $\mathbf{M}_+ \mathbf{R}^d$ -valued solution  $\mathbf{u}$  of equation (1.1) in the sense of Definition 1.3.1 such that  $\mathbf{u}(T) = \mu$  and  $\mathbf{u}(0)$  belongs to  $\mathbf{C}$ . We consider now a filtered probability space  $\Omega, \mathbf{F}, (\mathbf{F}_t)_{t \in [0, T]}$ ,  $\mathbf{P}$  equipped with an  $(\mathbf{F}_t)_{t \in [0, T]}$ -Brownian motion  $W$ . Let  $X_0$  be a r.v. distributed according to  $\mathbf{u}(0)$ . Under Assumption 1, it is well-known that there is a solution  $X$  to

$$X_t = X_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s, \quad t \in [0, T]. \quad (4.8)$$

Now, by Proposition 1.3.2,  $t \mapsto \mathbf{L}(X_t)$  is a  $\mathbf{P} \mathbf{R}^d$ -valued solution of equation (3.2) in the sense of (3.3) with initial value  $\mathbf{u}(0) \in \mathbf{C}$ . Then Assumption 4 gives

$$\mathbf{L}(X_t) = \mathbf{u}(t), \quad t \in [0, T], \quad (4.9)$$

since  $\mathbf{u}$  solves also (3.2) with initial value  $\mathbf{u}(0) \in \mathbf{C}$ . This implies in particular that  $\mathbf{u}$  is probability valued and that for all  $t \in ]0, T[$ ,  $X_t$  has  $u(t, \cdot)$  as a density fulfilling condition (4.5) in Assumption 6.

Combining this observation with Assumption 1, Theorem 2.1 in [55] states that there exists a filtered probability space  $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t \in [0, T]}, \mathbb{Q})$  equipped with the Brownian motion  $\beta$  and a copy of  $\hat{X}$  (still denoted by the same letter) such that  $X$  fulfills the first line of (1.3) with  $\beta$  and

$$\mathbf{p}(t) := \mathbf{u}(T - t), \quad t \in ]0, T[. \quad (4.10)$$

Finally, existence in law for (1.3) in the sense of Definition 1.4.1 holds since  $(X, \mathbf{u}(T - \cdot))$  is a solution of (1.3) on the same filtered probability space and the same Brownian motion above. This occurs in  $\mathbf{A}_C$  since  $\mathbf{L} X_T \in \mathbf{C}$  thanks to equality (4.9) for  $t = T$ .

We discuss rapidly the *in particular* point.

- Suppose the validity of Assumption 7. Then  $\mathbf{u}(T - \cdot)$  belongs to  $\mathbf{A}_C \cap \mathbf{A}_1$  and we also have existence in law in  $\mathbf{A}_C \cap \mathbf{A}_1$ .
- Suppose the validity of Assumption 8. Then, taking into account (4.10), strong existence and pathwise uniqueness for the first line of (1.3) holds by classical arguments since the coefficients are locally Lipschitz with linear growth, see [96] Exercise (2.10), and Chapter IX.2 and [96], Th. 12. section V.12. of [99]. By Yamada-Watanabe theorem this implies uniqueness in law, which shows that  $\mathbf{u}(T - \cdot)$  constitutes the marginal laws of the considered strong solutions. This concludes the proof of strong existence in  $\mathbf{A}_C \cap \mathbf{A}_2$  since  $\mathbf{u}(T - \cdot)$  belongs to  $\mathbf{A}_C \cap \mathbf{A}_2$ , by Assumption 8.

□

**Remark 1.4.7.** By (4.10), the second component  $\mathbf{p}$  of the solution of (1.3) is given by  $\mathbf{u}(T - \cdot)$ .

### 1.4.3 PDE with terminal condition and uniqueness for the McKean SDE

In this subsection we discuss some questions related to uniqueness for equation (1.3). We state the following hypothesis related to  $(\mu, \mathbf{C})$  where  $\mathbf{C}$  is a given subset of  $\mathbf{P} \mathbb{R}^d$ .

**Assumption 9.** *The equation (1.1) with terminal condition  $\mu$  admits at most a  $\mathbf{P} \mathbb{R}^d$ -valued solution  $\mathbf{u}$  in the sense of Definition 1.3.1 such that  $\mathbf{u}(0)$  belongs to  $\mathbf{C}$ .*

We recall that Section 1.3.2 provides various classes of examples where Assumption 9 holds.

**Proposition 1.4.8.** *Suppose the validity of Assumption 9 with respect to  $(\mu, \mathbf{C})$  and suppose  $b, \sigma$  to be locally bounded.*

*Let  $Y^i, \mathbf{p}^i$ ,  $i \in \{1, 2\}$  be two solutions of equation (1.3) in the sense of Definition 1.4.1 such that  $\mathbf{p}^1(T), \mathbf{p}^2(T)$  belong to  $\mathbf{C}$ . Then,*

$$\mathbf{p}^1 = \mathbf{p}^2.$$

*Proof.* Proposition 1.4.3 shows that  $\mathbf{p}^1(T - \cdot), \mathbf{p}^2(T - \cdot)$  are  $\mathbf{P} \mathbb{R}^d$ -valued solutions of equation (1.2) in the sense of Definition 1.3.1 with terminal value  $\mu$ . Assumption 9 gives the result since  $\mathbf{p}^1(T), \mathbf{p}^2(T)$  belong to  $\mathbf{C}$ . □

As a corollary, we establish some consequences about uniqueness in law and pathwise uniqueness results for equation (1.3) in the classes  $\mathbf{A}_1$  and  $\mathbf{A}_2$ .

**Corollary 1.4.9.** *Suppose the validity of Assumption 9 with respect to  $(\mu, \mathbf{C})$ . Then, the following results hold.*

1. *If  $b$  is locally bounded,  $\sigma$  is continuous and if the non-degeneracy Assumption 3 holds then (1.3) admits uniqueness in law in  $\mathbf{A}_C \cap \mathbf{A}_1$ .*
2. *If  $(b, \sigma)$  are locally Lipschitz with linear growth in space, then (1.3) admits pathwise uniqueness in  $\mathbf{A}_C \cap \mathbf{A}_2$ .*

*Proof.* If  $(Y, \mathbf{p})$  is a solution of (1.3) and is such that  $\mathbf{p}(T)$  belongs to  $\mathbf{C}$ , then by Proposition 1.4.8  $\mathbf{p}$  is determined by  $\mu = L(Y_0)$ .

To show that item 1. (resp. 2.) holds, it suffices to show that the classical SDE

$$dX_t = b(t, X_t; \mathbf{p}_t) dt + \sigma(t, X_t) dW_t, \quad t \in [0, T], \quad (4.11)$$

where  $b$  was defined in (4.4) and  $W$  an  $m$ -dimensional Brownian motion, admits uniqueness in law (resp. pathwise uniqueness). The mentioned uniqueness in law is a consequence of Theorem 10.1.3 in [104] and pathwise uniqueness holds by [96] Exercise (2.10), and Chapter IX.2 and [99] Th. 12. Section V.12.  $\square$

#### 1.4.4 Well-posedness for the McKean SDE: the bounded coefficients case

In this section, we state a significant result related to existence and uniqueness in law together with pathwise uniqueness for equation (1.3). In particular we obtain existence and uniqueness in law for (1.3) in the class  $\mathbf{A}_1$ .

We formulate the following hypotheses.

**Assumption 10.** 1. *Assumption 3 holds.*

2. *The functions  $\sigma$  is Lipschitz (in space).*

3. *The functions  $\sigma, b, (\nabla_x b_i)_{i \in \llbracket 1, d \rrbracket}, (\nabla_x \Sigma_{ij})_{i, j \in \llbracket 1, d \rrbracket}$  are continuous bounded and  $\nabla_x^2 \Sigma$  is Hölder continuous with exponent  $\alpha \in ]0, 1[$  in space uniformly in time.*

**Assumption 11.**  $\Sigma$  is supposed to be Hölder continuous in time

**Remark 1.4.10.** *Under Assumption 10, for every  $v \in \mathbf{P}(\mathbb{R}^d)$  there exists a unique  $\mathbf{P} \mathbb{R}^d$ -valued solution  $\mathbf{u}^v$  of (3.2).*

*Indeed the assumptions of Lemma 1.3.4 are fulfilled.*

We continue with a fundamental lemma whose proof will appear in the Appendix.

**Lemma 1.4.11.** *Suppose the validity of Assumptions 10 and 11. Then, for all  $v \in \mathbf{P}(\mathbb{R}^d)$   $\mathbf{u}^v(t)$  admits a density  $u^v(t, \cdot) \in C^1(\mathbb{R}^d)$  for all  $t \in ]0, T]$ . Furthermore, for each compact  $K$  of  $]0, T] \times \mathbb{R}^d$ , there are strictly positive constants  $C_1^K, C_2^K, C_3^K$ , also depending on  $v$  such that*

$$C_1^K \leq u^v(t, x) \leq C_2^K \quad (4.12)$$

$$|\partial_i u^v(t, x)| \leq C_3^K, \quad i \in \llbracket 1, d \rrbracket, \quad (4.13)$$

for all  $(t, x) \in K$ .

**Lemma 1.4.12.** *Let  $\mu$  be the probability measure introduced at the beginning of Section 1.4.1. Suppose that  $\mu$  equals  $\mathbf{u}^v(T)$  for some  $v \in \mathbf{P}(\mathbb{R}^d)$ . We assume the following.*

1. Assumptions 10.
2.  $\mathbf{u}^v(t)$  admits a density  $u^v(t, \cdot) \in W_{\text{loc}}^{1,1}(\mathbb{R}^d)$ , for all  $t \in ]0, T]$ .
3. For each compact  $K$  of  $]0, T] \times \mathbb{R}^d$ , there are strictly positive constants  $C_1^K, C_2^K, C_3^K$ , also depending on  $v$  such that (4.12) and (4.13) hold  $\forall (t, x) \in K$ .

Then equation (1.3) admits existence in law in  $\mathbf{A}_1$ .

**Corollary 1.4.13.** *We suppose the validity of Assumptions 10 and 11.*

1. Suppose the existence of  $v \in \mathbf{P}(\mathbb{R}^d)$  such that  $\mathbf{u}^v(T) = \mu$ . Then, equation (1.3) admits existence in law in  $\mathbf{A}_1$ . Moreover, if  $v$  is a Dirac mass, existence in law occurs in  $\mathbf{A}_{(\delta_x)_{x \in \mathbb{R}^d}} \cap \mathbf{A}_1$ .
2. Otherwise (1.3) does not admit existence in law.

*Proof.*

1. The first part is a direct consequence of Lemma 1.4.11, Lemma 1.4.12 and expression (4.4). If in addition,  $v$  is a Dirac mass, then  $\mathbf{u}^v(0)$  belongs to  $\mathbf{C} := (\delta_x)_{x \in \mathbb{R}^d}$ , hence existence in law occurs in  $\mathbf{A}_C \cap \mathbf{A}_1$  again by Proposition 1.4.6.
2. Otherwise suppose *ab absurdo* that  $(Y, \mathbf{p})$  is a solution of (1.3). By Proposition 1.4.3  $\mathbf{p}(T - \cdot)$  is a solution of (2.3). We set  $v_0 = \mathbf{p}(T)$  so that  $\mathbf{p}(T - \cdot)$  verifies also (3.2) with initial value  $v_0$ . Since, by Lemma 1.3.4 uniqueness holds for (3.2), it follows that  $\mathbf{p}(T - \cdot) = \mathbf{u}^{v_0}$  which concludes the proof of item 2.

□

**Proof** (of Lemma 1.4.12). Suppose  $\mu = \mathbf{u}^v(T)$  for some  $v \in \mathbf{P}(\mathbb{R}^d)$ .

We recall that Assumption 4 holds with respect to  $\mathbf{C} := \mathbf{P}(\mathbb{R}^d)$  by Remark 1.3.3 1.

In view of applying Proposition 1.4.6, we need to check that Assumptions 6 and 7 hold with respect to  $(\mu, \mathbf{C})$ .

Assumption 6 is verified by  $\mathbf{u} = \mathbf{u}^\nu$ . Indeed the function  $\mathbf{u}^\nu$  is a  $\mathbb{P}$   $\mathbb{R}^d$ -valued solution of (1.2) with terminal value  $\mu$  and such that  $\mathbf{u}^\nu(0)$  belongs to  $\mathbf{C}$ . Condition (4.5) appearing in Assumption 6 is satisfied with  $\mathbf{u} = \mathbf{u}^\nu$  thanks to the right-hand side of inequalities (4.12) and (4.13) and the fact that  $\sigma$  is bounded. Hence Assumption 6 holds with respect to  $(\mu, \mathbf{C})$ .

It remains to show Assumption 7 holds i.e. that

$$(t, x) \mapsto \frac{\operatorname{div}_x \Sigma_i(t, x) u^\nu(T - t, x)}{u^\nu(T - t, x)}$$

is locally bounded on  $[0, T] \times \mathbb{R}^d$ . To achieve this, we fix  $i \in \llbracket 1, d \rrbracket$  and a bounded open subset  $\mathbf{O}$  of  $[0, T] \times \mathbb{R}^d$ . For  $(t, x) \in \mathbf{O}$  we have

$$\frac{\operatorname{div}_x \Sigma_i(t, x) u^\nu(T - t, x)}{u^\nu(T - t, x)} \leq \operatorname{div}_x \Sigma_i(t, x) + \Sigma_i(t, x) \frac{|\nabla_x u^\nu(T - t, x)|}{u^\nu(T - t, x)}.$$

The latter quantity is locally bounded in  $t, x$  thanks to the boundedness of  $\Sigma$ ,  $\operatorname{div}_x \Sigma_i$  and inequalities (4.12) and (4.13). Hence, Assumption 7 holds. This ends the proof.  $\square$

**Proposition 1.4.14.** *Suppose the validity of Assumption 10 and 11. The following results hold.*

1. Let us suppose  $d = 1$ . Suppose  $\mu$  equals  $\mathbf{u}^{\delta_{x_0}}(T)$  for some  $x_0 \in \mathbb{R}$ . Then (1.3) admits existence and uniqueness in law in  $\mathbf{A}_{(\delta_x)_{x \in \mathbb{R}^d}} \cap \mathbf{A}_1$ , pathwise uniqueness in  $\mathbf{A}_{(\delta_x)_{x \in \mathbb{R}^d}} \cap \mathbf{A}_2$ .
2. Let  $d \geq 2$ . There is a maturity  $T$  sufficiently small (only depending on the Lipschitz constant of  $b, \sigma$ ) such that the following result holds. Suppose  $\mu$  equals  $\mathbf{u}^{\delta_{x_0}}(T)$  for some  $x_0 \in \mathbb{R}^d$ . Then (1.3) admits existence and uniqueness in law in  $\mathbf{A}_{(\delta_x)_{x \in \mathbb{R}^d}} \cap \mathbf{A}_1$ , pathwise uniqueness in  $\mathbf{A}_{(\delta_x)_{x \in \mathbb{R}^d}} \cap \mathbf{A}_2$ .

*Proof.* By Assumptions 10 and 11, Corollary 1.4.13 implies that (1.3) admits existence in law in the two cases in the specific classes. To check the uniqueness in law and pathwise uniqueness results, we wish to apply Corollary 1.4.9. It suffices to check Assumption 9 because the other hypotheses are included in Assumption 10. Below we verify Assumption 9 with respect to  $(\mu, (\delta_x)_{x \in \mathbb{R}})$ , for the separate two cases.

1. Fix  $x_0 \in \mathbb{R}^d$ . This will follow from Proposition 1.3.9 that holds under Assumption 1 which is a consequence of Assumption 10.
2. We proceed as for previous case but applying Theorem 1.3.10 instead of Proposition 1.3.9.

$\square$

We state now the most important results of the section.

**Theorem 1.4.15.** *Suppose  $b, \sigma$  are time-homogeneous, Assumption 10 and suppose there is  $\nu \in \mathbb{P} \mathbb{R}^d$  (a priori not known) such that  $\mu = \mathbf{u}^\nu(T)$ .*



1. (1.3) admits existence and uniqueness in law. Moreover existence in law holds in  $\mathbf{A}_1$ .
2. (1.3) admits pathwise uniqueness in  $\mathbf{A}_2$ .

*Proof.*

1. (a) First, Assumption 11 trivially holds since  $b, \sigma$  are time-homogeneous. Then, point 1 of Corollary 1.4.13 implies that (1.3) admits existence in law (in  $\mathbf{A}_1$ ) since Assumption 10 holds.
- (b) Let  $(Y, \mathbf{p})$  be a solution of (1.3). Proceeding as in the proof of item 2. of Corollary 1.4.13, we obtain that  $\mathbf{p}(T - \cdot) = \mathbf{u}^{v_0}$  with  $v_0 = \mathbf{p}(T)$ . Then, Lemma 1.4.11 and the fact that  $\sigma$  is bounded allow to show that  $\mathbf{p}$  belongs to  $\mathbf{A}_1$ , see (4.4) in Notation 1.
- (c) To conclude it remains to show uniqueness in law in  $\mathbf{A}_1$ . For this we wish to apply point 1. of Corollary 1.4.9. To achieve this, we check Assumption 9 with respect to  $\mu, \mathbf{P} \in \mathbf{R}^d$ . This is a consequence of Assumptions 3 and 5 and Theorem 1.3.13 This concludes the proof of item 1.

2. Concerning pathwise uniqueness in  $\mathbf{A}_2$ , we proceed as for uniqueness in law but applying point 2 of Corollary 1.4.9. This is valid since Assumption 10 implies that  $b, \sigma$  are bounded and Lipschitz. □

In the result below we extend Theorem 1.4.15 to the case when the coefficients  $b, \sigma$  are piecewise time-homogeneous.

**Theorem 1.4.16.** *Let  $n \in \mathbf{N}^*$ . Let  $0 = t_0 < \dots < t_n = T$  be a partition. For  $k \in \llbracket 2, n \rrbracket$  (resp.  $k = 1$ ) we denote  $I_k = ]t_{k-1}, t_k]$  (resp.  $[t_0, t_1]$ ). Suppose that the following holds.*

1. *For all  $k \in \llbracket 1, n \rrbracket$  the restriction of  $\sigma$  (resp.  $b$ ) to  $I_k \times \mathbf{R}^d$  is a time-homogeneous function  $\sigma^k : \mathbf{R}^d \rightarrow M_d(\mathbf{R})$  (resp.  $b^k : \mathbf{R}^d \rightarrow \mathbf{R}^d$ ).*
2. *Assumption 3.*
3.  *$\sigma$  is Lipschitz in space uniformly in time.*
4. *The functions  $\sigma^k, b^k, \nabla_x b_i^k$   $_{i \in \llbracket 1, d \rrbracket}, \nabla_x \Sigma_{ij}^k$   $_{i, j \in \llbracket 1, d \rrbracket}$  are continuous bounded and  $\nabla_x^2 \Sigma^k$  is Hölder continuous with exponent  $\alpha \in ]0, 1[$ .*

*Suppose  $\mu$  equals  $\mathbf{u}^v(T)$  for some  $v \in \mathbf{P}(\mathbf{R}^d)$ . Then equation (1.3) admits existence and uniqueness in law. Existence in law holds in  $\mathbf{A}_1$ .*

**Remark 1.4.17.** *A similar remark as in Corollary 1.4.13 holds for the Theorems 1.4.15 and 1.4.16. If there is no  $v \in \mathbf{P}(\mathbf{R}^d)$  such that  $\mathbf{u}^v(T) = \mu$ , then (1.3) does not admit existence in law.*

**Proof** of Theorem 1.4.16). We recall that by Lemma 1.3.4,  $\mathbf{u}^{v_0}$  is well-defined for all  $v_0 \in \mathbf{P}(\mathbf{R}^d)$ .

1. We first show that  $\mathbf{u}^{v_0}$  verifies (4.12) and (4.13). Indeed, fix  $k \in \llbracket 1, n \rrbracket$ . The restriction  $\mathbf{u}_k$  of  $\mathbf{u}^{v_0}$  to  $\bar{I}_k$  is a solution  $\mathbf{v}$  of the first line (3.2) replacing  $[0, T]$  with  $\bar{I}_k$ ,  $L$  by  $L^k$  defined in (3.25), with initial condition  $\mathbf{v}(t_{k-1}) = \mathbf{u}^{v_0}(t_{k-1})$ . That restriction is even the unique solution, using Lemma 1.3.4 replacing  $[0, T]$  with  $\bar{I}_k$ . We apply Lemma 1.4.11 replacing  $[0, T]$  with  $\bar{I}_k$ , taking into account Assumptions 10 and 11, which holds trivially replacing  $\sigma, b, \Sigma$  with  $\sigma^k, b^k, \Sigma^k$ . This implies that  $\mathbf{u}^{v_0}$  verifies (4.12) and (4.13) replacing  $[0, T]$  with  $\bar{I}_k$ , and therefore on the whole  $[0, T]$ .
2. Existence in law in  $\mathbf{A}_1$ , follows now by Lemma 1.4.12.
3. It remains to show uniqueness in law. Let  $(Y, \mathbf{p})$  be a solution of (1.3). We set  $v_0 := \mathbf{p}(T)$ . Since  $\mathbf{u}^{v_0}$  and  $\mathbf{p}(T - \cdot)$  solve (3.2), Lemma 1.3.4 implies that  $\mathbf{p}$  is uniquely determined. Similarly as in item 1.(b) of the proof of Theorem 1.4.15, item 1. of the present proof and Lemma 1.4.11 allow to show that  $\mathbf{p}$  belongs to  $\mathbf{A}_1$ .
4. It remains to show uniqueness in law in  $\mathbf{A}_1$ . For this, Corollary 1.3.16 implies Assumption 9 with  $\mathbf{C} = \mathbf{P}(\mathbb{R}^d)$ . Uniqueness of (1.3) in the class  $\mathbf{A}_1$  follows now by Corollary 1.4.9, which ends the proof.

□

### 1.4.5 Well-posedness for the McKean SDE: the Ornstein-Uhlenbeck semigroup

In this section we consider the case  $b : (s, x) \mapsto C(s)x$  with  $C$  continuous from  $[0, T]$  to  $\mathbb{R}^d$  and  $\sigma$  continuous from  $[0, T]$  to  $M_{d,m}(\mathbb{R})$ . We also suppose that for all  $t \in [0, T]$ ,  $\sigma(t)$  is invertible. We denote by  $\mathbf{C}(t) := \mathbf{D}(t)^{-1} \cdot^{-T}$ ,  $t \in [0, T]$ , where  $\mathbf{D}$  is the unique solution of (3.27). Evaluating the transpose on both sides of (3.28), we remark that  $\mathbf{C}$  is solution of the matrix-valued ODE

$$\mathbf{C}(t) = I + \int_0^t C(s)\mathbf{C}(s)ds, \quad t \in [0, T].$$

For a given  $x_0 \in \mathbb{R}^d$  and a given  $t \in ]0, T]$ , we denote by  $p_t^{x_0}$  the density of a Gaussian random vector with mean  $m_t^{x_0} = \mathbf{C}(t)x_0$  and covariance matrix  $Q_t = \mathbf{C}(t) \int_0^t \mathbf{C}^{-1}(s)\Sigma(s)\mathbf{C}^{-1}(s)^T ds \mathbf{C}(t)^T$ . Note that for all  $t \in ]0, T]$ ,  $Q_t$  is strictly positive definite, in particular it is invertible. Indeed, for every  $t \in [0, T]$ ,  $\Sigma(t)$  is strictly positive definite. By continuity in  $\int_0^t \mathbf{C}^{-1}(s)\Sigma(s) \mathbf{C}^{-1}(s)^T ds$  is also strictly positive definite and finally the same holds for  $Q_t$ . For a given  $\nu \in \mathbf{P}(\mathbb{R}^d)$ ,  $t \in ]0, T]$ , we set the notation

$$p_t^\nu : x \mapsto \int_{\mathbb{R}^d} p_t^{x_0}(x) \nu(dx_0). \tag{4.14}$$

At this level, we need a lemma.

**Lemma 1.4.18.** *Let  $\nu \in \mathbf{P}(\mathbb{R}^d)$ . The measure-valued function  $t \mapsto p_t^\nu(x)dx$  is the unique solution of (3.2) with initial value  $\nu$  and we denote it by  $\mathbf{u}^\nu$ . Furthermore,  $\mathbf{u}^\nu(T - \cdot)$  belongs to  $\mathbf{A}_2$ .*

*Proof.* 1. We denote immediately  $\mathbf{u}^\nu(t)(dx) := p^\nu(x_t)dx$ ,  $t \in ]0, T]$ . By Chapter 5, Section 5.6 in [69], for every  $t \in ]0, T]$ ,  $p^{x_0}$  is the density of the random variable  $X_t^{x_0}$ , where  $X^{x_0}$  is the unique strong solution of (3.4) with initial value  $x_0$ . The mapping  $t \mapsto p_t^{x_0}(x)dx$  is a solution of (3.2) by Proposition 1.3.2, with initial condition  $\delta_{x_0}$ . Consequently, by superposition,  $\mathbf{u}^\nu$  is a solution of (3.2) with initial value  $\nu$ .

2.  $\mathbf{u}^\nu$  is the unique solution of (3.2) because of Proposition 1.3.17.

3. It remains to show that  $\mathbf{u}^\nu(T - \cdot)$  belongs to  $\mathbf{A}_2$ , namely that for all  $i \in \llbracket 1, d \rrbracket$

$$(t, x) \mapsto \frac{\operatorname{div}_x \Sigma(T-t)_i p_{T-t}^\nu(x)}{p_{T-t}^\nu(x)},$$

is locally Lipschitz with linear growth in space on  $[0, T[ \times \mathbb{R}^d$ .

Fix  $i \in \llbracket 1, d \rrbracket$ ,  $t \in [0, T[$  and  $x \in \mathbb{R}^d$ . Remembering the fact,  $p_{T-t}^{x_0}$  is a Gaussian law with mean  $m_{T-t}^{x_0}$  and covariance matrix  $Q_{T-t}$  for a given  $x_0 \in \mathbb{R}^d$ , we have

$$\frac{\operatorname{div}_x \Sigma(T-t)_i p_{T-t}^\nu(x)}{p_{T-t}^\nu(x)} = - \frac{1}{p_{T-t}^\nu(x)} \Sigma(T-t)_i Q_{T-t}^{-1} (x - m_{T-t}^{x_0}) p_{T-t}^{x_0}(x) \nu(dx_0). \quad (4.15)$$

Let  $K$  be a compact subset of  $]0, T] \times \mathbb{R}^d$ ; then there is  $M_K > 0$  such that for all  $(t, x) \in K$ ,  $x_0 \in \mathbb{R}^d$ ,

$$|\Sigma(T-t)_i \cdot Q_{T-t}^{-1} (x - m_{T-t}^{x_0}) p_{T-t}^{x_0}(x)| \leq |\Sigma(T-t)_i| |Q_{T-t}^{-1} (x - m_{T-t}^{x_0}) p_{T-t}^{x_0}(x)| \leq M_K.$$

This follows because  $t \mapsto \Sigma(T-t)$  and  $t \mapsto Q_{T-t}^{-1}$  are continuous on  $[0, T[$  and, setting

$$c_K := \inf\{t \mid (t, x) \in K\}, \quad m_K := \sup_{a \in \mathbb{R}} |a| \exp -c_K \frac{a^2}{2},$$

we have

$$|x - m_{T-t}^{x_0}| p_{T-t}^{x_0}(x) \leq m_K, \quad \forall (t, x) \in K.$$

To show that left-hand side of (4.15) is locally bounded on  $[0, T[ \times \mathbb{R}^d$  it remains to show that  $(t, x) \mapsto \int_{\mathbb{R}^d} p_{T-t}^{x_0}(x) \nu(dx_0)$  is lower bounded on  $K$ . Indeed, let  $I$  be a compact of  $\mathbb{R}^d$ . Since  $(t, x, x_0) \mapsto p_{T-t}^{x_0}(x)$  is strictly positive and continuous is lower bounded by a constant  $c(K, I)$ . The result follows choosing  $I$  such that  $\nu(I) > 0$ .

To conclude, it remains to show that the functions  $(t, x) \mapsto \frac{\operatorname{div}_x(\Sigma(T-t)_i p_{T-t}^\nu(x))}{p_{T-t}^\nu(x)}$ ,  $i \in \llbracket 1, d \rrbracket$  defined on  $[0, T[ \times \mathbb{R}^d$  has locally bounded spatial derivatives, which implies that they are Lipschitz with linear growth on each compact of  $[0, T[ \times \mathbb{R}^d$ . By technical but easy computations, the result follows using the fact the real functions  $a \mapsto |a|^m \exp -\frac{a^2}{2}$ ,  $m = 1, 2$ , are bounded.  $\square$

We give now a global well-posedness result for equation (1.3).

**Theorem 1.4.19.** 1. Suppose the initial condition  $\mu$  equals  $\mathbf{u}^v(T)$  for some  $v \in \mathbf{P} \mathbb{R}^d$ . Then, equation (1.3) admits existence in law, strong existence, uniqueness in law and pathwise uniqueness.

2. Otherwise (1.3) does not admit any solution.

*Proof.* Item 2. can be proved using similar arguments as for the proof of Corollary 1.4.13. Let  $(Y, \mathbf{p})$  be a solution of (1.3) and set  $v_0 = \mathbf{p}(T)$ . By Proposition 1.4.3,  $\mathbf{p}(T - \cdot)$  is a solution of (2.3), so that  $\mathbf{p}(T - \cdot)$  verifies also (3.2) with initial value  $v_0$ . Since, by Proposition 1.3.17, uniqueness holds for (3.2), it follows that  $\mathbf{p}(T - \cdot) = \mathbf{u}^{v_0}$  which concludes the proof of item 2.

We prove now item 1. For this, taking into account Proposition 1.4.8 and Yamada-Watanabe theorem and related results for classical SDEs, it suffices to show strong existence and pathwise uniqueness.

We set  $\mathbf{C} := \mathbf{P} \mathbb{R}^d$

1. Concerning the strong existence statement, we want to apply Proposition 1.4.6. For this we have to check the validity of Assumption 1, Assumption 4 with respect to  $\mathbf{C}$  and Assumptions 6, 8 hold with respect to  $(\mu, \mathbf{C})$ .

Since  $b, \sigma$  are affine, Assumption 1 trivially holds. Furthermore, Assumption 4 holds with respect to  $\mathbf{C}$  thanks to Proposition 1.3.17.

Now,  $\mathbf{u}^v$  is a probability valued solution of (1.1) with terminal value  $\mu$ . Furthermore, Lemma 1.4.18 shows that  $\mathbf{u}^v$ , being the unique solution of solution of (3.2), is such that, for all  $t \in ]0, T]$ ,  $\mathbf{u}^v(t)$  admits  $p_t^v$  (see (4.14)) for density. Then, relation (4.5) holds since, by the considerations above (4.14)  $(t, x) \mapsto p_t^v(x)$  is locally bounded with locally bounded spatial derivatives. Hence, Assumption 6 holds with respect to  $(\mu, \mathbf{C})$ . Finally, Lemma 1.4.18 implies that  $\mathbf{u}^v(T - \cdot)$  belongs to  $\mathbf{A}_2$ . Hence, Assumption 8 holds with respect to  $(\mu, \mathbf{C})$ . At this point Proposition 1.4.6 implies existence in law.

2. Let  $(Y, \mathbf{p})$  be a solution of equation (1.3). Proposition 1.4.3 implies that  $\mathbf{p}(T - \cdot)$  solves (1.2). Then, Proposition 1.3.17 gives  $\mathbf{p}(T - \cdot) = \mathbf{u}^{v_0}$  with  $v_0 = \mathbf{p}(T)$ . Lemma 1.4.18 implies  $\mathbf{p}$  belongs to  $\mathbf{A}_2$ .
3. It remains to show pathwise uniqueness in  $\mathbf{A}_2$ . Assumption 9 holds with respect to  $(\mu, \mathbf{C})$  thanks to Theorem 1.3.19. Now, point 2 of Corollary 1.4.9 implies pathwise uniqueness in  $\mathbf{A}_2$  since  $b, \sigma$  are locally Lipschitz with linear growth in space.

□

## Appendix

### 1.4.6 Proof of Lemma 1.4.11

Let  $\nu \in \mathcal{P}(\mathbb{R}^d)$  For each given  $t \in [0, T]$ , we denote by  $G_t$  the differential operator such that for all  $f \in C^2(\mathbb{R}^d)$

$$G_t f = \frac{1}{2} \sum_{i,j=1}^d \partial_{ij} (\Sigma_{ij}(t, \cdot)) f - \sum_{i=1}^d \partial_i (b_i(t, \cdot)) f.$$

Assumption 10 implies that for a given  $f \in C^2(\mathbb{R}^d)$ ,  $G_t f$  can be rewritten in the two following ways:

$$G_t f = \frac{1}{2} \sum_{i,j=1}^d \Sigma_{ij}(t, \cdot) \partial_{ij} f + \sum_{i=1}^d \left( \sum_{j=1}^d \partial_i \Sigma_{ij}(t, \cdot) - b_i(t, \cdot) \right) \partial_i f + c^1(t, \cdot) f, \quad (4.16)$$

with

$$c^1 : (t, x) \mapsto \frac{1}{2} \sum_{i,j=1}^d \partial_{ij} \Sigma_{ij}(t, x) - \sum_{i=1}^d \partial_i b_i(t, x).$$

$$G_t f = \frac{1}{2} \sum_{i,j=1}^d \partial_j (\partial_i \Sigma_{ij}(t, \cdot)) f + \sum_{i,j=1}^d \Sigma_{ij}(t, \cdot) \partial_{ij} f - \sum_{i=1}^d b_i(t, \cdot) \partial_i f - \sum_{i=1}^d \partial_i b_i(t, \cdot) f. \quad (4.17)$$

On the one hand, combining identity (4.16) with Assumption 10, there exists a fundamental solution  $\Gamma$  (in the sense of Definition stated in Section 1. p.3 of [47]) of  $\partial_t u = G_t u$ , thanks to Theorem 10. Section 6 Chap. 1. in the same reference. Furthermore, there exists  $C_1, C_2 > 0$  such that for all  $i \in \llbracket 1, d \rrbracket$ ,  $x, \xi \in \mathbb{R}^d$ ,  $\tau \in [0, T]$ ,  $t > \tau$ ,

$$|\Gamma(x, t, \xi, \tau)| \leq C_1 (t - \tau)^{-\frac{d}{2}} \exp \left[ -\frac{C_2 |x - \xi|^2}{4(t - \tau)} \right], \quad (4.18)$$

$$|\partial_{x_i} \Gamma(x, t, \xi, \tau)| \leq C_1 (t - \tau)^{-\frac{d+1}{2}} \exp \left[ -\frac{C_2 |x - \xi|^2}{4(t - \tau)} \right], \quad (4.19)$$

thanks to identities (6.12), (6.13) in Section 6 Chap. 1 in [47].

On the other hand, combining Identity (4.17) with Assumption 10, there exists a weak fundamental solution  $\Theta$  of  $\partial_t u = G_t u$  thanks to Theorem 5 in [3]. In addition, there exists  $K_1, K_2, K_3 > 0$  such that for almost every  $x, \xi \in \mathbb{R}^d$ ,  $\tau \in [0, T]$ ,  $t \geq \tau$

$$\frac{1}{K_1} (t - \tau)^{-\frac{d}{2}} \exp \left[ -\frac{K_2 |x - \xi|^2}{4(t - \tau)} \right] \leq \Theta(x, t, \xi, \tau) \leq K_1 (t - \tau)^{-\frac{d}{2}} \exp \left[ -\frac{K_3 |x - \xi|^2}{4(t - \tau)} \right], \quad (4.20)$$

thanks to point (ii) of Theorem 10 in [3].

Our goal is now to show that  $\Gamma$  and  $\Theta$  coincide. To this end, we adapt the argument developed at the beginning of Section 7 in [3]. Fix a function  $H$  from  $[0, T] \times \mathbb{R}^d$  belonging to  $C_c^\infty([0, T] \times \mathbb{R}^d)$ . Identity (7.6) in Theorem 12 Chap 1. Section 1. of [47] implies in particular that the function

$$u : (t, x) \mapsto \int_0^t \int_{\mathbb{R}^d} \Gamma(x, t, \xi, \tau) H(\tau, \xi) d\xi d\tau,$$

is continuously differentiable in time, two times continuously differentiable in space and is a solution of the Cauchy problem

$$\begin{aligned} \square \partial_t u(t, x) &= G_t u(t, x) + H(t, x), \quad (t, x) \in ]0, T] \times \mathbb{R}^d \\ \square u(0, \cdot) &= 0. \end{aligned} \quad (4.21)$$

It is consequently also a weak (i.e. distributional) solution of (4.21), which belongs to  $E^2(]0, T] \times \mathbb{R}^d)$  (see definition of that space in [3]) since  $u$  is bounded thanks to inequality (4.18) and the fact that  $H$  is bounded. Then, point (ii) of Theorem 5 in [3] says that

$$(t, x) \mapsto \int_0^t \int_{\mathbb{R}^d} \Theta(x, t, \xi, \tau) H(\tau, \xi) d\xi d\tau$$

is the unique weak solution in  $E^2(]0, T] \times \mathbb{R}^d)$  of (4.21). This implies that for every  $(t, x) \in ]0, T] \times \mathbb{R}^d$  we have

$$\int_0^t \int_{\mathbb{R}^d} (\Gamma - \Theta)(x, t, \xi, \tau) H(\tau, \xi) d\xi d\tau = 0.$$

Point (i) of Theorem 5 in [3] (resp inequality (4.18)) implies that  $\Theta$  (resp.  $\Gamma$ ) belongs to  $L^p(]0, T] \times \mathbb{R}^d)$  as a function of  $(\xi, \tau)$ , for an arbitrary  $p \geq d + 2$ . Then, we conclude that for all  $(t, x) \in ]0, T] \times \mathbb{R}^d$ ,

$$\Theta(x, t, \xi, \tau) = \Gamma(x, t, \xi, \tau), \quad d\xi d\tau a.e. \quad (4.22)$$

for all  $(\tau, \xi) \in [0, t] \times \mathbb{R}^d$ . This happens by density of  $C_c^\infty([0, T] \times \mathbb{R}^d)$  in  $L^q(]0, T] \times \mathbb{R}^d)$ ,  $q$  being the conjugate of  $p$ .

This, together with (4.20) and the fact that  $\Gamma$  is continuous in  $(\tau, \xi)$  implies that (4.20) holds for all  $(\tau, \xi) \in [0, t] \times \mathbb{R}^d$  and therefore

$$\frac{1}{K_1} (t - \tau)^{-\frac{d}{2}} \exp\left[-\frac{K_2 |x - \xi|^2}{4(t - \tau)}\right] \leq \Gamma(x, t, \xi, \tau) \leq K_1 (t - \tau)^{-\frac{d}{2}} \exp\left[-\frac{K_3 |x - \xi|^2}{4(t - \tau)}\right]. \quad (4.23)$$

We introduce

$$q_t := x \mapsto \int_{\mathbb{R}^d} \Gamma(x, t, \xi, 0) \nu(d\xi).$$

By (4.23), with  $\tau = 0$  we get

$$q_t(x) \geq \frac{1}{K_1} t^{-\frac{d}{2}} \int_{\mathbb{R}^d} \exp\left[-\frac{K_2 |x - \xi|^2}{4t}\right] \nu(d\xi). \quad (4.24)$$

We denote now by  $\mathbf{v}^\nu$  the measure-valued mapping such that  $\mathbf{v}^\nu(0, \cdot) = \nu$  and for all  $t \in ]0, T]$ ,  $\mathbf{v}^\nu(t)$  has density  $q_t$  with respect to the Lebesgue measure on  $\mathbb{R}^d$ . We want to show that  $\mathbf{v}^\nu$  is a solution of (3.2) with initial value  $\nu$  to conclude  $\mathbf{u}^\nu = \mathbf{v}^\nu$  thanks to the validity of Assumption 4 because of Remark 1.3.3 1. and 3. To this end, we remark that the definition of a fundamental solution for  $\partial_t u = G_t u$  says that  $u$  is a  $C^{1,2}$  solution and consequently also a solution in the sense of distributions. In particular for all  $\varphi \in C_c^\infty(\mathbb{R}^d)$ , for all  $t \geq E > 0$

$$\int_{\mathbb{R}^d} \varphi(x) \mathbf{v}^\nu(t)(dx) = \int_{\mathbb{R}^d} \varphi(x) \mathbf{v}^\nu(E)(dx) + \int_E^t \int_{\mathbb{R}^d} L_s \varphi(x) \mathbf{v}^\nu(s)(dx) ds. \quad (4.25)$$

To conclude, it remains to send  $E$  to  $0^+$ . Theorem 15 section 8. Chap 1. and point (ii) of the definition stated p. 27 in [47] imply in particular that for all  $\varphi \in C_c^\infty \mathbb{R}^d$ ,  $\xi \in \mathbb{R}^d$ ,

$$\int_{\mathbb{R}^d} \Gamma(x, E, \xi, 0) \varphi(x) dx \xrightarrow{E \rightarrow 0^+} \varphi(\xi).$$

Fix now  $\varphi \in C_c^\infty \mathbb{R}^d$ . In particular thanks to Fubini's theorem, (4.20) and Lebesgue's dominated convergence theorem we have

$$\begin{aligned} \int_{\mathbb{R}^d} \varphi(x) \mathbf{v}^\nu(E)(dx) &= \int_{\mathbb{R}^d} \varphi(x) \int_{\mathbb{R}^d} \Gamma(x, E, \xi, 0) \nu(d\xi) dx \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \Gamma(x, E, \xi, 0) \varphi(x) dx \nu(d\xi) \\ &\xrightarrow{E \rightarrow 0^+} \int_{\mathbb{R}^d} \varphi(\xi) \nu(d\xi). \end{aligned}$$

By (4.25)  $\mathbf{v}^\nu$  is a solution of (3.2) and consequently  $\mathbf{u}^\nu = \mathbf{v}^\nu$ , so that, for every  $t \in ]0, T]$ ,  $\mathbf{u}^\nu(t)$  admits  $u^\nu(t, \cdot) = q_t$  for density with respect to the Lebesgue measure on  $\mathbb{R}^d$ . Now, integrating the inequalities (4.18), (4.19) with respect to  $\nu$  and combining this with inequality (4.24), we obtain the existence of  $K_1, K_2, C_1, C_2 > 0$  such that for all  $t \in ]0, T]$ , for all  $x \in \mathbb{R}^d$ , for all  $i \in \llbracket 1, d \rrbracket$

$$\begin{aligned} \frac{1}{K_1} t^{-\frac{d}{2}} \int_{\mathbb{R}^d} \exp\left[-\frac{K_2 |x - \xi|^2}{4t}\right] \nu(d\xi) &\leq u^\nu(t, x) \leq K_1 t^{-\frac{d}{2}}, \\ |\partial_i u^\nu(t, x)| &\leq C_1 t^{-\frac{d+1}{2}}. \end{aligned}$$

Consequently, the upper bounds in (4.12) and (4.13) hold. Concerning the lower bound in (4.12), let  $I$  be a compact subset of  $\mathbb{R}^d$  such that  $\nu(I) > 0$ , the result follows since  $(t, x, \xi) \mapsto \exp\left[-\frac{K_2 |x - \xi|^2}{4t}\right]$  is strictly positive, continuous and therefore lower bounded by a strictly positive constant on  $K \times I$  for each compact  $K$  of  $]0, T] \times \mathbb{R}^d$ .

## Chapter 2

# A fully backward representation of semilinear PDEs applied to the control of thermostatic loads in power systems

This chapter is the object of the paper [60].

### 2.1 Introduction

The numerical resolution of non-linear PDEs is a crucial issue in many applications. In particular, stochastic control problems can be formulated by mean of the Hamilton-Jacobi-Bellman (HJB) equations with terminal condition. In this paper, we focus more particularly on control problems raised by demand-side management in power systems. The difficulties come especially from the high dimensionality of the state space, which motivates the use of probabilistic representations. The main issue of numerical schemes is then to concentrate the computing effort in specific regions of interest in the state space. In classical regression Monte-Carlo approaches, the solution is evaluated backwardly in time from the final time to the initial time, while the regression grid is generated forwardly from the initial time to the final one. In this paper, we propose a fully backward probabilistic approach which allows to generate adaptively the regression grid, as the solution is evaluated, taking advantage of the calculations already performed. Besides, there is no need to store the entire grid, since the points are generated as they are used for calculations. Our grid will be indeed simulated according to the time-reversal of some diffusion starting from a judicious terminal distribution.

We are interested in semilinear PDEs of the type

$$\begin{aligned} \partial_t v(t, x) + H(t, x, v(t, x), \nabla_x v(t, x)) + \frac{1}{2} \text{Tr}[\sigma \sigma^{-T}(t) \nabla_x^2 v(t, x)] &= 0, \quad (t, x) \in [0, T] \times \mathbb{R}^d \\ v(T, x) &= g(x), \end{aligned} \quad (1.1)$$

where in particular  $\sigma$  is a deterministic non-degenerate matrix-valued function. Under suitable conditions, there exists a unique viscosity solution  $v$  of (1.1) in the class of continuous functions with



polynomial growth. One classical probabilistic representation of  $v$  is provided by Forward-Backward SDEs (FBSDEs), see e.g. [91]. First a forward diffusion is fixed, with an arbitrary drift  $\tilde{b}$

$$dX_t = b(t, X_t)dt + \sigma(t)dW_t. \tag{1.2}$$

Then the solution of (1.1) is represented by  $v(s, x) = \mathbb{Y}^{s,x}$ , where  $(Y, Z) = (Y^{s,x}, Z^{s,x})$  is the unique solution of the BSDE

$$Y_t = g(X_T) + \int_t^T F(r, X_r, Y_r, Z_r)dr - \int_t^T Z_r dW_r, \tag{1.3}$$

with  $X = X^{s,x}$  being the solution of (1.2) starting at time  $s$  with value  $x$  and  $F$  being related to  $H$  by

$$F(t, x, y, z) := H(t, x, y, \sigma^{-1}(t)^{-T} z) - b(t, x), \sigma^{-1}(t)^{-T} z. \tag{1.4}$$

Considering a time discretization mesh  $t_k = k\delta t$ , with  $\delta t = \frac{T}{n}$  and  $k = 0, \dots, n$ , for a given positive integer  $n$ , [48] proved that one can approximate  $(Y_{t_k}, Z_{t_k})$  by  $(\hat{Y}_k, \hat{Z}_k)$  such that  $\hat{Y}_n = g(X_T)$  and for  $k = 0, \dots, n - 1$

$$\begin{aligned} \square \hat{Y}_k &= \mathbb{E} \int_{t_{k+1}}^n F(t, X_t, \hat{Y}, \hat{Z}_{-1})\delta t + g(X_T) X_{t_k} \\ \square \hat{Z}_k &= \frac{1}{\delta t} \mathbb{E} \hat{Y}_{k+1} (W_{t_{k+1}} - W_{t_k}) X_{t_k}. \end{aligned} \tag{1.5}$$

Most of probabilistic numerical schemes (see e.g. regression Monte-Carlo [50, 20], Kernel Monte-Carlo [25], Quantization [38]) rely on that representation. The common idea is then articulated in two steps. First, one generates a grid discretizing the forward process (1.2) in space and time on  $[0, T]$ , (by Monte-Carlo simulations or Quantization, etc.). Then, one calculates the conditional expectations (1.5) on the grid points in order to estimate  $(\hat{Y}, \hat{Z})$ . These techniques have generally two limitations.

1. The degree of freedom in the choice of the forward diffusion  $X$  is difficult to exploit although it has a major impact on the numerical scheme efficiency: how to chose a reasonable drift  $\tilde{b}$  without a priori information on  $v$  ?
2. The entire grid discretizing the forward process has to be stored in memory to be revisited backwardly in time in order to compute the solution process  $(Y, Z)$ . This approach naturally raises some huge memory issues which in general limit drastically the number of Monte-Carlo runs and time steps, hence the accuracy of the procedure.

To overcome such limitations some approaches were proposed in the domain of mathematical finance, in particular for the evaluation of American style options. One technique, intended to deal with the memory problem, relies on bridge simulation, see e.g. [97, 101]. However, this approach requires specific developments for each price model (based for instance on the Brownian bridge for Brownian prices or on the gamma bridge for variance gamma prices) and remains difficult to generalize to a wide class of models. To address the efficiency issue, [18] developed a scheme based on Picard's type iterations that avoids the use of nested conditional expectations backwardly in time,

which are replaced by nested conditional expectations along the iterations. In the same line, [49] proposes an adaptive variance reduction technique which combines Picard's iterations and control variate to solve the BSDE. A parallel version of that algorithm was proposed in [73]. However, those approaches require, at each iteration, to approximate the solution on the whole time horizon. Similarly, importance sampling and Girsanov's theorem, were considered to force the exploration of the space towards areas of interest [19]. In particular, this type of approach was derived in the case of stochastic control in [41] providing an iterative scheme that is capable of learning the optimally controlled drift. Here again, that method requires several estimations of the value function on the whole time horizon. Besides [52] proposed an adaptive importance sampling scheme for FBSDEs allowing to select the drift adaptively, as the calculations are performed backwardly. Unfortunately, that approach is limited to situations where the driver  $F$  does not depend on  $Z$ . In the present paper, we introduce a new adaptive approach to address both the memory problem and the efficiency issue (related to the drift selection) in the general case where the driver may depend on  $X, Y$  and  $Z$ .

We propose to choose adaptively the drift  $\tilde{b}$  at the same time as we discover the function  $v$  such that

$$v(t, X_t) = \mathbf{E} \int_t^T H(s, X_s, v(s, X_s), \nabla_x v(s, X_s)) - b(s, X_s), \nabla_x v(s, X_s) ds + g(X_T) | X_t, \quad (1.6)$$

by simulating the time-reversal of a solution  $X$  of (1.2) starting from the distribution of  $X_T$ . More specifically, to take advantage of the Ornstein-Uhlenbeck setting, we choose the drift  $\tilde{b}$  to be affine w.r.t. the space variable. We fix a Gaussian distribution  $\nu$  and look for solutions  $\xi$  of the McKean SDE

$$\begin{aligned} & \square \xi_0 \sim \nu, \\ & \square \xi_t = \xi_0 - \int_0^t b(T-s, \xi_s) + \sigma \sigma^\top (T-s) Q(T-s)^{-1} (\xi_s - m(T-s)) ds + \int_0^t \sigma(T-s) d\beta_s, \\ & m(T-t) = \mathbf{E}(\xi_t), \\ & \square Q(T-t) = \text{Cov}(\xi_t) \quad \text{for } t \in ]0, T]. \end{aligned} \quad (1.7)$$

By Proposition 2.3.7, (1.7) admits exactly one solution  $\xi$ , provided Assumption 12 in Section 2.3.1 is verified. That assumption depends on the covariance matrix of  $\nu$ , the drift  $\tilde{b}$  and the volatility  $\sigma$ . Indeed, one important limitation is that the covariance matrix should be chosen carefully to ensure that the process is well-defined until  $T$ . Point 2. of Proposition 2.3.7 and Lemma 2.3.6 say that the time-reversal process  $\hat{\xi}$ , i.e.  $\hat{\xi}_t := \xi_{T-t}$ , is an Ornstein-Uhlenbeck process solution of (1.2) such that the law of  $X_0$  is Gaussian with mean  $m(0)$  and covariance  $Q(0)$ . This leads to the first result of this paper which consists of the fully backward representation stated in Theorem 2.3.10. The proof is based on Feynman-Kac type formula instead of BSDEs and it does not require explicitly the uniqueness of viscosity solution of the PDE (1.1). The second contribution of the paper is Corollary 2.4.4 which is the "instantiation" of Theorem 2.3.10 in the framework of stochastic control, i.e. the representation of its value function (solution of a Hamilton-Jacobi-Bellman equation). This holds when the running and terminal cost have polynomial growth with respect to the state space variable. We also suppose that the value function is of class  $C^{0,1}$  whose gradient has polynomial growth. In particular, we derive in

Corollary 2.4.6, a representation involving the gap between the optimally controlled drift and the instrumental drift  $\tilde{b}$ . In Section 2.5, we present a fully backward Monte-Carlo regression scheme, where the instrumental drift is adaptively updated in order to mimic the optimally controlled dynamics, see Algorithm 1. We expect that this approach is particularly well-suited when the final cost has a strong impact on the global cost and when the terminal cost function is localized in a small region of the space, so that the initial distribution  $\nu$  can be chosen in an appropriate way. Finally, in Section 2.6 we illustrate the interest of this new algorithm applied to the problem of controlling the consumption of a large number of thermostatic loads in order to minimize an aggregative cost. We compare our approach to the classical regression Monte-Carlo scheme based on a forward grid.

## 2.2 Notations

Let us fix  $T > 0, d, k \in \mathbf{N}^*$ . For a given  $p \in \mathbf{N}^*$ ,  $[[1, p]]$  denotes the set of all integers between 1 and  $p$  included.  $\cdot, \cdot$  denotes the usual scalar product on  $\mathbf{R}^d$  and  $|\cdot|$  the associated norm. Elements of  $\mathbf{R}^d$  are supposed to be column vectors.  $M_d(\mathbf{R})$  stands for the set of  $d \times d$  matrices,  $S_d(\mathbf{R})$  for the subset of symmetric matrices,  $S_d^+(\mathbf{R})$  the subset of symmetric positive semi-definite matrices (in particular with non-negative eigenvalues) and  $S_d^{++}(\mathbf{R})$  for the subset of strictly positive definite symmetric matrices. For a given  $A \in M_d(\mathbf{R})$ ,  $A^T$  will denote its transpose,  $Tr(A)$  its trace,  $Sp(A)$  its spectrum, i.e. the set of its eigenvalues,  $e^A := \sum_{k=0}^{\infty} \frac{A^k}{k!}$  its exponential and  $\|A\| := \sup_{x \in \mathbf{R}^d, |x|=1} |Ax|$ . For a given  $A \in S_d^+(\mathbf{R})$ ,  $\sqrt{A}$  denotes the unique element of  $S_d^+(\mathbf{R})$  such that  $(\sqrt{A})^2 = A$ .

For a given continuous function  $f : [0, T] \mapsto \mathbf{R}^d$  (resp.  $g : [0, T] \mapsto M_d(\mathbf{R})$ ), we set  $\|f\|_{\infty} := \sup_{t \in [0, T]} |f(t)|$  (resp.  $\|g\|_{\infty} := \sup_{t \in [0, T]} \|g(t)\|$ ).  $C^{1,2}([0, T], \mathbf{R}^d)$  (resp.  $C^{0,1}([0, T], \mathbf{R}^d)$ ) denotes the set of real-valued functions defined on  $[0, T] \times \mathbf{R}^d$  being continuously differentiable in time and twice continuously differentiable in space (resp. continuous in time and continuously differentiable in space).  $C^0([0, T] \times \mathbf{R}^d)$  (resp.  $C^1(\mathbf{R}^d)$ ) denotes the set of continuous (resp. continuously differentiable) real-valued functions defined on  $[0, T] \times \mathbf{R}^d$  (resp.  $\mathbf{R}^d$ ).  $\nabla_x$  will denote the gradient operator and  $\nabla_x^2$  the Hessian matrix. For each  $p \in \mathbf{N}$ ,  $P_p \mathbf{R}^d$  denotes the set of polynomial functions on  $\mathbf{R}^d$  with degree  $p$ .

In the whole paper, we say that a function  $v : [0, T] \times \mathbf{R}^d \mapsto \mathbf{R}$  has *polynomial growth* if there exists  $q, K > 0$  such that for all  $(t, x) \in [0, T] \times \mathbf{R}^d$

$$|v(t, x)| \leq K(1 + |x|^q).$$

When  $v$  verifies previous property with  $q = 1$ , we say that it has *linear growth*.

For a given random vector  $X$  defined on a probability space  $(\Omega, \mathcal{F}, \mathbf{P})$ ,  $\mathbf{E}_{\mathbf{P}}(X)$  (resp.  $\text{COV}_{\mathbf{P}}(X) := \mathbf{E}_{\mathbf{P}}(X - \mathbf{E}_{\mathbf{P}}(X))(X - \mathbf{E}_{\mathbf{P}}(X))^T$ ) will denote its expectation (resp. its covariance matrix) under  $\mathbf{P}$ . When self-explanatory, the subscript will be omitted in the sequel. For a given  $(m, Q) \in \mathbf{R}^d \times S_d^+(\mathbf{R})$ ,  $\mathbf{N}(m, Q)$  denotes the Gaussian probability on  $\mathbf{R}^d$  with mean  $m$  and covariance matrix  $Q$ .

For any stochastic process  $X$ ,  $\mathbb{F}^X$  will denote its canonical filtration.  $\bar{X}$  will denote the time-reversal process  $X_{T-\cdot}$ .

## 2.3 Representation of semilinear PDEs

### 2.3.1 Around two backward ODEs

Let  $a$  (resp.  $c$ ) be Borel bounded functions from  $[0, T]$  to  $M_d(\mathbb{R})$  (resp.  $\mathbb{R}^d$ ).

In the sequel we will fix a Gaussian Borel probability  $\nu$  on  $\mathbb{R}^d$  with mean  $\bar{m}^\nu$  and covariance matrix  $\bar{Q}^\nu$ . We consider the functions  $m^\nu : [0, T] \rightarrow \mathbb{R}^d$  and  $Q^\nu : [0, T] \rightarrow S_d(\mathbb{R})$  denoting respectively the unique solutions of the backward ODEs

$$\begin{aligned} \square \frac{d}{dt} m^\nu(t) &= a(t) m^\nu(t) + c(t), \quad t \in [0, T] \\ \square m^\nu(T) &= \bar{m}^\nu, \end{aligned} \tag{3.1}$$

$$\begin{aligned} \square \frac{d}{dt} Q^\nu(t) &= Q^\nu(t) a(t)^{-T} + a(t) Q^\nu(t) + \Sigma(t), \quad t \in [0, T] \\ \square Q^\nu &= \bar{Q}^\nu \end{aligned} \tag{3.2}$$

for which existence and uniqueness hold since they are linear.

We introduce an hypothesis on  $\nu$  which will be used in the sequel.

**Assumption 12.**  $Q^\nu(0) \in \mathbb{S}_d^+(\mathbb{R})$ .

Easy computations imply for all  $t \in [0, T]$

$$m^\nu(t) = \mathbf{A}(t) \mathbf{A}(T)^{-1} \bar{m}^\nu - \int_t^T \mathbf{A}(s)^{-1} c(s) ds, \tag{3.3}$$

$$Q^\nu(t) = \mathbf{A}(t) \mathbf{A}(T)^{-1} \bar{Q}^\nu \mathbf{A}(T)^{-T} - \int_t^T \mathbf{A}(s)^{-1} \Sigma(s) \mathbf{A}(s)^{-T} ds \mathbf{A}(t)^{-T}, \tag{3.4}$$

where  $\mathbf{A}(t), t \in [0, T]$  is the unique solution of the matrix ODE

$$\begin{aligned} \square \frac{d}{dt} \mathbf{A}(t) &= a(t) \mathbf{A}(t), \quad t \in [0, T] \\ \square \mathbf{A}(0) &= I_d. \end{aligned} \tag{3.5}$$

We recall that for all  $t \in [0, T]$ ,  $\mathbf{A}(t)$  is invertible and the matrix valued function  $t \mapsto \mathbf{A}(t)^{-1}$  solves the ODE

$$\begin{aligned} \square \frac{d}{dt} \mathbf{A}(t)^{-1} &= -\mathbf{A}(t)^{-1} a(t), \quad t \in [0, T] \\ \square \mathbf{A}(0)^{-1} &= I_d \end{aligned} \tag{3.6}$$

see Chapter 8 in [28] for similar and further properties.

Note that in the case  $a(t) = a$ ,  $t \in [0, T]$  for a given  $a \in M_d(\mathbb{R})$ , then  $\mathbf{A} : t \rightarrow e^{at}$  and identities (3.3), (3.4) simplify as follows:

$$m^v(t) = e^{-a(T-t)} \bar{m}^v - \int_t^T e^{-a(s-t)} c(s) ds, \quad (3.7)$$

$$Q^v(t) = e^{-a(T-t)} \bar{Q}^v - \int_t^T e^{-a(s-t)} \Sigma(s) e^{-a^\top(s-t)} ds, \quad (3.8)$$

for all  $t \in [0, T]$ .

**Remark 2.3.1.** Suppose that  $Q^v(0)$  belongs to  $S_d^+(\mathbb{R})$ . Identity (3.4) gives in particular

$$Q^v(t) = \mathbf{A}(t) Q^v(0) + \int_0^t \mathbf{A}(s)^{-1} \Sigma(s) \mathbf{A}(s)^{-1} ds \mathbf{A}(t)^{-\top}, \quad t \in [0, T]. \quad (3.9)$$

Combining (3.9) and the fact  $\sigma(t)$  is invertible for all  $t \in [0, T]$ , we remark that  $Q^v(t)$  belongs to  $S_d^{++}(\mathbb{R})$  for all  $t \in ]0, T]$ .

Finally we give a condition depending on  $\mathbf{A}$ ,  $\sigma$ ,  $\bar{Q}^v$  and  $T$  to ensure the measure  $\nu$  fulfills Assumption 12.

**Proposition 2.3.2.** Suppose that

$$\min Sp Q^{-\nu} \geq \int_0^T \|\sigma(s)\|^{-2} \mathbf{A}(T) \mathbf{A}(s)^{-1} ds. \quad (3.10)$$

Then,

$$Q^v(0) \in S_d^+(\mathbb{R}). \quad (3.11)$$

*Proof.* Since  $\mathbf{A}(T)$  is invertible and  $Q^v(0)$  belongs to  $S_d(\mathbb{R})$ , (3.11) is equivalent to

$$\mathbf{A}(T) Q^v(0) \mathbf{A}(T)^{-\top} \in S_d^+(\mathbb{R}). \quad (3.12)$$

To prove (3.12), taking into account (3.4), it suffices to show that the matrix

$$\bar{Q}^v - \int_0^T \mathbf{A}(T) \mathbf{A}(s)^{-1} \Sigma(s) \mathbf{A}(T) \mathbf{A}(s)^{-1} ds \in S_d^+(\mathbb{R}),$$

or, equivalently, that for all  $x \in \mathbb{R}^d$

$$\lambda := x^{-\top} \bar{Q}^v x - \int_0^T x^{-\top} \mathbf{A}(T) \mathbf{A}(s)^{-1} \Sigma(s) \mathbf{A}(T) \mathbf{A}(s)^{-1} x ds \geq 0. \quad (3.13)$$

Let  $x \in \mathbb{R}^d$ ,

$$\begin{aligned} \lambda &\geq \min Sp \bar{Q}^v |x|^2 - \int_0^T \sigma(s)^{-\top} \mathbf{A}(T) \mathbf{A}(s)^{-1} x^2 ds, \\ &\geq \min Sp Q^{-\nu} - \int_0^T \sigma(s)^{-\top} \mathbf{A}(T) \mathbf{A}(s)^{-1} ds |x|^2, \\ &\geq 0, \end{aligned}$$

since (3.10) holds. This ends the proof. □

**Remark 2.3.3.** In the case  $a(t) = a$ ,  $t \in [0, T]$  for a given  $a \in M_d(\mathbb{R})$ , Condition (3.10) is satisfied in particular if

$$\min Sp \bar{Q}^\nu \geq \sigma^{-T} \int_0^T e^{\alpha^\top s} ds \quad (3.14)$$

is verified.

**Remark 2.3.4.** Let  $X$  be a solution of

$$X_t = X_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s) dW_s, \quad t \in [0, T], \quad (3.15)$$

where  $\sigma$  is a deterministic matrix-valued function and  $b$  the piecewise affine function

$$b(t, x) = a(t)x + c(t), \quad t \in [0, T],$$

and  $X_0$  be a square integrable r.v. It is well-known that  $X$  is a square integrable process. Let, for every  $t \in [0, T]$ ,  $m(t) = \mathbf{E}(X_t)$  and  $Q(t)$  the covariance matrix of  $X_t$ . Setting  $\bar{m}^\nu = \mathbf{E}(X_T)$  and  $\bar{Q}^\nu$  the covariance matrix of  $X_T$ . Then

$$m = m^\nu, \quad Q = Q^\nu. \quad (3.16)$$

Indeed, by Problem 6.1 in Chapter 5 in [69]  $m$  (resp.  $Q$ ) is solution of (3.1) (resp. (3.2)). (3.16) follows by uniqueness of previous ODEs.

### 2.3.2 The representation formula for a general semilinear PDE

In the whole paper  $\sigma$  will be a continuous function defined on  $[0, T]$  with values in  $M_d(\mathbb{R})$  such that for all  $t \in [0, T]$ ,  $\sigma(t)$  is invertible. We will set  $\Sigma := \sigma\sigma^{-T}$ .

Let  $b : [0, T] \times \mathbb{R}^d \mapsto \mathbb{R}^d$  and  $b_c : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times S_d^{++}(\mathbb{R}) \mapsto \mathbb{R}^d$  defined by

$$b_c : (t, x, m, Q) \mapsto \Sigma(t)Q^{-1}(x - m), \quad b : (t, x) \mapsto a(t)x + c(t), \quad (3.17)$$

where  $a, c$  were defined at Section 2.3.1. Let  $H : [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}$  and  $g : \mathbb{R}^d \rightarrow \mathbb{R}$ . The goal of this subsection is to provide a probabilistic representation of viscosity solutions, being continuous in time and continuously differentiable in space, of the semilinear PDE

$$\begin{aligned} \partial_t v(t, x) + \frac{1}{2} \text{Tr} \Sigma(t) \nabla_x^2 v(t, x) + H(t, x, v(t, x), \nabla_x v(t, x)) &= 0, \quad (t, x) \in [0, T] \times \mathbb{R}^d \\ v(T, \cdot) &= g. \end{aligned} \quad (3.18)$$

To formulate the result we consider the following assumption.

**Assumption 13.**  $g$  is continuous and has polynomial growth.

Let  $\nu$  be a Gaussian Borel probability on  $\mathbb{R}^d$  with mean  $\bar{m}^\nu$  and covariance  $\bar{Q}^\nu$ . Let  $t \mapsto m^\nu(t)$  defined in (3.3),  $t \mapsto Q^\nu(t)$  be given by (3.4) and suppose that  $\nu$  fulfills Assumption 12.

We fix a filtered probability space  $\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P}$  on which are defined a  $d$ -dimensional Brownian motion  $\beta$  and a random vector  $\xi_0$  distributed according to  $\nu$  and independent of  $\beta$ .

Let  $\xi$  be the unique strong solution of

$$\xi_t = \xi_0 - \int_0^t b(T-s, \xi_s) + b_c(T-s, \xi_s, m^\nu(T-s), Q^\nu(T-s)) ds + \int_0^t \sigma(T-s) d\beta_s, \quad t \in [0, T[. \quad (3.19)$$

**Remark 2.3.5.** (3.19) admits a unique strong solution on  $[0, T[$  since its drift is affine with time-dependent continuous coefficients.

**Lemma 2.3.6.** 1. The process  $\xi := \xi_{T-}$  solves the SDE

$$X_t = X_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s) dW_s, \quad t \in [0, T[, \quad (3.20)$$

where  $W$  is an  $\mathbb{F}^\xi$ -Brownian motion and  $X_0 \sim \mathbf{N}(m^\nu(0), Q^\nu(0))$ .

2.  $\xi$  extends continuously to  $[0, T]$ .

*Proof.* 1. The SDE (3.20) admits in particular existence in law. Let  $X$  be a solution of (3.20). To prove the first statement, it suffices to show that the laws of  $\xi$  and  $X$  coincide. For this it is enough to prove that  $X = X_{T-}$  and the solution  $\xi$  of (3.19) are identically distributed. By Problem 6.1 in Chapter 5 in [69] and by uniqueness of the ODE (3.1) (resp. (3.2)) with initial condition  $m^\nu(0)$  (resp.  $Q^\nu(0)$ ), we get  $\mathbf{E}(X_t) = m^\nu(t)$  and  $\text{Cov}(X_t) = Q^\nu(t)$  for all  $t \in [0, T]$ . By Problem 6.2, Chapter 5 in [69])  $X$  is a Gaussian process so

$$\xi_t \sim \mathbf{N}(m^\nu(t), Q^\nu(t)), \quad t \in [0, T]. \quad (3.21)$$

By (3.21) and Theorem 2.1 in [55],  $X$  is a solution (in law) of (3.19) on  $[0, T[$ . Pathwise uniqueness for (3.19) implies uniqueness in law on  $[0, T[$  and the first statement of Lemma 2.3.6 is established.

2. It remains to prove the second statement. For this we show

$$\mathbf{E} \int_0^T b_c(s, \xi_s, m^\nu(s), Q^\nu(s)) ds < \infty. \quad (3.22)$$

On the one hand, for all  $t \in ]0, T]$ ,

$$\begin{aligned} b_c(t, \xi_t, m^\nu(t), Q^\nu(t)) &= \Sigma(t) \begin{matrix} \mathbf{J} & \mathbf{J} \\ \hline Q^\nu(t)^{-1} & Q^\nu(t)^{-1} \end{matrix} \xi_t - m^\nu(t) \\ &\leq \|\Sigma\|_\infty \begin{matrix} \mathbf{J} \\ \hline Q^\nu(t)^{-1} & Q^\nu(t)^{-1} \end{matrix} \xi_t - m^\nu(t) \\ &= \frac{\|\Sigma\|_\infty}{\|Q^\nu(t)\|} \begin{matrix} \mathbf{J} \\ \hline Q^\nu(t)^{-1} \end{matrix} \xi_t - m^\nu(t), \end{aligned}$$

remembering that  $Q^\nu(t)$  belongs to  $S_d^{++}(\mathbb{R})$ .

On the other hand, by (3.21)  $Q^v(t)^{-1} \xi_t - m^v(t) \sim |Z|$  where  $Z \sim \mathbf{N}(0, I_d)$ . Then, (3.22) is verified if we show

$$\int_0^T \frac{1}{\|Q^v(t)\|} dt < \infty. \tag{3.23}$$

If  $Q^v(0) = 0$ , then for all  $t \in ]0, T]$ , for all  $t \in ]0, T]$ , Remark 2.3.1 implies

$$\frac{Q^v(t)}{t} = \mathbf{A}(t) \frac{1}{t} \int_0^t \mathbf{A}(s)^{-1} \Sigma(s) \mathbf{A}(s)^{-1} ds \mathbf{A}(t)^{-T} \xrightarrow{t \rightarrow 0} \Sigma(0).$$

If  $Q^v(0) \neq 0$ , then for all  $]0, T]$ , again Remark 2.3.1 yields

$$\frac{\|Q^v(t)\|}{t} \geq \mathbf{A}(t) \frac{Q^v(0)}{t} \mathbf{A}(t)^{-T} - \mathbf{A}(t) \frac{1}{t} \int_0^t \mathbf{A}(s)^{-1} \Sigma(s) \mathbf{A}(s)^{-1} ds \mathbf{A}(t)^{-T} \xrightarrow{t \rightarrow 0} +\infty,$$

where we have also used the fact  $\mathbf{A}(0) = I_d$  and the fact  $\int_0^t \mathbf{A}(s)^{-1} \Sigma(s) \mathbf{A}(s)^{-1} ds$  tends to  $\Sigma(0)$  as  $t$  tends to 0 thanks to the continuity of  $\Sigma, \mathbf{A}^{-1}$  on  $[0, T]$ .

Hence, for all  $t \in ]0, T]$ ,

$$\lim_{t \rightarrow 0} \frac{\sqrt{t}}{\|Q^v(t)\|} = \begin{cases} \sqrt{\frac{1}{\|\Sigma(0)\|}}, & \text{if } Q^v(0) = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{3.24}$$

This yields (3.23) which implies (3.22); consequently the solution  $X$  of (3.19) prolongates to  $t = T$  and item 2. is proved. □

Though, this will not be exploited in the algorithm proposed at Section 2.5, it is interesting to note that the process  $\xi$  introduced in (3.19) can also be seen as the solution of a McKean SDE. Proposition 2.3.7 below shows that (1.7) admits existence and uniqueness if and only if Assumption 12 is verified. In particular we have the following.

**Proposition 2.3.7.** 1. *There is at most one solution  $(\xi, m, Q)$  of (1.7).*

2. *Suppose the validity of Assumption 12. Let  $\xi$  be the unique solution of (3.19). Then  $(\xi, m^v, Q^v)$  is a solution of (1.7).*

*Proof.* 1. Let  $(\xi, m, Q)$  be a solution of (1.7). By definition,  $\xi$  solves an SDE of type (3.20) replacing  $a$  by  $a^\Sigma : s \mapsto -a(T-s) - \Sigma(T-s)Q(T-s)^{-1}$  and  $c$  by  $c^\Sigma : s \mapsto -c(T-s) + \Sigma(T-s)Q(T-s)^{-1}m(T-s)$ .

By Problem 6.1 Section 5 in [69], the function  $t \mapsto \mathbf{E}(\xi_t) (= m(T-t))$  (resp.  $t \mapsto \text{Cov}(\xi_t) (= Q(T-t))$ ) solves the first line of (3.1) (resp. (3.2)) replacing  $a$  by  $a^\Sigma$  and  $c$  by  $c^\Sigma$ . Then, the following identities hold for all  $t \in ]0, T]$ :

$$m(T-t) = \mathbf{E}(\xi_0) - \int_0^t a(T-s)m(T-s) + c(T-s) ds, \tag{3.25}$$



$$Q(T-t) = \text{Cov}(\xi_0) - \int_0^t Q(T-s)a(T-s)^T + a(T-s)Q(T-s)ds - \int_0^t \Sigma(T-s)ds, \quad (3.26)$$

remarking that

$$a^\Sigma(t)m(T-t) + c^\Sigma(t) = -a(T-t)m(T-t) - c(T-t),$$

$$Q(T-t)a^\Sigma(t)^T + a^\Sigma(T-t)Q(t) = -Q(T-t)a(T-t)^T - a(T-t)Q(T-t) - 2\Sigma(T-t).$$

Applying the change of variable  $t \mapsto T - t$  in identities (3.25) and (3.26), we show that  $m$  (resp.  $Q$ ) solves the backward ODE (3.1) (resp. (3.2)), which is well-posed. We recall that  $\xi_0$  is distributed according to  $\nu$ . Then,  $m = m^\nu$  and  $Q = Q^\nu$ , see the beginning of Section 2.3.1. As a consequence,  $\xi$  solves (3.19) and is uniquely determined thanks to Remark 2.3.5. This shows the validity of item 1.

2. Let  $\xi$  be the unique solution of (3.19). Then, the time-reversed process  $\xi$  solves (3.20) and  $\xi_T \sim \mathbf{N}(m^\nu(0), Q^\nu(0))$ , thanks to item 1. of Lemma 2.3.6. Now, by Remark 2.3.4, we have  $\mathbf{E} \xi_t = m^\nu(t)$ ,  $\text{Cov} \xi_t = Q^\nu(t)$  for all  $t \in [0, T[$ . This concludes the proof of item 2. □

**Remark 2.3.8.** 1. In [62] we have discussed existence and uniqueness of more general McKean problems involving the densities of the marginal laws instead of expectation and covariance matrix, where the solution is the time-reversal of some (not necessarily Gaussian) diffusion.

2. In particular, in Section 4.5 of [62] we have investigated existence and uniqueness of

$$\begin{aligned} & \square Y_t = Y_0 - \int_0^t \tilde{b}(T-r, Y_r) dr + \int_0^t \frac{\text{div}_y(\Sigma_i(T-r)p_r(Y_r))}{p_r(Y_r)} \mathbf{1}_{i \in \llbracket 1, d \rrbracket} dr + \int_0^t \sigma(T-r) d\beta_r, \\ & \square p_t \text{ density law of } \mathbf{p}_t = \text{law of } Y_t, t \in ]0, T[, \\ & \square Y_0 \sim \mathbf{p}_T = \nu, \end{aligned} \quad (3.27)$$

where  $\beta$  is a  $m$ -dimensional Brownian motion and  $\Sigma = \sigma\sigma^T$ , whose solution is the couple  $(Y, \mathbf{p})$ . Moreover, when the solution exists, there is a probability-valued function  $\mathbf{u}$  defined on  $[0, T]$  solution of the Fokker-Planck equation

$$\begin{aligned} & \square \partial_t \mathbf{u} = \frac{1}{2} \sum_{i,j=1}^d \partial_{ij}^2 (\sigma\sigma^T)_{ij}(t)\mathbf{u} - \text{div} \tilde{b}(t, x)\mathbf{u} \\ & \square \mathbf{u}(T) = \nu. \end{aligned} \quad (3.28)$$

3. Suppose that  $\nu$  is a Gaussian law on  $\mathbf{R}^d$ . It is possible to show that Assumption 12 is equivalent to the existence of a probability-valued solution  $\mathbf{u}$  of (3.28). In this case the McKean problems (1.7) and (3.27) are equivalent. In particular the component  $Y$  of the solution of (3.27) is Gaussian.

We continue with a preliminary lemma. Let  $W$  be a Brownian motion. For each  $(s, x) \in [0, T] \times \mathbb{R}^d$ ,  $X^{s,x}$  will denote below the process

$$X_t^{s,x} := x + \int_s^t \sigma(r) dW_r, \quad t \in [s, T].$$

**Lemma 2.3.9.** *Suppose the validity of Assumption 13. Let  $v : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  of class  $C^{0,1} [0, T], \mathbb{R}^d$ , with polynomial growth and such that the function  $H^v : (t, x) \mapsto H(t, x, v(t, x), \nabla_x v(t, x))$  is continuous with polynomial growth. Then, the following assertions are equivalent.*

1.  $v$  is a viscosity solution of (3.18).

2. For each  $(s, x) \in [0, T] \times \mathbb{R}^d$ ,

$$v(s, x) = \mathbb{E} \int_s^T H(r, X_r^{s,x}, v(r, X_r^{s,x}), \nabla_x v(r, X_r^{s,x})) dr + g(X_T^{s,x}). \quad (3.29)$$

3.  $v$  is of class  $C^{1,2} [0, T], \mathbb{R}^d$  and is a (classical) solution of (3.18).

*Proof.* Let  $v$  as in the lemma statement.

a) We set

$$w^v(s, x) := \mathbb{E} \left[ g(X_T^{s,x}) + \int_s^T H^v(r, X_r^{s,x}) dr \right], \quad (s, x) \in [0, T] \times \mathbb{R}^d. \quad (3.30)$$

We show first that  $w^v$  is a (classical) solution in  $C^{1,2} [0, T], \mathbb{R}^d \cap C^0 [0, T] \times \mathbb{R}^d$  with polynomial growth of the linear PDE

$$\begin{aligned} \partial_t w(t, x) + \frac{1}{2} \text{Tr}[\sigma \sigma^T(t) \nabla_x^2 w(t, x)] + H^v(t, x) &= 0, \quad (t, x) \in [0, T] \times \mathbb{R}^d \\ w(T, \cdot) &= g. \end{aligned} \quad (3.31)$$

Indeed  $w^v$  can be rewritten as

$$w(s, x) = \int_{\mathbb{R}^d} g(z) p_T(s, z - x) dz + \int_s^T \int_{\mathbb{R}^d} H^v(r, z) p_r(s, z - x) dz dr, \quad (s, x) \in [0, T] \times \mathbb{R}^d, \quad (3.32)$$

where for each  $r \in [0, T], s \in [0, r], p_r(s, \cdot)$  is the density of the r.v.  $\int_s^r \sigma(u) dW_u$ , i.e. a Gaussian r.v. with mean zero and covariance  $\int_s^r \Sigma(u) du$ . Moreover, it is well-known, see e.g. Remark 3.2 in [39], that for each  $r \in [0, T], p_r : [0, r] \times \mathbb{R}^d \rightarrow \mathbb{R}$  is a smooth solution of

$$\partial_t p_r(t, z) + \frac{1}{2} \text{Tr} \left[ \Sigma(t) \nabla_x^2 p_r(t, z) \right] = 0, \quad (t, z) \in [0, r] \times \mathbb{R}^d. \quad (3.33)$$

Consequently, by usual integration theorems allowing to commute derivation and integrals, one shows (3.31).

b) Consequently  $w^v$  is a viscosity solution (3.31).

- c) If 1. holds then  $v$  is also a viscosity solution of (3.31). By point 1. of Remark 2.3.12, equation (3.31) admits at most one continuous viscosity solution with polynomial growth. So  $v = w^v$  which means 2.
- d) If 2. holds then  $v = w^v$  and by b)  $v$  is a viscosity solution of (3.31) and therefore of (3.18).
- e) 3. implies obviously 1. Viceversa, if item 1. holds, a) implies that  $w^v$  is a classical solution of (3.31); b) and the uniqueness of viscosity solutions for previous linear equation implies  $v = w^v$  and finally item 3.

□

We state now the announced representation result.

**Theorem 2.3.10.** *Suppose the validity of Assumption 13. Let  $v$  be a Gaussian probability fulfilling Assumption 12 with associated functions  $m^v$  and  $Q^v$ .*

*Let  $v \in C^{0,1} [0, T], \mathbb{R}^d; \mathbb{R}$  with polynomial growth and such that  $H^v : (t, x) \mapsto H(t, x, v(t, x), \nabla_x v(t, x))$  is continuous with polynomial growth. Then,  $v$  is a viscosity solution of (3.18) if and only if for all  $t \in [0, T]$*

$$\begin{aligned}
 & \square \quad \xi_t = \xi_0 - \int_0^t (b(T-s, \xi_s) + b_c(T-s, \xi_s, m^v(T-s), Q^v(T-s))) ds + \int_0^t \sigma(T-s) d\beta_s, \\
 & \square \quad \xi_0 \sim v, \\
 & \square \quad v(t, \xi_t) = \mathbf{E} \int_t^T H(s, \xi_s, v(s, \xi_s), \nabla_x v(s, \xi_s) - b(s, \xi_s), \nabla_x v(s, \xi_s)) ds + g(\xi_T, \xi_t).
 \end{aligned} \tag{3.34}$$

**Remark 2.3.11.** *The affine drift  $b$  remains a degree of freedom of the representation. In Section 2.5, in the framework of the Hamilton-Jacobi-Bellman PDEs are given elements to choose rationally  $b$ .*

**Remark 2.3.12.** *We remark that previous representation (3.34) is valid even if uniqueness does not hold for the semilinear PDE (3.18). In that case even the equation (3.34) does not admit uniqueness. However, we provide below some typical situations for which (3.18) admits at most one viscosity solution, within different classes of solutions.*

1. *Suppose the validity of Assumption 13. Suppose also that  $H$  is continuous with polynomial growth in  $x$  and linear growth in  $(y, z)$ . In addition, we suppose that  $H$  is Lipschitz in  $(y, z)$  uniformly in  $(t, x)$  and suppose that for all  $R > 0$ , there exists  $m_R : \mathbb{R} \rightarrow \mathbb{R}^+$ , tending to 0 at  $0^+$  such that*

$$|H(t, x, y, z) - H(t, x, y, z)| \leq m_R (|x| - x (1 + |z|)),$$

*for all  $t \in [0, T]$ ,  $z \in \mathbb{R}^d$  and  $|x|, |y| \leq R$ . Then, by Theorem 5.1 in [93], implies that (3.18) admits at most one continuous viscosity solution with polynomial growth. In fact that theorem states uniqueness even in a wider class of solutions.*

2. The first theorem in [65] formulates a uniqueness result in a suitable class of bounded uniformly continuous solutions. Alternative assumptions are available to ensure uniqueness in different classes of unbounded functions, for fully non-linear parabolic Cauchy problems. See for instance Corollary 2 in [59], Theorem 3.1 in [88], [33], [58].

**Proof** (of Theorem 2.3.10).

Let  $v$  as in the statement.

1. Lemma 2.3.6 implies that there exists an  $\mathbf{F}^\xi$ -Brownian motion  $W$  such that, under  $\mathbf{P}$ ,

$$\xi_t = \xi_0 + \int_0^t b(s, \xi_s) ds + \int_0^t \sigma(s) dW_s, \quad t \in [0, T], \quad (3.35)$$

where  $\xi_0 \sim \mathbf{N}(m^v(0), Q^v(0))$ . In particular

$$\mathbf{E} \sup_{t \in [0, T]} \xi_s^p < \infty, \quad \forall p \geq 1. \quad (3.36)$$

This, together with Assumption 13 and the polynomial growth of  $H^v$  also imply that the r.v.

$$\int_0^T H(s, \xi_s, v(s, \xi_s), \nabla_x v(s, \xi_s)) - b(s, \xi_s), \nabla_x v(s, \xi_s) ds + g(\xi_T)$$

is square integrable.

2. We give now an equivalent formulation of (3.34) using a change of probability measure.

We set  $L_s := \sigma(s)^{-1} b(s, \xi_s)$ ,  $s \in [0, T]$ . We denote by  $\mathbf{Q}$ , the probability equivalent to  $\mathbf{P}$  on  $\mathbf{F}_T^\xi$  defined by  $\frac{d\mathbf{Q}}{d\mathbf{P}} = \mathbf{E} - \int_0^T L_s^i dW_s^i$ , being well-defined thanks to Lemma 2.7.1.

The goal is to show that  $v$  fulfills (3.34) if and only if it fulfills for all  $t \in [0, T]$

$$v(t, \xi_t) = \mathbf{E}_{\mathbf{Q}} \int_t^T H(s, \xi_s, v(s, \xi_s), \nabla_x v(s, \xi_s)) ds + g(\xi_T) - \xi_t. \quad (3.37)$$

We remark that,

$$\xi_t = \xi_0 + \int_0^t \sigma(s) d\mathbb{W}_s, \quad t \in [0, T], \quad (3.38)$$

where

$$\mathbb{W} := W + \int_0^\cdot L_s ds, \quad (3.39)$$

which is a Brownian motion under  $\mathbf{Q}$  thanks to Girsanov's Theorem 5.1 in [69]. By item 1.

$$\int_0^T H(s, \xi_s, v(s, \xi_s), \nabla_x v(s, \xi_s)) ds + g(\xi_T),$$

is obviously also square integrable under  $\mathbf{Q}$ .

We set  $H_s := H(s, \xi_s, v(s, \xi_s), \nabla_x v(s, \xi_s))$ ,  $s \in [0, T]$ , for the sake of brevity.

We remark first that for each given  $s \in [0, T]$ ,

$$b(s, \xi_s), \nabla_x v(s, \xi_s) = \sigma(s) \sigma(s)^{-1} b(s, \xi_s), \nabla_x v(s, \xi_s) = L_s, \sigma(s)^{-T} \nabla_x v(s, \xi_s). \quad (3.40)$$

Then, (3.40) combined with the Markov property of  $\xi$  implies that (3.34) is equivalent to

$$v(t, \xi_t) = \mathbf{E} \int_t^T H_s - L_s \sigma(s)^{-T} \nabla_x v(s, \xi_s) ds + g(\xi_T) \mathbf{F}_t^\xi,$$

which can be rewritten

$$v(t, \xi_t) = M_t - \int_0^t H_s - L_s \sigma(s)^{-T} \nabla_x v(s, \xi_s) ds,$$

where  $M$  is the P-martingale

$$M_t = \mathbf{E} \int_0^t H_s - L_s \sigma(s)^{-T} \nabla_x v(s, \xi_s) ds + g(\xi_T) \mathbf{F}_t^\xi, \quad t \in [0, T]. \quad (3.41)$$

Similarly, (3.37) is equivalent to

$$v(t, \xi_t) = \bar{M}_t - \int_0^t H_s ds,$$

where  $\bar{M}$  is the Q-martingale

$$\bar{M}_t = \mathbf{E}_Q \int_0^t H_s ds + g(\xi_T) \mathbf{F}_t^\xi, \quad t \in [0, T]. \quad (3.42)$$

To show the aforementioned equivalence, it suffices now to show

$$M_t - \bar{M}_t = \int_0^t L_s \sigma(s)^{-T} \nabla_x v(s, \xi_s) ds, \quad t \in [0, T].$$

On the one hand, Theorem 1.7 Chapter 8 in [96] implies that the process  $\mathbb{M} := M + \sum_{i=1}^d [M, \int_0^\cdot L_s^i dW_s^i]$  is a Q-local martingale. On the other hand, for each  $i \in \llbracket 1, d \rrbracket$  by Proposition 3.10 in [54] we have

$$\begin{aligned} [M, \int_0^\cdot L_s^i dW_s^i] &= [v(\cdot, \xi), \int_0^\cdot L_s^i dW_s^i] \\ &= \int_0^\cdot L_s^i \sigma(s)^{-T} \nabla_x v(s, \xi_s) dW_s^i, \end{aligned}$$

combining (3.42) with the usual properties of covariation for semimartingales. This means that

$$\mathbb{M} = M + \int_0^\cdot L_s \sigma(s)^{-T} \nabla_x v(s, \xi_s) ds$$

is a Q-local martingale. Now,

$$\begin{aligned} \mathbb{M}_T &= M_T + \int_0^T L_s \sigma(s)^{-T} \nabla_x v(s, \xi_s) ds \\ &= \int_0^T H_s ds + g(\xi_T), \end{aligned}$$

thanks to (3.41). Since  $\bar{M}$  and  $\mathbb{M}$  are Q-local martingales being equal at  $t = T$ , we have  $\bar{M} = \mathbb{M}$ . This shows the validity of point 2.

3. For each  $(s, x) \in [0, T] \times \mathbb{R}^d$ , we set  $X^{s,x} := x + \int_s^T \sigma(r) dW_r$  where  $W$  is the  $Q$ -Brownian motion defined in (3.39). Associated with  $v$ , we consider the continuous function

$$w^v(t, x) := \mathbb{E}_Q \int_t^T H(r, X_r^{t,x}, v(r, X_r^{t,x}), \nabla_x v(r, X_r^{t,x})) dr + g(X_T^{t,x}), \quad (t, x) \in [0, T] \times \mathbb{R}^d.$$

We observe that  $v$  fulfills (3.37) if and only if for all  $(t, x) \in [0, T] \times \mathbb{R}^d$

$$v(t, x) = w^v(t, x). \quad (3.43)$$

Indeed this follows by the freezing lemma of the conditional expectation, the fact that  $\xi_t$  is independent of the random field  $X_s^{t,x}$  for  $t \leq s \leq T, x \in \mathbb{R}^d$  and the flow property

$$X_s^{t,\xi_t} = \xi_s, \quad s \in [t, T].$$

4. It remains to show that (3.43) is satisfied if and only if  $v$  is a viscosity solution of (3.18). This is the object of Lemma 2.3.9 applied under the probability  $Q$ , in particular to the equivalence between item 1. and item 3.

□

## 2.4 Representation of stochastic control problems

Let us briefly recall the link between stochastic control and non-linear PDEs given by the Hamilton-Jacobi-Bellman (HJB) equation. We refer for instance to [42, 94, 108] for more details.

Let  $A \subset \mathbb{R}^k$  compact and denote by  $\mathbf{A}_0$  the set of all  $A$ -valued progressively measurable processes  $(\alpha_t)_{t \in [0, T]}$ , namely the set of *admissible controls*.

We consider now *state processes*  $(X_t^{s,x,\alpha})_{s \leq t \leq T, \alpha \in \mathbf{A}}$  starting at time  $s \in [0, T]$  with value  $x \in \mathbb{R}^d$ , solutions of the controlled SDE

$$dX_t = b(t, X_t, \alpha_t) dt + \sigma(t) dW_t, \quad (4.1)$$

where  $W$  is a  $d$ -dimensional Brownian motion and  $b : [0, T] \times \mathbb{R}^d \times A \rightarrow \mathbb{R}^d$  is supposed to fulfill the following.

**Assumption 14.** *The function  $b$  is continuous and there exists  $K \geq 0$  such that*

$$|b(t, x_2, a) - b(t, x_1, a)| \leq K |x_2 - x_1|, \quad (t, x_1, x_2, a) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times A.$$

Note that Assumption 14 implies  $b$  to have linear growth in space uniformly in time and in the control. Consequently, (4.1) starting at time  $s$  with value  $x$  admits a unique solution for each  $\alpha \in \mathbf{A}_0$ , for each  $(s, x) \in [0, T] \times \mathbb{R}^d$ , by the same arguments as in Theorem 3.1 in [108].

We also introduce the *cost function*  $J : [0, T] \times \mathbb{R}^d \times \mathbf{A}_0 \rightarrow \mathbb{R}$  defined by

$$J(s, x, \alpha) := \mathbf{E} \left[ g(X_T^{s,x,\alpha}) + \int_s^T f(r, X_r^{s,x,\alpha}, \alpha_r) dr \right], \quad (s, x, \alpha) \in [0, T] \times \mathbb{R}^d \times \mathbf{A}_0, \quad (4.2)$$

where the function  $f : [0, T] \times \mathbb{R}^d \times A \rightarrow \mathbb{R}$  (*running cost*) is supposed to fulfill what follows.

**Assumption 15.** *The function  $f$  is continuous and there exists  $m, M \geq 0$  such that*

$$|f(t, x, a)| \leq M(1 + |x|^m), \quad (t, x, a) \in [0, T] \times \mathbb{R}^d \times A.$$

Supposing the validity of Assumptions 14 and 15 together with Assumption 13 on the function  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  (*terminal cost*), we are interested in minimizing, over control processes  $\alpha \in \mathbf{A}_0$  the functions  $\alpha \mapsto J(0, x, \alpha)$  for every  $x \in \mathbb{R}^d$ .

To tackle this finite horizon stochastic control problem, the usual approach consists in introducing the associated *value* (or *Bellman*) function  $v : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  representing the minimum expected costs, starting from any time  $t \in [0, T]$  at any state  $x \in \mathbb{R}^d$ , i.e.

$$v(t, x) := \inf_{\alpha \in \mathbf{A}_0} J(t, x, \alpha), \quad (t, x) \in [0, T] \times \mathbb{R}^d. \quad (4.3)$$

Note that the terminal condition is known, which fixes  $v(T, \cdot) = g$ , whereas  $v(0, \cdot)$  corresponds to the solution of the original minimization problem.

**Remark 2.4.1.** *Suppose the validity of Assumptions 13, 14 and 15.*

1. *The function  $v$  is continuous on  $[0, T] \times \mathbb{R}^d$  and has polynomial growth, see Theorem 5. Chapter 3. in [71].*
2. *The value function  $v$  is a viscosity solution of the Hamilton-Jacobi-Bellman equation*

$$\begin{aligned} \partial_t v(t, x) + H(t, x, \nabla_x v(t, x)) + \frac{1}{2} \text{Tr}[\sigma \sigma^\top(t) \nabla_x^2 v(t, x)] &= 0, \quad (t, x) \in [0, T] \times \mathbb{R}^d \\ v(T, \cdot) &= g, \end{aligned} \quad (4.4)$$

where  $H$  denotes the real-valued function defined on  $[0, T] \times \mathbb{R}^d \times \mathbb{R}^d$  by

$$H(t, x, \delta) := \inf_{a \in A} \{ f(t, x, a) + b(t, x, a, \delta) \}, \quad (t, x, \delta) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d, \quad (4.5)$$

see for example Theorem 7.4 in [108].

3. *By definition, it is obvious that  $(x, z) \mapsto H(t, x, z)$  has polynomial growth uniformly with respect to  $t$ . It is also clear that  $H$  is continuous.*
4. *Under Assumptions 13, 14 and 15, the PDE (4.4) admits at most one viscosity solution in the class of continuous solutions with polynomial growth, see Theorem II.3 in [80]. Since  $v$  has polynomial growth, the value function  $v$  is the unique viscosity solution of (4.4) in the considered class.*

We formulate below another assumption for the value function  $v$ .

**Assumption 16.**  $v$  is of class  $C^{0,1}$   $[0, T], \mathbb{R}^d$  such that  $\nabla_x v$  has polynomial growth.

**Remark 2.4.2.** Under Assumption 16, using Remark 2.4.1 3. that the function  $(t, x) \mapsto H^v(t, x) := H(t, x, \nabla_x v(t, x))$  is continuous with polynomial growth.

**Remark 2.4.3.** 1. Assumption 16 is not so restrictive, since whenever  $g$  and  $f$  are locally Lipschitz with polynomial growth gradient (in space), then  $v$  is locally Lipschitz in the space variable. To prove this, it suffices to show that  $J$  is locally Lipschitz in  $x$  uniformly in  $t$  and  $\alpha$ . A proof of this fact is given in Lemma 2.7.2 stated in the Appendix.

In that context, the value function  $v$  is in particular absolutely continuous and for every  $t \in [0, T]$ , for almost every  $x \in \mathbb{R}^d$ ,  $v(t, \cdot)$  is differentiable and  $\nabla_x v(t, \cdot)$  exists.

2. Suppose in addition that the functions  $f$ ,  $g$  and  $b$  are of class  $C^1$  (in the space variable) and the validity of Assumption 17. Then  $\nabla_x v(t, \cdot)$  has polynomial growth as we show below. Indeed, by usual dominated convergence arguments, we can show that for each  $(t, \alpha) \in [0, T] \times \mathbf{A}_0$ ,  $x \mapsto J(t, x, \alpha)$  is differentiable with gradient

$$\nabla_x J(t, x, \alpha) = \mathbf{E} \left[ Y_T^{t,x,\alpha} \nabla_x g(X_T^{t,x,\alpha}) + \int_t^T Y_r^{t,x,\alpha} \nabla_x f(r, X_r^{t,x,\alpha}, \alpha_r) dr \right], \quad (4.6)$$

where  $Y^{t,x,\alpha}$  is the unique matrix-valued process fulfilling

$$Y_r^{t,x,\alpha} = I_d + \int_t^r \nabla_x b(s, X_s^{t,x,\alpha}, \alpha_s) Y_s^{t,x,\alpha} ds, \quad r \in [t, T],$$

where  $\nabla_x b := \partial_x b^i$   $(i,j) \in [[1,d]]^2$ .

Combining what precedes with Lemma 2.7.3 stated in the Appendix, we deduce that for all  $t \in [0, T]$ , for almost every  $x \in \mathbb{R}^d$

$$\nabla_x v(t, x) = \nabla_x J(t, x, \alpha^*(t, x)), \quad (4.7)$$

where  $\alpha^*$  is the Borel function introduced in Assumption 17. In view of (4.6) and (4.7),  $\nabla_x v$  has polynomial growth.

**Corollary 2.4.4.** Let  $\nu$  be a Gaussian probability measure fulfilling Assumption 12 with associated functions  $m^\nu$  and  $Q^\nu$ . We suppose the validity of Assumptions 13, 14, 15. Among the functions  $v : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  fulfilling Assumption 16, the value function is the unique one which is solution of (3.34). (In this framework  $H$  only depends on  $\nabla_x v$  and not on  $v$ ).

*Proof.* We recall that  $H^v$  has polynomial growth by Remark 2.4.2 1. Otherwise, on the one hand, by Remark 2.4.1 and the direct implication in Theorem 2.3.10,  $v$  fulfills (3.34). On the other hand, if a function  $v$  fulfills (3.34) then, by the converse implication of Theorem 2.3.10  $v$  is a viscosity solution of (4.4). By Remark 2.4.1 3.,  $v$  can only be the value function. □



We introduce a supplementary hypothesis on the value function  $v$ .

**Assumption 17.** *There exists a Borel function  $\alpha^* : [0, T] \times \mathbb{R}^d \rightarrow A$  such that*

$$H(t, x, \nabla_x v(t, x)) = b(t, x, \alpha^*(t, x)), \nabla_x v(t, x) + f(t, x, \alpha^*(t, x)), (t, x) \in [0, T] \times \mathbb{R}^d.$$

We state (and show below) a *verification* type result involving  $\alpha^*$  without any further regularity assumptions on the value function. That result is somehow classical, but it is not obvious to find it in the literature (see e.g. Chapter 5 of [108] or [53]), with our assumptions. So, for the consistency of the paper we provide a proof. Note to begin that the Borel function  $b^* : (t, x) \rightarrow b(t, x, \alpha^*(t, x))$  has linear growth thanks to Assumption 14. As a consequence, the *closed loop* equation

$$d\bar{X}_t = b^*(t, \bar{X}_t) dt + \sigma(t) dW_t, \tag{4.8}$$

admits a unique strong solution  $\bar{X}^x$  starting at time 0 with value  $x$ , for each  $x \in \mathbb{R}^d$ , see Theorem 6 in [110].

**Proposition 2.4.5.** *Suppose the validity of Assumptions 13, 14, 15. Let  $v$  be the value function defined in (4.3) supposed to be of class  $C^{0,1}$  such that  $(t, x) \mapsto H(t, x, \nabla_x v(t, x))$  has polynomial growth.*

*Then, the Borel function  $\alpha^*$  introduced in Assumption 17 defines an optimal feedback function for the considered control problem in the sense that for each  $x \in \mathbb{R}^d$ ,*

$$v(0, x) = J(0, x, \alpha^*, \bar{X}^x). \tag{4.9}$$

**Proof** (of Proposition 2.4.5). Let  $x \in \mathbb{R}^d$ .

1.  $v$  is a continuous viscosity solution with polynomial growth of (4.4) and so of (3.18), with  $(t, x, y, z) \mapsto H(t, x, z)$  for the non linearity. By Remark 2.4.1 we know that  $H^v$  is continuous and by assumption it has polynomial growth. So we apply Lemma 2.3.9 to deduce that  $v$  is of class  $C^{1,2} [0, T[, \mathbb{R}^d$  and is a classical solution of (4.4).
2. Applying Itô's formula to  $v(\cdot, \bar{X}^x)$  between 0 and  $T_0 \in [0, T[$  and using the fact  $v$  is a classical solution of (4.4) combined with Assumption 17, we obtain

$$v(0, x) = v(T_0, \bar{X}_{T_0}^x) + \int_0^{T_0} f(r, \bar{X}_r^x, \alpha^*(r, \bar{X}_r^x)) dr - M_{T_0}, \tag{4.10}$$

where

$$M_t = \int_0^t \nabla_x v(r, \bar{X}_r^x)^T \sigma(r) dW_r, t \in [0, T[.$$

By the usual BDG (Burkholder-Davies-Gundy) and Jensen's arguments,  $\sup_{t \in [0, T]} |\bar{X}_t^x|$  has all its moments. So, (4.10) implies that the local martingale  $M$  extends continuously to a true martingale on  $[0, T]$  still denoted by  $M$  verifying  $\sup_{t \in [0, T]} |M_t| \in L^1$ . Indeed  $v$  is continuous on  $[0, T] \times \mathbb{R}^d$  and  $v$  (resp.  $f$ ) has polynomial growth in space (resp. in the second and third

variable). Therefore  $M$  is a true martingale. Sending  $T_0$  to  $T$ , (4.10) holds with  $T_0$  replaced by  $T$  and  $v|_{T_0}, \bar{X}_{T_0}^x$  replaced by  $g|_T, \bar{X}_T^x$ . Taking the expectation, we obtain

$$v(0, x) = \mathbf{E} \left[ g|_T + \int_0^T f(r, \bar{X}_r^x, \alpha^*(r, \bar{X}_r^x)) dr \right]. \quad (4.11)$$

3. The process  $\alpha_t^* := \alpha^*(t, \bar{X}_t^x)$ ,  $t \in [0, T]$ , belongs to the set  $\mathbf{A}_0$  of admissible controls and  $X = \bar{X}^x$ , is a solution of (4.1). Invoking pathwise uniqueness for (4.1), we obtain  $X^{0,x,\alpha^*}$  coincides with  $\bar{X}^x$ . Then, (4.11) implies (4.9). □

We formulate now a corollary in which is given a representation formula for the value function  $v$  involving the optimal feedback function  $\alpha^*$ .

**Corollary 2.4.6.** *Let  $\nu$  be a Gaussian probability measure fulfilling Assumption 12 with associated functions  $m^\nu$  and  $Q^\nu$ . We suppose the validity of Assumptions 13, 14, 15. Among the functions fulfilling Assumptions 16 and 17, the value function  $v$  is the unique one which is solution of*

$$\begin{aligned} & \square \quad \xi_t = \xi_0 - \int_0^t (b(T-s, \xi_s) + b_c(T-s, \xi_s, m^\nu(T-s), Q^\nu(T-s))) ds + \int_0^t \sigma(T-s) d\beta_s, \\ & \square \quad \xi_0 \sim \nu, \\ & \square \quad v(t, \xi_t) = \mathbf{E} \int_t^T f(s, \xi_s, \alpha^*(s, \xi_s)) - b(s, \xi_s) - b^*(s, \xi_s) \cdot \nabla_x v(s, \xi_s) ds + g(\xi_T) \quad (4.12) \end{aligned}$$

for all  $t \in [0, T]$ .

*Proof.* The result is a direct consequence of Corollary 2.4.4, replacing the function  $H$  by its expression given in Assumption 17. □

## 2.5 A heuristic algorithm

In this section, we propose a heuristic algorithm to solve the control problem described in Section 2.4. In what follows, the *terminal cost* function  $g$  is supposed to belong to  $C^1(\mathbb{R}^d)$ .

Consider a regular time grid with time step  $\delta t := \frac{T}{n}$  and grid instants  $t_k = k\delta t$  for any  $k \in \llbracket 0, n \rrbracket$ . For  $k = n-1, n-2, \dots, 0$ , select arbitrarily  $\bar{m}_{k+1}, c_{k+1} \in \mathbb{R}^d$  and  $\bar{Q}_{k+1} \in S_d^+(\mathbb{R}), a_{k+1} \in M_d(\mathbb{R})$  such that

$Q_k(t_k) := e^{-a_{k+1}\delta t} \bar{Q}_{k+1} e^{-a_{k+1}^\top \delta t} - \int_{t_k}^{t_{k+1}} e^{-a_{k+1}(s-t_k)} \Sigma(s) e^{-a_{k+1}^\top (s-t_k)} ds \in S_d^+(\mathbb{R})$ . By Corollary 2.4.6, applied substituting  $[0, T]$  with  $[t_k, t_{k+1}]$ , the solution of (4.4) on  $[t_k, t_{k+1}]$ , with terminal condition

$v(t_{k+1}, \cdot)$ , can be represented for  $t \in [t_k, t_{k+1}]$  by

$$\begin{aligned}
 \square \bar{\xi}_{k+1} &\sim \mathbf{N}(\bar{m}_{k+1}, \bar{Q}_{k+1}) \\
 \square Y_{k+1} &= v(t_{k+1}, \bar{\xi}_{k+1}) \\
 \square m_k(t) &= e^{-a_{k+1}(t_{k+1}-t)} \bar{m}_{k+1} - c_{k+1} \int_t^{t_{k+1}} e^{-a_{k+1}(s-t)} ds \\
 \square Q_k(t) &= e^{-a_{k+1}(t_{k+1}-t)} \bar{Q}_{k+1} e^{-a_{k+1} \int_t^{t_{k+1}} \Sigma(s) ds} - \int_t^{t_{k+1}} e^{-a_{k+1}(s-t)} \Sigma(s) e^{-a_{k+1} \int_t^s \Sigma(s') ds'} ds \\
 \square \xi_{k,T-t} &= \bar{\xi}_{k+1} - \int_{T-t}^{t_{n-(k+1)}} a_{k+1} \xi_{k,s} + c_{k+1} + b_c(T-s, \xi_{k,s}, m_k(T-s), Q_k(T-s)) ds \\
 &\quad + \int_{T-t}^{t_{n-(k+1)}} \sigma(T-s) d\beta_s \\
 \square \hat{\xi}_{k,t} &= \xi_{k,T-t} \Big|_{t_{k+1}}^{t_{n-(k+1)}} \\
 \square v(t, \hat{\xi}_{k,t}) &= \mathbf{E} \int_t^{t_{k+1}} F_k(s, \hat{\xi}_{k,s}, \nabla_x v(s, \hat{\xi}_{k,s})) ds + Y_{k+1} \hat{\xi}_{k,t} .
 \end{aligned} \tag{5.1}$$

In the above recursion,  $\beta$  denotes a  $d$ -dimensional Brownian motion on  $[0, T]$ ; for any  $k \in \llbracket 0, n-1 \rrbracket$ ,  $(\xi_{k,t})_t$  is a  $d$ -dimensional process defined on  $[t_{n-(k+1)}, t_{n-k}]$  while  $(\hat{\xi}_{k,t})_t$  denotes the associated time reversal defined on  $[t_k, t_{k+1}]$ ; the driver  $F_k$  defined on  $[t_k, t_{k+1}] \times \mathbf{R}^d \times \mathbf{R}^d$  is such that,

$$F_k(t, x, \delta) := H(t, x, \delta) - a_{k+1}x + c_{k+1}, \delta = \min_{a \in A} \{f(t, x, a) + b(t, x, a), \delta\} - a_{k+1}x + c_{k+1}, \delta . \tag{5.2}$$

The idea now is to apply a classical numerical method based on linear regressions to approximate the solution to (5.1) recursively in time from  $k = n-1$  to  $k = 0$ . For each time instant  $k$ , select arbitrarily  $\bar{m}_{k+1}, c_{k+1} \in \mathbf{R}^d$  and  $\bar{Q}_{k+1} \in S_d^+(\mathbf{R}), a_{k+1} \in M_d(\mathbf{R})$  such that

$$Q_k = e^{-a_{k+1} \delta t} \bar{Q}_{k+1} e^{-a_{k+1} \int_t^{t_{k+1}} \Sigma(s) ds} \in S_d^+(\mathbf{R}) . \tag{5.3}$$

Then we propose to approximate  $v(t_k, \cdot)$  by  $v_k$  obtained by an explicit time discretization scheme of (5.1) with time step  $\delta t = \frac{T}{n}$  as follows.

$$\begin{aligned}
 \square \bar{\xi}_{k+1} &\sim \mathbf{N}(\bar{m}_{k+1}, \bar{Q}_{k+1}) \\
 \square Y_{k+1} &= v_{k+1}(\bar{\xi}_{k+1}) \\
 \square \xi_k &= \bar{\xi}_{k+1} - a_{k+1} \bar{\xi}_{k+1} + c_{k+1} + b_c(t_{k+1}, \bar{\xi}_{k+1}, \bar{m}_{k+1}, \bar{Q}_{k+1}) \delta t + \sigma(t_{k+1}) \sqrt{\delta t} \varepsilon_k \\
 \square v_k(\xi_k) &= \mathbf{E} \int_{t_{k+1}}^{t_{k+1}} F_k(t_{k+1}, \bar{\xi}_{k+1}, \nabla_x v_{k+1}(\bar{\xi}_{k+1})) \delta t + Y_{k+1} \xi_k ,
 \end{aligned} \tag{5.4}$$

where  $(\varepsilon_k)_{0 \leq k \leq n-1}$  are i.i.d.  $d$ -dimensional standard Gaussian variables. As in the classical literature, see e.g. [50], we propose to approximate the conditional expectation appearing in (5.4) using Monte-Carlo least squares regression based on a grid constituted by  $N$  independent simulations  $(\xi_k^i, \bar{\xi}_{k+1}^i)_{1 \leq i \leq N}$  for  $k \in \llbracket 0, n-1 \rrbracket$ . In that literature, one generally simulates forwardly that grid. The interest of such *fully backward* representations (5.1)-(5.4), where the grid  $(\xi_k^i, \bar{\xi}_{k+1}^i)_{1 \leq i \leq N}$  is defined backwardly in time, (like the value function), is twofold.

- In terms of computer memory: at each time instant  $k+1$ , the values of the grid are generated on the fly,  $(\xi_k^i, \bar{\xi}_{k+1}^i)_{1 \leq i \leq N}$ . Contrary to the standard approach, there is no need to store the whole grid over the whole set of grid instants  $k \in \llbracket 0, n-1 \rrbracket$ .

- In terms of the relevance of the grid: at each grid instant,  $k + 1$  the information acquired on the value function  $v(t_{k+1}, \cdot)$  and optimal control strategy  $\alpha^*(t_{k+1}, \cdot)$  can be used to adaptively optimize the grid parameter  $(a_{k+1}, c_{k+1}, \bar{m}_{k+1}, \bar{Q}_{k+1})$  in order to explore relevant regions of the state space.

We develop some arguments to justify the relevance mentioned above. Indeed, as already announced, the target idea is to generate the grid used for regression computations according to the optimally controlled process dynamics. If this were possible, the sensitivity of the driver  $F_k$  w.r.t. the third variable  $\nabla_x v$  would vanish. In fact the driver sensitivity w.r.t.  $\nabla_x v$  is known to be one major cause of the propagation of numerical errors in approximation schemes, see e.g. [51]. Replacing  $\nabla_x v_{k+1}(\bar{\xi}_{k+1})$  by a perturbation  $\nabla_x v_{k+1}(\bar{\xi}_{k+1}) + h$  in the last equation of (5.4) we obtain

$$v_k^h(\xi_k) := \mathbf{E} F_k(t_{k+1}, \bar{\xi}_{k+1}, \nabla_x v_{k+1}(\bar{\xi}_{k+1}) + h, \delta t + Y_{k+1} | \xi_k).$$

The impact on  $v_k(\xi_k)$  can crudely be evaluated by computing the error  $\mathbf{E}[|v_k^h(\xi_k) - v_k(\xi_k)|^2]$ . Supposing that no perturbation is impacting  $Y_{k+1}$ , fact which will be heuristically justified in Remark 2.5.2 1., we have

$$\mathbf{E}(|v_k^h(\xi_k) - v_k(\xi_k)|^2) \leq \mathbf{E} F_k(t_{k+1}, \bar{\xi}_{k+1}, \nabla_x v_{k+1}(\bar{\xi}_{k+1}) - F_k(t_{k+1}, \bar{\xi}_{k+1}, \nabla_x v_{k+1}(\bar{\xi}_{k+1}) + h)^2.$$

Suppose from now on the existence of a Borel function  $(t, x, \delta) \mapsto a^*(t, x, \delta)$ , such that

$$H(t, x, \delta) := \{f(t, x, a^*(t, x, \delta)) + b(t, x, a^*(t, x, \delta)), \delta\}, (t, x, \delta) \in [0, T] \times \mathbf{R}^d \times \mathbf{R}^d. \quad (5.5)$$

In this case one has  $a^*(t, x) = a^*(t, x, \nabla_x v(t, x))$ ,  $(t, x) \in [0, T] \times \mathbf{R}^d \times \mathbf{R}^d$ , where  $a^*$  was defined in Assumption 17. Coming back to (5.2) we get

$$F_k(t, x, \delta) := H(t, x, \delta) - a_{k+1}x + c_{k+1}, \delta = \{f(t, x, a^*(t, x, \delta)) + b(t, x, a^*(t, x, \delta)), \delta\} - a_{k+1}x + c_{k+1}, \delta. \quad (5.6)$$

A suitable application of the envelope theorem gives

$$\frac{\partial F_k}{\partial \delta}(t, x, \delta) = b(t, x, a^*(t, x, \delta)) - (a_{k+1}x + c_{k+1}), \quad (5.7)$$

which yields

$$\begin{aligned} \mathbf{E} |v_k^h(\xi_k) - v_k(\xi_k)|^2 &\leq \mathbf{E} \int_0^1 \frac{\partial F_k}{\partial \delta}(t_{k+1}, \bar{\xi}_{k+1}, \nabla_x v_{k+1}(\bar{\xi}_{k+1}) + \theta h) d\theta, h^2 \\ &= \mathbf{E} \int_0^1 b(t_{k+1}, \bar{\xi}_{k+1}, a^*(t_{k+1}, \bar{\xi}_{k+1}, \nabla_x v_{k+1}(\bar{\xi}_{k+1}) + \theta h)) d\theta - (a_{k+1}\bar{\xi}_{k+1} + c_{k+1}), h^2 \\ &\leq |h|^2 \mathbf{E} \int_0^1 b(t_{k+1}, \bar{\xi}_{k+1}, a^*(t_{k+1}, \bar{\xi}_{k+1}, \nabla_x v_{k+1}(\bar{\xi}_{k+1}) + \theta h)) d\theta - (a_{k+1}\bar{\xi}_{k+1} + c_{k+1})^2. \end{aligned}$$

The above relation highlights the fact that the original idea consisting in generating the grid according to a dynamics approaching the optimally controlled process dynamics reduces the propagation of the error induced by the Monte-Carlo regression scheme in terms of least square criteria.

**Remark 2.5.1.** *The above relation also shows that previous idea can be read in the more general perspective of the probabilistic representation of a solution  $v$  to a semilinear PDE of the type (1.1), via an FBSDE. In that general context, one expects the selected drift of the forward process in the FBSDE to reduce the impact of the sensitivity of the FBSDE driver with respect to  $\nabla_x v$ .*

Based on that observation, we propose a heuristic algorithm where parameters  $(a_{k+1}, c_{k+1})$  are adaptively chosen as

$$(a_{k+1}, c_{k+1}) \in \arg \min_{a,c} \mathbf{E} [ b(t_{k+1}, \bar{\xi}_{k+1}, a^*(t_{k+1}, \bar{\xi}_{k+1}, \nabla_x v_{k+1}(\bar{\xi}_{k+1})) - (a\bar{\xi}_{k+1} + c))^2 ]. \quad (5.8)$$

In the above algorithm, the random variables  $(\varepsilon^i, k \in [[0, n-1]], i \in [[1, N]])$  are i.i.d. according to

---

**Algorithm 1** Fully Backward Monte-Carlo Regression scheme

**Initialization** Set  $v_n = g$ ;  $k = n - 1$ ; select arbitrarily  $(\bar{m}_n, \bar{Q}_n) \in \mathbf{R}^d \times \mathbf{S}_d^+(\mathbf{R})$ ; generate  $(\xi_n^i)_{1 \leq i \leq N}$  i.i.d.  $\sim \mathbf{N}(\bar{m}_n, \bar{Q}_n)$ ; set  $Y_n^i = g(\xi_n^i)$ , for all  $i \in [[1, N]]$ .

**while**  $k \geq 0$  **do**

1.  $\alpha_{k+1}^i = \arg \min_{a \in \mathbf{A}} f(t_{k+1}, \xi_{k+1}^i, a) + b(t_k, \xi_{k+1}^i, a, \nabla_x v_{k+1}(\xi_{k+1}^i))$ , for all  $i \in [[1, N]]$ .
2.  $(a_{k+1}, c_{k+1}) = \arg \min_{(a,c) \in \mathcal{M}_d(\mathbf{R}) \times \mathbf{R}^d} \frac{1}{N} \sum_{i=1}^N a \xi_{k+1}^i + c - b(t_{k+1}, \xi_{k+1}^i, \alpha_{k+1}^i)^2$ .
3.  $\bar{m}_k = e^{-ak+1\delta t} \bar{m}_{k+1} - c_{k+1} \delta t$ .
4.  $Q_k = e^{-ak+1\delta t} \bar{Q}_{k+1} e^{-ak+1\delta t} - \Sigma(t_{k+1}) \delta t$ .
  - **If**  $Q_k \in \mathbf{S}_d^+(\mathbf{R})$ : set  $\bar{Q}_k = Q_k$ ,
  - **Else** : set  $\bar{Q}_k = Proj_{\mathbf{S}_d^+(\mathbf{R})}(Q_k)$ ; recompute  $\bar{Q}_{k+1} = e^{ak+1\delta t} \bar{Q}_k + \Sigma(t_{k+1}) \delta t e^{ak+1\delta t}$ ; regenerate  $(\xi_{k+1}^i)_{1 \leq i \leq N}$  i.i.d.  $\sim \mathbf{N}(\bar{m}_{k+1}, \bar{Q}_{k+1})$ ; set  $Y_{k+1}^i = v_{k+1}(\xi_{k+1}^i)$ , for all  $i \in [[1, N]]$ .
5. Set  $e_{k+1}^i = a_{k+1} \xi_{k+1}^i + c_{k+1} - b(t_{k+1}, \xi_{k+1}^i, \alpha_{k+1}^i)$ , for all  $i \in [[1, N]]$ .
6.  $\xi_k^i = \xi_{k+1}^i - a_{k+1} \xi_{k+1}^i + c_{k+1} + b(t_{k+1}, \xi_{k+1}^i, \bar{m}_{k+1}, \bar{Q}_{k+1}) \delta t + \sigma(t_{k+1}) \varepsilon_k^i \sqrt{\delta t}$ , for all  $i \in [[1, N]]$ .
7.  $v_k = \arg \min_{P \in P_p(\mathbf{R}^d)} \frac{1}{N} \sum_{i=1}^N Y_{k+1}^i + f(t_{k+1}, \xi_{k+1}^i, \alpha_{k+1}^i) - e_{k+1}^i \nabla_x v_{k+1}(\xi_{k+1}^i) \delta t - P(\xi_k^i)^2$ .
8.  $Y_k^i = Y_{k+1}^i + f(t_{k+1}, \xi_{k+1}^i, \alpha_{k+1}^i) - e_{k+1}^i \nabla_x v_{k+1}(\xi_{k+1}^i) \delta t$ , for all  $i \in [[1, N]]$ .
9.  $k - 1 \leftarrow k$ .

**end while**

---

$\mathbf{N}(0, I_d)$ ;  $Proj_{\mathbf{S}_d^+(\mathbf{R})} : \mathbf{S}_d(\mathbf{R}) \rightarrow \mathbf{S}_d^+(\mathbf{R})$  denotes the Frobenius projection operator on the closed and convex space of semidefinite matrices; for each  $p \in \mathbf{N}$ ,  $P_p(\mathbf{R}^d)$  denotes the set of polynomial functions on  $\mathbf{R}^d$  with degree  $p$ .

- Remark 2.5.2.** 1. Note that in Step 4, as soon as  $Q_k \in S_d^+(\mathbf{R})$  then  $(Y_{k+1}^i)_{1 \leq i \leq N}$  results from the update made at previous iteration at Step 8. That updating rule corresponds to the multi-step forward dynamic programming approach [51] which is well-known for not inducing any additional bias error that would propagate backwardly during iterations. However, when  $Q_k \notin S_d^+(\mathbf{R})$ , in Step 4, then we have to modify  $\bar{Q}_{k+1}$ , re-generate new variables  $(\xi_{k+1}^i)_{1 \leq i \leq N}$  i.i.d.  $\sim \mathbf{N}(\bar{m}_{k+1}, \bar{Q}_{k+1})$  and use the update  $Y_{k+1}^i = v_{k+1}(\xi_{k+1}^i)$  which adds a bias error. Fortunately, in our numerical simulations it appeared easy to chose a first covariance matrix  $\bar{Q}_n$  so that for all  $k \in \llbracket 0, n-1 \rrbracket$  we had  $Q_k \in S_d^+$ . In that situation, the error propagation is only due to the sensitivity of the driver w.r.t.  $\nabla_x v$  which is precisely minimized by our heuristics.
2. The complexity of Algorithm 1, is comparable to the traditional Monte-Carlo Regression scheme using a **forward grid**. Indeed, Algorithm 1 requires an additional linear regression calculation of order  $\mathcal{O}(d^2N)$  at Step 2 which is negligible w.r.t. the polynomial regression computations at Step 7 (operated by both algorithms) inducing  $\mathcal{O}(d^3N)$  operations in the specific case considered in simulations where the maximum degree of polynomials is  $p = 2$ . When  $Q_k \notin S_d^+$ , Algorithm 1 requires in addition, at Step 4, to implement: a Frobenius projection  $\text{Proj}_{S_d^+(\mathbf{R})}(Q_k)$  ( $\mathcal{O}(d^3)$ ),  $N$  multiplications of matrices  $d \times d$  with vectors  $d \times 1$  ( $\mathcal{O}(d^2N)$ ); and  $N$  independent generations of  $d$ -dimensional Gaussian random variables. These additional operations induce a complexity of  $\mathcal{O}(d^3N)$  which does not increase the original  $\mathcal{O}(d^3N)$  complexity.
3. In terms of memory, as already mentioned, we do not have to store the whole regression grid on the whole time horizon constituted of  $ndN$  reals but only to consider  $dN$  reals at each instant.

**Remark 2.5.3.** Suppose that at each time step  $k \in \llbracket 0, n-1 \rrbracket$  the matrix  $Q_k$  belongs to  $S_d^+(\mathbf{R})$ . Then, Algorithm 1 is based on the representation formula appearing in Corollary 2.4.6, on the whole time interval  $[0, T]$  with piecewise constant coefficients  $a, c$  such that  $a(t), c(t) = a_{k+1}, c_{k+1}$  for each  $t \in ]t_k, t_{k+1}]$ , for each  $k \in \llbracket 0, n-1 \rrbracket$ .

## 2.6 Stochastic control of thermostatically controlled loads

### 2.6.1 Model description

With the massive integration of variable renewable energies (like wind farms or solar panels) into power systems, balancing supply and demand in a real time basis requires to develop new leverages. A technical solution is to develop load control schemes in order to automatically adapt consumption to generation. In this section, we propose to apply Algorithm 1 in order to control a large heterogeneous population of air-conditioners on a time horizon  $[0, T]$  such that the overall consumption of the population follows a given target profile, while preserving the rooms temperatures within users comfort bounds.

We consider a hierarchical control scheme introduced in [30], where the population is aggregated into  $d$  clusters of  $N^i$  homogeneous loads (with same air-conditioners and rooms characteristics)

for  $i \in \llbracket 1, d \rrbracket$ . For each cluster  $i \in \llbracket 1, d \rrbracket$ , a *local controller* decides at each time step to turn ON or OFF optimally some air-conditioners of cluster  $i$ , in order to satisfy a *prescribed proportion* of devices with status ON in the cluster. The *prescribed proportion* of devices ON in each cluster, at each time step, is computed by a *central controller* controlling the average rooms temperatures in each cluster,  $X^i := \frac{1}{N_i} \sum_{j=1}^{N_i} X^{ij}$ , where  $X^{ij}_t$  is the room temperature associated to load  $j \in \llbracket 1, N_i \rrbracket$  of cluster  $i \in \llbracket 1, d \rrbracket$ .  $(X^{ij})_{0 \leq t \leq T}$  is supposed to follow the usual thermal dynamics (see [102] and references therein)

$$X_{t,j}^i = x_{0,j}^i + \int_0^t -\theta_i (X_{s,j}^i - x_{out}^i) - \kappa_i P_{\max}^i \alpha_{s,i,j} ds + \sigma_{i,j} W_{t,j}^i, \quad t \in [0, T], \quad (6.1)$$

where for any  $j \in \llbracket 1, N_i \rrbracket$ ,  $\sigma^{ij} > 0$ ,  $W^{ij}$  are independent real Brownian motions representing model errors and temperature fluctuations inside the room due to local behavior (window, door opening etc.);  $x_0^{ij}$  is the initial temperature;  $\kappa^i$  is the heat exchange parameter;  $x_{out}^i$  denotes the outdoor air temperature;  $1/\theta^i > 0$  is the thermal time constant;  $P_{\max}^i > 0$  denotes the maximal power consumption;  $\alpha_s^{ij} \in \{0, 1\}$  is the status OFF or ON of load  $(i, j)$  at time instant  $s \in [0, T]$ .

We are interested in the problem of the *central controller* who considers the aggregated state process  $X := (X^i)_{1 \leq i \leq d}$ , whose dynamics is obtained by averaging dynamics (6.1) over  $j \in \llbracket 1, N_i \rrbracket$ , for any  $i \in \llbracket 1, d \rrbracket$ ,

$$X_t^i = x_0^i + \int_0^t -\theta (X_s^i - x_{out}^i) - \kappa^i P_{\max}^i \alpha_s^i ds + \sigma^i W_t^i, \quad t \in [0, T], \quad (6.2)$$

where the control process  $\alpha_s = (\alpha_s^i)_{1 \leq i \leq d}$ ,  $s \in [0, T]$  taking values in  $[0, 1]$  prescribes the proportions of devices ON in each cluster;  $x_0^i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_0^{ij}$ ;  $(\sigma^i)^2 = \frac{1}{N_i^2} \sum_{j=1}^{N_i} (\sigma^{ij})^2$ ;  $(W^i)_{1 \leq i \leq d}$  is a  $d$ - $d$  dimensional Brownian motion. The *central controller* problem can be formulated as a specific instantiation of problem (4.1)-(4.2) with the following:

- the controlled process  $X$  driven by a drift coefficient  $b := (b^i)_{1 \leq i \leq d}$  defined on  $[0, T] \times \mathbb{R}^d \times [0, 1]^d$  s.t. for any  $i \in \llbracket 1, d \rrbracket$   $b^i(t, x, a) = -\theta^i (x^i - x_{out}^i) - \kappa^i P_{\max}^i a^i$ , with the notation  $a := (a^i)_{1 \leq i \leq d}$  and  $x := (x^i)_{1 \leq i \leq d}$ ;
- the terminal cost  $g(x) := \frac{1}{d} \sum_{i=1}^d |x^i - \bar{x}^i|^2$  where  $\bar{x} \in \mathbb{R}^d$  denotes given *target* values for the final average temperatures of each cluster;
- the running cost defined on  $[0, T] \times \mathbb{R}^d \times [0, 1]^d$ ,

$$f(t, x, a) := \lambda \sum_{i=1}^d \rho^i a^i - r_t + \frac{1}{d} \sum_{i=1}^d \gamma^i (\rho^i a^i)^2 + \eta^i (x^i - x_{\max}^i)^2 + \eta^i (x^i - x_{\min}^i)^2,$$

where  $\rho^i := \frac{N^i P_{\max}^i}{\sum_{j=1}^d N^j P_{\max}^j}$ ;  $\sum_{i=1}^d \rho^i a^i$  gives the overall current consumption of the population as a proportion of the maximum consumption  $\sum_{j=1}^d N^j P_{\max}^j$ ;  $r : [0, T] \rightarrow \mathbb{R}_*^+$  denotes the target consumption profile for the overall consumption as a proportion of the maximum consumption  $\sum_{j=1}^d N^j P_{\max}^j$ ;  $\lambda > 0$  quantifies the incentive for the overall consumption to track the target consumption profile  $r$ ;  $\gamma^i > 0$  quantifies the quadratic penalty favoring smooth consumption

profiles for cluster  $i$ ;  $\eta^i > 0$  is a parameter penalizing excursions outside of the comfort interval  $[x_{\min}^i, x_{\max}^i]$  for cluster  $i$  average temperature.

Note that  $b$  verifies Assumption 14,  $f$  verifies Assumption 15 and  $g$  Assumption 13.

## 2.6.2 Simulation results

Consider the *central controller* problem on a time horizon  $T = 3600s$ , with a population of heterogeneous air-conditioners composed of  $d = 1, 2, 5, 10, 15, 20$  clusters with  $N^i = 20$  identical loads in each cluster. We specify the chosen parameters. In each case,  $\kappa = 2.5 \cdot C/J$  and  $\sigma^i = 0.1 \cdot C \cdot S^{\frac{1}{2}}$ ;  $x_{\text{out}} = 27 \cdot C$ ;  $\theta^i [s^{-1}]$  is chosen arbitrarily in  $[0.1, 0.97]$ ;  $P_{\max}^i [W]$  is chosen arbitrarily in  $[0.5, 5]$ ;  $x_0 = \bar{x} [C]$  is chosen arbitrarily in  $[16, 27]$ ;  $x_{\min} = \bar{x} - 1.5 \cdot C$ ;  $x_{\max} = \bar{x} + 1.5 \cdot C$ ;  $\eta = 1 \cdot (C)^{-2}$ ;  $\lambda = 20$ ;  $\gamma^i$  is chosen arbitrarily in  $[0.5, 1.5]$ . The target profile,  $r$ , used in simulations is obtained as the sum of a nominal profile corresponding to the standard (uncontrolled) behavior of air-conditioners and a deviation:  $r = r^{\text{nom}} + dev$ . The standard dynamics of an (uncontrolled) air-conditioner is driven by a cycling rule of ON/OFF decisions intended to keep the room temperature in  $[x_{\min}^i, x_{\max}^i]$ . When the air-conditioner is ON, it stays ON at  $P_{\max}^i$  until the temperature reaches  $x_{\min}^i$  then it switches OFF until the temperature reaches  $x_{\max}^i$ . Then, the air-conditioner turns ON again and begins a new cycle. The nominal profile  $r^{\text{nom}}$  has been generated by averaging the consumption of 1000 sets of  $d$  clusters of  $N^i$  heterogeneous air-conditioners simulated independently according to (6.1), with  $(q^{ij})_{0 \leq t \leq T}$  following the cycling rule of ON/OFF decisions and with independent initial conditions for temperature  $x_t^{ij} \sim N(x_0^i, 1)$  and ON/OFF status  $\alpha_t^{ij} \in \{0, 1\}$ . The deviation profile  $dev_t = \frac{20}{100} * \sin(\frac{2\pi t}{T})$  induces a maximal deviation of 20% from the nominal profile and integrates to zero on the time horizon  $[0, T]$  so that the target profile corresponds to the same energy consumed on the period  $[0, T]$  as the nominal profile. The time step is  $\delta t = 60s$ . We have implemented Algorithm 1 with a *backward grid* initiated with  $N(m_n = \bar{x}, Q_n = I_d)$ . For comparison, we have also implemented the standard Monte-Carlo regression scheme using a *forward grid* simulated according to (6.2) with a deterministic control  $\alpha_s$  approximating the nominal dynamics (according to the ON/OFF cycling rule) described previously. In both cases, we have used second order polynomials ( $p = 2$ ) as basis functions for regressions. We have considered  $N = 10^2, 10^3, 5 \times 10^3, 10^4, 2 \times 10^4, 5 \times 10^4, 10^5$  Monte-Carlo paths for the regression grids. To evaluate the statistical performances of the *forward* and *backward grids*, we have implemented each algorithm independently  $N_{\text{grid}} = 100$  times for each value of  $N$ . For each run,  $i = 1, \dots, N_{\text{grid}}$ , the value functions estimate  $(u^i)_{0 \leq k \leq n}$  (and the corresponding gradients) was used to implement the associated strategy  $\alpha^i = (\alpha^i)_{0 \leq k \leq n}$  on  $M = 1000$  i.i.d. simulations of the Brownian motion  $W, \omega^1, \dots, \omega^j, \dots, \omega^M$ . Then the resulting cost  $J(\alpha^i, \omega^j) := g(X_T^{0, \alpha^i}(\omega^j)) + \int_0^T f(r, X^{0, \alpha^i}(\omega^j), \alpha) dr$  has been computed. The expected cost has been estimated as  $E[J(\alpha^i, \omega^j)] \approx \hat{f} := \frac{1}{MN_{\text{grid}}} \sum_{i=1}^{N_{\text{grid}}} \sum_{j=1}^M J(\alpha^i, \omega^j)$ . The variance of  $\hat{f}$  is estimated by  $\hat{\sigma}_f^2$  obtained by replacing, expectations and variances by their empirical approximation based on the sample,  $J(\alpha^i, \omega^j), i \in \llbracket 1, N_{\text{grid}} \rrbracket j \in \llbracket 1, M \rrbracket$ , in the expression  $\hat{\sigma}_f^2 \approx Var(\hat{f}) = \frac{1}{MN_{\text{grid}}} E Var J(\alpha^i, \omega^j) | \alpha^i +$



$\frac{1}{N_{\text{grid}}} \text{Var} \mathbf{E} J(\alpha^i, \omega^j) | \alpha^i$ , for each  $i$  and  $j$ . We have reported on Table 2.1 (resp. Table 2.2) the empirical mean  $\hat{J}$  and within parenthesis the empirical standard deviation  $\hat{\sigma}_{\hat{J}}$  obtained for each considered pair  $(d, N)$  for the *forward grid* (resp. *backward grid*).

One can observe that the *backward grid* performs surprisingly well providing with high precision the lowest expected cost achieved by both methods (or almost) with only  $N = 5 \times 10^3$  paths whatever the dimension  $d$  of the control problem. This is consistent with our intuition based on the idea that localizing the grid around the optimally controlled process paths would bring efficiency and reduce the impact of dimension. The particularity of this problem is that the optimally controlled process is naturally localized in a small region of the state space because, on the one hand a target value,  $\bar{x}$ , is prescribed for the terminal temperatures (by the terminal cost) and on the other hand a target profile is assigned for the overall power consumption. The *backward grid* has the advantage of being initiated around the target state and of following dynamics approaching the optimal strategy. This allows to concentrate the *backward grid* in the small region of interest so that restricting the regression basis to polynomials of order  $p = 2$  seems already enough to obtain reasonable results. However, one can observe some cases where the *forward grid* (for  $N = 10^5$  and  $d \leq 5$ ) has performed slightly better than the *backward grid*. This can be interpreted by the fact that the *forward grid* knows the initial condition  $x_0$  while the *backward grid* has no information about it. To further improve the performances Algorithm 1, an idea would be to find a way to exploit that information on the initial condition. This could constitute the subject of future research.

<b>N</b>	<b>d=1</b>	<b>d=2</b>	<b>d=5</b>	<b>d=10</b>	<b>d=15</b>	<b>d=20</b>
<b>10<sup>2</sup></b>	8.68(0.98)	17.28(1.01)	42.04(1.32)	34.79(0.66)	21.27(0.12)	18.97(0.09)
<b>10<sup>3</sup></b>	7.61(6e <sup>-4</sup> )	8.24(0.07)	14.83(0.64)	28.14(0.64)	37.91(0.60)	34.83(0.45)
<b>5 × 10<sup>3</sup></b>	7.60(3e <sup>-4</sup> )	7.78(2e <sup>-3</sup> )	8.98(0.21)	19.84(0.52)	35.31(0.71)	33.57(0.52)
<b>10<sup>4</sup></b>	7.60(3e <sup>-4</sup> )	7.77(1e <sup>-3</sup> )	7.69(0.06)	16.06(0.38)	32.20(0.63)	30.66(0.59)
<b>2 × 10<sup>4</sup></b>	7.60(3e <sup>-4</sup> )	7.77(2e <sup>-4</sup> )	7.37(0.02)	13.58(0.40)	28.97(0.71)	28.17(0.67)
<b>5 × 10<sup>4</sup></b>	7.60(3e <sup>-4</sup> )	7.79(2e <sup>-4</sup> )	7.28(2e <sup>-3</sup> )	7.96(0.25)	26.69(0.65)	26.21(0.69)
<b>10<sup>5</sup></b>	7.61(3e <sup>-4</sup> )	7.78(1e <sup>-4</sup> )	7.27(8e <sup>-4</sup> )	6.12(0.08)	22.54(0.56)	23.26(0.59)

Table 2.1: Mean,  $\hat{f}$  (standard deviation,  $\hat{\sigma}_{\hat{f}}$ ) of the simulated cost with the *forward grid* strategy

<b>N</b>	<b>d=1</b>	<b>d=2</b>	<b>d=5</b>	<b>d=10</b>	<b>d=15</b>	<b>d=20</b>
<b>10<sup>2</sup></b>	7.61(3e <sup>-4</sup> )	7.78(7e <sup>-4</sup> )	7.41(6e <sup>-3</sup> )	7.31(0.12)	28.14(0.18)	26.01(0.12)
<b>10<sup>3</sup></b>	7.61(3e <sup>-4</sup> )	7.77(2e <sup>-4</sup> )	7.39(1e <sup>-3</sup> )	6.18(3e <sup>-3</sup> )	8.19(6e <sup>-3</sup> )	7.87(1e <sup>-2</sup> )
<b>5 × 10<sup>3</sup></b>	7.61(3e <sup>-4</sup> )	7.77(2e <sup>-4</sup> )	7.38(8e <sup>-4</sup> )	6.17(1e <sup>-3</sup> )	8.15(2e <sup>-3</sup> )	7.74(3e <sup>-3</sup> )
<b>10<sup>4</sup></b>	7.61(3e <sup>-4</sup> )	7.77(2e <sup>-4</sup> )	7.38(5e <sup>-4</sup> )	6.17(1e <sup>-3</sup> )	8.15(2e <sup>-3</sup> )	7.73(3e <sup>-3</sup> )
<b>2 × 10<sup>4</sup></b>	7.61(3e <sup>-4</sup> )	7.77(2e <sup>-4</sup> )	7.38(3e <sup>-4</sup> )	6.17(8e <sup>-4</sup> )	8.15(1e <sup>-3</sup> )	7.73(2e <sup>-3</sup> )
<b>5 × 10<sup>4</sup></b>	7.60(3e <sup>-4</sup> )	7.79(1e <sup>-4</sup> )	7.38(2e <sup>-4</sup> )	6.16(5e <sup>-4</sup> )	8.14(8e <sup>-4</sup> )	7.72(1e <sup>-3</sup> )
<b>10<sup>5</sup></b>	7.61(3e <sup>-4</sup> )	7.79(1e <sup>-4</sup> )	7.39(2e <sup>-4</sup> )	6.16(4e <sup>-4</sup> )	8.14(7e <sup>-4</sup> )	7.72(9e <sup>-4</sup> )

Table 2.2: Mean  $\hat{f}$  (standard deviation,  $\hat{\sigma}_{\hat{f}}$ ) of the simulated cost with the *backward grid* strategy

## 2.7 Appendix

### 2.7.1 A sufficient condition to obtain an equivalent probability

**Lemma 2.7.1.** *We recall that  $b$  was defined in (3.17). Let  $W$  be an  $(\mathbf{F}_t)_{t \in [0, T]}$ -Brownian motion and  $X$  be a solution of*

$$X_t = X_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s) dW_s, \quad t \in [0, T], \quad (7.1)$$

where  $X_0$  is a Gaussian random vector independent of  $W$ . Set  $L_t := \sigma(t)^{-1} b(t, X_t)$ ,  $t \in [0, T]$ . Then, the Doléans exponential  $\mathbf{E} \exp \left( \int_0^t L_s^i dW_s^i - \frac{1}{2} \int_0^t |L_s|^2 ds \right)$  is an  $(\mathbf{F}_t)_{t \in [0, T]}$ -martingale.

*Proof.* Following Corollary 5.14 in [69], it is sufficient to find a constant time step subdivision  $(t_n)_{n \in \mathbf{N}}$  of  $[0, T]$  such that, for all  $n \in \mathbf{N}$ ,

$$\mathbf{E} \exp \left( \frac{1}{2} \int_{t_n}^{t_{n+1}} |L_s|^2 ds \right) < \infty.$$

Combining Jensen's inequality and Fubini's theorem, this is fulfilled in particular if for all  $n \in \mathbb{N}$ ,

$$\frac{1}{\delta} \int_{t_n}^{t_{n+1}} \mathbf{E} \exp \frac{\delta |L_s|^2}{2} ds < \infty,$$

where  $\delta := t_{n+1} - t_n$ . Let  $s \in [0, T]$ . Then,

$$|L_s|^2 \leq 2\delta \sigma^{-1} \int_0^s \|a\|_\infty^2 |X_s|^2 + \|c\|_\infty^2, \quad \mathbf{P} - \text{a.s.},$$

since  $a, c$  are bounded and  $\sigma^{-1}$  is also bounded being continuous on  $[0, T]$ . Furthermore, by item 1. of Lemma 2.3.6 and (3.21),  $X$  is a Gaussian process with mean function  $m^X$  (resp. covariance function  $Q^X$ ) solving the first line of equation (3.1) (resp. (3.2)) with initial condition  $\mathbf{E}(X_0)$  (resp.  $\text{Cov}(X_0)$ ). Taking into account the fact that  $m^X$  is bounded (since continuous), it suffices to find a subdivision such that

$$\mathbf{E} \exp \frac{1}{2} K\delta |Z|^2 < \infty,$$

where  $Z \sim \mathbf{N}(0, I_d)$  and  $K := 4 \sigma^{-1} \int_0^s \|a\|_\infty^2 Q^X > 0$ . This is the case in particular if  $K\delta < 1$ , which ends the proof.  $\square$

### 2.7.2 Proof of the local Lipschitz property of the cost functional $J$

**Lemma 2.7.2.** *Suppose the validity of Assumption 14. Suppose in addition that the functions  $g$  and  $x \mapsto f(t, x, \alpha), (t, \alpha) \in [0, T] \times \mathbf{A}_0$  are locally Lipschitz with polynomial growth gradient (uniformly in  $t$  and  $\alpha$ ). Then, for each  $(t, \alpha) \in [0, T] \times \mathbf{A}_0$ ,*

$$x \mapsto J(t, x, \alpha)$$

*is locally Lipschitz, uniformly in  $t$  and  $\alpha$ .*

*Proof.* We give here a proof of the local Lipschitz property for the term involving the function  $g$  since the other term can be treated in the same way.

Let  $(t, \alpha) \in [0, T] \times \mathbf{A}_0$  and  $x, y$  in a compact set of  $\mathbb{R}^d$ . Let  $K$  be the Lipschitz constant of  $b$ . Using in particular the Cauchy-Schwarz inequality, we get

$$\begin{aligned} \mathbf{E} g(X_T^{t,x,\alpha}) - \mathbf{E} g(X_T^{t,y,\alpha}) &\leq \int_0^1 \mathbf{E} \nabla_x g(aX_T^{t,x,\alpha} + (1-a)X_T^{t,y,\alpha}) (X_T^{t,x,\alpha} - X_T^{t,y,\alpha}) da \\ &\leq e^{KT} \int_0^1 \mathbf{E} \nabla_x g(aX_T^{t,x,\alpha} + (1-a)X_T^{t,y,\alpha}) da |x - y| \end{aligned} \quad (7.2)$$

where we have used the estimate  $X_T^{t,x,\alpha} - X_T^{t,y,\alpha} \leq e^{KT} |x - y|$ , following from the identity

$$X_r^{t,x,\alpha} - X_r^{t,y,\alpha} \leq |x - y| + K \int_t^r (X_s^{t,x,\alpha} - X_s^{t,y,\alpha}) ds, \quad r \in [t, T],$$

together with Gronwall's lemma. In view of (7.2), the point is proved if

$$\int_0^1 \mathbf{E} \nabla_x g(aX_T^{t,x,\alpha} + (1-a)X_T^{t,y,\alpha}) da$$

is bounded uniformly in  $t, x, y, \alpha$ . This follows from polynomial growth of  $\nabla_x g$ , classical moment estimates for  $\sup_{s \in [t, T]} X^{t, z, \alpha}$ ,  $z \in \mathbb{R}^d$  (see for example Corollary 2.5.12 in [71]) and the fact  $x, y$  lie in a compact set.  $\square$

### 2.7.3 A simplified version of the envelope theorem

**Lemma 2.7.3.** *Let  $\Lambda$  be an arbitrary set and  $O$  be an open subset of  $\mathbb{R}^d$ . Let  $x \in \mathbb{R}^d$ . Let  $F : O \times \Lambda \rightarrow \mathbb{R}$  such that for all  $\lambda \in \Lambda$ ,  $F(\cdot, \lambda)$  and  $V : x \mapsto \sup_{\lambda \in \Lambda} F(x, \lambda)$  are differentiable at the point  $x$ . Suppose also that  $\Lambda^*(x) = \{\lambda \in \Lambda, V(x) = F(x, \lambda)\}$  is not empty. Then,*

$$\nabla_x V(x) = \nabla_x F(x, \lambda_x^*),$$

for every  $\lambda_x^* \in \Lambda^*(x)$ .

*Proof.* Let  $x$  as in the proposition statement and  $h \in \mathbb{R}^d$ . Let  $\lambda_x^* \in \Lambda^*(x)$ . Then, using in particular the differentiability of  $F(\cdot, \lambda_x^*)$  at the point  $x$ , we get

$$\begin{aligned} V(x+h) - V(x) &\geq F(x+h, \lambda_x^*) - F(x, \lambda_x^*) \\ &= \nabla_x F(x, \lambda_x^*) \cdot h + o(|h|). \end{aligned} \tag{7.3}$$

By the differentiability of  $V$  at the point  $x$ , (7.3) implies

$$\nabla_x V(x) \cdot h - \nabla_x F(x, \lambda_x^*) \cdot h \geq o(|h|). \tag{7.4}$$

Setting  $h$  to  $-h$  in (7.3) and proceeding as before, we obtain

$$\nabla_x V(x) \cdot h - \nabla_x F(x, \lambda_x^*) \cdot h \leq o(|h|). \tag{7.5}$$

Combining (7.4) and (7.5), we get

$$\frac{\nabla_x V(x) \cdot h - \nabla_x F(x, \lambda_x^*) \cdot h}{|h|} \xrightarrow{h \rightarrow 0} 0,$$

which forces  $\nabla_x V(x) = \nabla_x F(x, \lambda_x^*)$ . This ends the proof.  $\square$



## Chapter 3

# McKean Feynman-Kac probabilistic representations of non-linear partial differential equations

This Chapter is the object of the paper [61].

### 3.1 Introduction and motivations

#### 3.1.1 General considerations

The idea of the present article is to focus on models which have a double macroscopic-microscopic face in the form of *perturbation* of a so called Fokker-Planck type equation that we call *generalized* Fokker-Planck equation. Our ambition is driven by two main reasons.

1. A *modeling* reason: the idea is to observe both from a macroscopic-microscopic point of view phenomena arising from physics, biology, chemistry or complex systems.
2. A *numerical simulation* reason: to provide Monte-Carlo suitable algorithms to approach PDEs.

The *target macroscopic* Fokker-Planck equation is

$$\begin{aligned} \square \quad \partial_t u &= \frac{1}{2} \sum_{i,j=1}^d \partial_{ij}^2 (\sigma \sigma^T)_{ij}(t, x, u) u - \operatorname{div} (b(t, x, u, \nabla u) u) \\ &\quad + \Lambda(t, x, u, \nabla u) u, \quad \text{for } t \in ]0, T], \\ \square \quad u(0, \cdot) &= \mathbf{u}_0, \end{aligned} \tag{1.1}$$

where  $\mathbf{u}_0$  is a Borel probability measure  $\sigma : [0, T] \times \mathbb{R}^d \times \mathbb{R} \rightarrow M_{d,p}(\mathbb{R})$ ,  $b : [0, T] \times \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^d$ ,  $\Lambda : [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}$  and  $\nabla$  denotes the gradient operator. The initial condition in (1.1) means that for every continuous bounded real function  $\phi$  we have  $\int \phi(x) u(t, x) dx \rightarrow \int \phi(x) \mathbf{u}_0(dx)$  when

$t \rightarrow 0$ . When  $\mathbf{u}_0$  admits a density, we denote it by  $u_0$ . The unknown function  $u : ]0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  is supposed to run in  $L^1(\mathbb{R}^d)$  considered as a subset of the space of finite Radon measures  $\mathbf{M}(\mathbb{R}^d)$ . The idea consists in finding a probabilistic representation via the solution of a *Stochastic Differential Equation (SDE)* whose coefficients do not depend only on time and the position of the *particle* but also on its probability law. The *target microscopic equation* we have in mind is

$$\begin{aligned} \square & Y_t = Y_0 + \int_0^t \sigma(s, Y_s, u(s, Y_s)) dW_s + \int_0^t b(s, Y_s, u(s, Y_s)) ds \\ & Y_0 \sim \mathbf{u}_0 \\ \square & \phi(x)u(t, x)dx = \mathbf{E} \left\{ \phi(Y_t) \exp \int_0^t \Lambda(s, Y_s, u(s, Y_s), \nabla u(s, Y_s)) ds \right\}, \quad \text{for } t \in ]0, T], \end{aligned} \tag{1.2}$$

for any continuous bounded real valued test function  $\phi$ . Sometimes we denominate the third line equation of (1.2) the *linking equation*. When  $\Lambda = 0$ , in equation (1.2), the linking equation simply says that  $u(t, \cdot)$  coincides with the density of the marginal distribution  $\mathbf{L}(Y_t)$ . In this specific case, equation (1.2) reduces to a *McKean Stochastic Differential Equation (MSDE)*, which is in general an SDE whose coefficients, at time  $t$ , depend, not only on  $(t, Y_t)$ , but also on the marginal law  $\mathbf{L}(Y_t)$ . With more general functions  $\Lambda$ , the role of the linking equation is more intricate since the whole history of the process  $(Y_s)_{0 \leq s \leq t}$  is involved. This fairly general type of equations will be called *McKean Feynman-Kac Equation (MFKE)* to emphasize the fact that  $u(t, x)dx$  now corresponds to a non-conservative Feynman-Kac measure.

An interesting feature of MSDEs (which means  $\Lambda = 0$ ) is that the law of the process  $Y$  can often be characterized as the limiting empirical distribution of a large number of interacting particles, whose dynamics are described by a coupled system of classical SDEs. When the number of particles grows to infinity, the particles behave closely to a system of independent copies of  $Y$ . This constitutes the so called *propagation of chaos* phenomenon, already observed in the literature when the drift and diffusion coefficients are Lipschitz dependent on the solution marginal law, with respect to the Wasserstein metric, see e.g. [68, 83, 84, 105, 86]. Propagation of chaos is a common phenomenon arising in many physical contexts, see for instance [2] concerning Nelson stochastic mechanics.

When  $\Lambda = 0$ , equation (1.1) is a *non-linear* Fokker-Planck equation, it is conservative and it is known that, under mild assumptions, it describes the dynamics of the marginal probability densities,  $u(t, \cdot)$ , of the process  $Y$ . This correspondence between PDE (1.1) with MSDE (1.2) and interacting particles has extensive interesting applications. In physics, biology or economics, it is a way to relate a microscopic model involving interacting particles to a macroscopic model involving the dynamics of the underlying density. Numerically, this correspondence motivates Monte-Carlo approximation schemes for PDEs. In particular, [24] has contributed to develop stochastic particle methods in the spirit of McKean to provide original numerical schemes approaching a PDE related to Burgers equation providing also the rate of convergence.

Below we list some situations of particular interest where such correspondence holds.

### 3.1.2 Some motivating examples

#### Burgers equation

We fix  $d = p = 1$  and let  $\nu > 0$  and  $u_0$  be a probability density on  $\mathbb{R}$ . We consider two equivalent specific cases of (1.1). The first  $\sigma \equiv \nu$ ,  $b \equiv 0$ ,  $\Lambda(t, x, u, z) = z$ . The second  $\sigma \equiv \nu$ ,  $b(t, x, u) = \frac{u}{2}$ ,  $\Lambda = 0$ . Both instantiations correspond to the the *viscid Burgers equation* in dimension  $d = 1$ , given by

$$\begin{aligned} \partial_t u &= \frac{\nu}{2} \partial_{xx} u - u \partial_x u, & (t, x) \in [0, T] \times \mathbb{R}, \\ u(0, \cdot) &= \mathbf{u}_0. \end{aligned} \quad (1.3)$$

#### Generalized Burgers-Huxley equation

We fix  $d = p = 1$  and let  $\nu > 0$  and  $u_0$  be a probability density on  $\mathbb{R}$ . We consider the particular cases of (1.1) where  $\sigma \equiv \nu$ ,  $b(t, x, u) = \alpha \frac{u^n}{n+1}$ ,  $\Lambda(t, x, u) = \beta(1 - u^n)(u^n - \gamma)$ , with fixed reals  $\alpha, \beta, \gamma$  and a non-negative integer  $n$ . This instantiation corresponds to a natural extension of Burgers equation called *Generalized Burgers-Huxley equation* or *Burgers-Fisher equation* which is of great importance to represent non-linear phenomena in various fields such as biology [4, 87], physiology [70] and physics [111]. These equations have the particular interest to describe the interaction between the reaction mechanisms, convection effect, and diffusion transport. Those are non-linear and non-conservative PDEs of the form

$$\begin{aligned} \partial_t u &= \frac{\nu}{2} \partial_{xx} u - \alpha u^n \partial_x u + \beta u(1 - u^n)(u^n - \gamma), & (t, x) \in [0, T] \times \mathbb{R}, \\ u(0, \cdot) &= \mathbf{u}_0. \end{aligned} \quad (1.4)$$

#### Fokker-Planck equation with terminal condition

The present example does not properly integrate the framework of (1.1). In terms of application, we are interested by inverse problems that can be formulated by a PDE with terminal condition

$$\begin{aligned} \square \quad \partial_t u &= \frac{1}{2} \sum_{i,j=1}^d \partial_{ij}^2 (\sigma \sigma^t)_{i,j}(t, x) u - \operatorname{div}(b(t, x)u) \\ &\quad + \Lambda(t, x)u, & \text{for } t \in ]0, T[, \\ \square \quad u(T, \cdot) &= \mathbf{u}_T, \end{aligned} \quad (1.5)$$

where  $\mathbf{u}_T$  is a prescribed probability measure. Solving that equation by analytical means constitutes a delicate task. A probabilistic representation may help for studying well-posedness or providing numerical schemes.

Backward simulation of diffusions is a subject of active research in various domains of physical sciences and engineering, as heat conduction [14], material science [95] or hydrology [5]. In particular, *hydraulic inversion* is interested in inverting a diffusion phenomenon representing the concentration of a pollutant to identify the pollution source location when the final concentration profile is observed. The problem is in general ill-posed because either the solution is not unique or the solution is not stable. For this type of problem, the existence is ensured by the fact that the observed contaminant



has necessarily originated from some place at a given time (as soon as the model is correct). To correct the lack of well-posedness two regularization procedures have been proposed in the literature: the first one relies on the notion of quasi-solution, introduced by Tikhonov [107], the second one on the method of quasi-reversibility, introduced by Lattes and Lions [74]. Besides well-posedness, a second crucial issue consists in providing a numerical approximating scheme to the backward diffusion equation. A probabilistic representation of (1.5) via the time-reversal of a diffusion could show those issues under a new light.

**The stochastic Fokker-Planck with multiplicative noise**

We fix  $p = d$ ,  $\sigma(t, x, u) = \Phi(u)Id_d$ , where  $\Phi : \mathbb{R} \rightarrow \mathbb{R}$  and  $b = \Lambda \equiv 0$ . Typical examples are the case of classical porous media type equation (resp. fast diffusion equation), when  $\Phi(u) = u^q$ ,  $1 \leq q$  (resp.  $0 < q < 1$ ). The (singular) case  $\Phi(u) = \gamma H(u - e_c)$ ,  $H$  being the Heaviside function and  $e_c$  a given threshold in  $\mathbb{R}$ , appears in the science of complex systems, more precisely in the so called *self-organized criticality*, see e.g. [6, 29, 7].

$$\begin{aligned} \partial_t u &= \frac{\nu}{2} \Delta(H(u - e_c)u) \\ u(0, \cdot) &= \mathbf{u}_0. \end{aligned} \tag{1.6}$$

The phenomenon of *self-organized criticality* often is described in two scale phases: a fast dynamics (of *avalanch type*) described by the PDE (1.6) and a slower motion of *sand storming* modeled by the addition of a supplementary stochastic noise  $\Lambda(t, x; \omega)$ . In that case the *target macroscopic* equation is

$$\begin{aligned} \partial_t u &= \frac{\nu}{2} \Delta(H(u - e_c)u) + \Lambda(t, x; \omega)u \\ u(0, \cdot) &= \mathbf{u}_0, \end{aligned} \tag{1.7}$$

where  $\Lambda(t, x; \omega)$  is a quenched realization of a space-time colored (ideally white) noise. The SPDE will be represented by a MSDE in random environment, see Section 3.6.

**3.1.3 Structure of the paper**

In the rest of the paper, to simplify notations, most of the results are stated in the one-dimensional setting. The generalization to the multi-dimensional case is straightforward.

The paper is organized as follows. Next section presents a brief review of basic situations where Fokker-Planck equations can be represented by MSDEs which in turn can be represented by interacting particles systems. Section 3.3, considers the case of generalized Fokker-Planck equations in the sense of (1.1) with a non-zero term  $\Lambda$  allowing to take into account non-conservative PDEs including a large class of semi-linear PDEs. Section 3.5 highlights the correspondence between MFKEs and MSDEs with jumps which paves the way to a great variety of numerical approximations schemes for non-linear PDEs. Section 3.4 is devoted to a particular inverse problem which consists in modeling backwardly in time the evolution of a Fokker-Planck equation with a given terminal condition. This problem can be related to a time-reversed SDE which in turn can be represented by a MSDE. In Section 3.6 we analyze the well-posedness of generalized Fokker-Planck equation where the term

$\Lambda$  in (1.1) may involve an exogenous noise resulting in a stochastic non-linear PDE. Finally, in Section 3.7, we consider a stochastic control problem for which the associated Hamilton-Jacobi-Bellman equation can be represented by a MFKE.

### 3.2 McKean representations of non-linear Fokker-Planck equations

In this section, we recall some standard situations where a Fokker-Planck PDE can be represented by an SDE which in turn can be approached by an interacting particles system.

#### 3.2.1 Probabilistic representation of linear Fokker-Planck equations

Suppose there exists a solution  $(Y_t)_{t \in [0, T]}$  (in law) to the SDE

$$\begin{aligned} \square & Y_t = Y_0 + \int_0^t \sigma(s, Y_s) dW_s + \int_0^t b(s, Y_s) ds, \quad t \in [0, T], \\ \square & Y_0 \sim \mathbf{u}_0, \end{aligned} \tag{2.1}$$

where  $W$  is a real valued Brownian motion on  $[0, T]$  and  $\mathbf{u}_0$  is a probability measure on  $\mathbb{R}$ . A direct application of Itô formula shows that the marginal probability laws  $(\mu(t, \cdot) := \mathbf{L}(Y_t))_{t \in [0, T]}$  generate a distributional solution of the linear Fokker-Planck PDE

$$\begin{aligned} \square & \partial_t \mu = \frac{1}{2} \partial_{xx}^2 (\sigma^2(t, x) \mu) - \partial_x (b(t, x) \mu) \\ \square & \mu(0, dx) = \mathbf{u}_0(dx). \end{aligned} \tag{2.2}$$

This naturally suggests a Monte Carlo algorithm to approximate the above linear PDE, consisting in simulating  $N$  i.i.d. particles  $(\xi^i)_{i=1, \dots, N}$  with  $N$  i.i.d. Brownian motions  $(W^i)_{i=1, \dots, N}$  i.e.

$$\begin{aligned} \square & \xi_t^i = \xi_0^i + \int_0^t \sigma(s, \xi_s^i) dW_s^i + \int_0^t b(s, \xi_s^i) ds \\ \square & \xi_0^i \text{ i.i.d. } \sim \mathbf{u}_0 \\ \square & \mu_t^N = \frac{1}{N} \sum_{j=1}^N \delta_{\xi_t^j}. \end{aligned} \tag{2.3}$$

Then the law of large numbers provides the convergence of the empirical approximation  $\mu_t^N \xrightarrow[N \rightarrow \infty]{} \mu(t, \cdot)$ , the solution of the Fokker-Planck equation (2.2).

#### 3.2.2 McKean probabilistic representation of non-linear Fokker-Planck equation

We consider the non-linear SDE in the sense of McKean (MSDE)

$$\begin{aligned} \square & Y_t = Y_0 + \int_0^t \sigma(s, Y_s, (K \star \mu)(s, Y_s)) dW_s + \int_0^t b(s, Y_s, (K \star \mu)(s, Y_s)) ds \\ \square & Y \sim \mathbf{u}_0 \\ \square & \mu(t, \cdot) \text{ is the probability law of } Y_t, \quad t \in [0, T], \end{aligned} \tag{2.4}$$

whose solution is a couple  $(Y, \mu)$ . Here  $\sigma, b$  are Lipschitz,  $K : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  denotes a Lipschitz continuous convolution kernel such that  $(K * \mu)(t, y) := \int K(y, z)\mu(t, dz)$  for any  $y \in \mathbb{R}$ . We emphasize that this type of regularized dependence of the drift and diffusion coefficients on  $\mu$  is essentially different (and in general easier to handle) from a pointwise dependence where the coefficients  $b$  or  $\sigma$  may depend on the value of the marginal density at the current particle position  $\frac{d\mu}{dx}(s, Y_s)$ . This regularized or non-local dependence on the time-marginals  $\mu(t, \cdot)$  is a particular case of the framework when the diffusion and drift coefficients are Lipschitz with respect to  $\mu(t, \cdot)$  according to the Wasserstein metric.

Again, by Itô formula, given a solution  $(Y, \mu)$  of (2.4),  $\mu$  solves the non-local non-linear PDE

$$\begin{aligned} \square \quad \partial_t \mu &= \frac{1}{2} \partial_{xx}^2 \sigma^2(t, x, K * \mu) \mu - \partial_x b(t, x, K * \mu) \mu \\ \square \quad \mu(0, dx) &= \mathbf{u}_0(dx), \end{aligned} \tag{2.5}$$

in the sense of distributions. In this setting, the well-posedness of (2.4) relies on a fixed point argument in the space of trajectories under the Wasserstein metric, see e.g. [105], at least in the case when the diffusion term does not depend on the law. We will denominate this situation as the *traditional* setting.

Deriving a Monte-Carlo approximation scheme from this probabilistic representation already becomes more tricky since it can no more rely on independent particles but should involve an interacting particles system as initially proposed in [68, 105]. Consider  $N$  interacting particles  $(\xi_t^{i,N})_{i=1, \dots, N}$  with  $N$  i.i.d. Brownian motions  $(W^i)$ , i.e.

$$\begin{aligned} \square \quad \xi_t^{i,N} &= \xi_0^{i,N} + \int_0^t \sigma(s, \xi_s^{i,N}, (K * \mu_s^N)(\xi_s^{i,N})) dW_s^i + \int_0^t b(s, \xi_s^{i,N}, (K * \mu_s^N)(\xi_s^{i,N})) ds \\ \square \quad \xi_0^{i,N} &\text{ i.i.d. } \sim \mathbf{u}_0 \\ \square \quad \mu_t^N &= \frac{1}{N} \sum_{j=1}^N \delta_{\xi_t^{j,N}} \end{aligned} \tag{2.6}$$

with  $(K * \mu_t^N)(y) = \frac{1}{N} \sum_{j=1}^N K(y, \xi_t^{j,N})$ . The above system defines a so-called *weakly interacting* particles system, as pointed out in [89]. This terminology underlines the fact that any particle interacts with the rest of the population with a vanishing impact of order  $1/N$ . In this setting, at least when the diffusion coefficient does not depend on the law, [105] proves the so called *chaos propagation* which means that  $(\xi_t^{i,N})_{i=1, \dots, N}$  asymptotically behaves as an i.i.d. sample according to  $\mu(t, \cdot)$  as the number of particles  $N$  grows to infinity, where  $\mu$  is the solution of the regularized non-linear PDE (2.5). This in particular implies the convergence of the empirical measures  $\mu_t^N \xrightarrow[N \rightarrow \infty]{} \mu(t, \cdot)$  with the rate  $C/\sqrt{N}$  inherited from the law of large numbers.

As already announced, the case where the coefficients depend pointwise on the density law  $u(t, \cdot)$  of  $\mu(t, \cdot)$ ,  $t > 0$ , is far more singular. Indeed the dependence of the coefficients on the law of  $Y$  is no more continuous with respect to the Wasserstein metric. In this context, well-posedness results rely generally on analytical methods. One important contribution in this direction is reported in [66], where strong existence and pathwise uniqueness are established when the diffusion coefficient  $\sigma$

and the drift  $b$  exhibit pointwise dependence on  $u$  but are assumed to satisfy strong smoothness assumptions together with the initial condition. In this case, the solution  $u$  is a classical solution of the PDE

$$\begin{aligned} \square & \partial_t u = \frac{1}{2} \partial_{xx}^2 \sigma^2(t, x, u(t, x)) u - \partial_x b(t, x, u(t, x)) u \\ \square & u(0, x) = \mathbf{u}_0(dx), \end{aligned} \tag{2.7}$$

which is formally derived from (2.5) setting  $K(x, y) = \delta_0(x - y)$ . Let us fix  $K^\varepsilon$  being a mollifier (depending on a window-width parameter  $\varepsilon$ ), such that  $K^\varepsilon(x, y) = \frac{1}{\varepsilon^d} \varphi\left(\frac{x-y}{\varepsilon}\right) \xrightarrow{\varepsilon \rightarrow 0} \delta_0(x - y)$ . As

in (2.6), we consider the  $N$  interacting particles  $(\xi_t^{i,N})_{i=1, \dots, N}$  solving

$$\begin{aligned} \square & \xi_t^{i,N} = \xi_0^{i,N} + \int_0^t \sigma(s, \xi_s^{i,N}, u_s^{N,\varepsilon}(\xi_s^{i,N})) dW_s^i + \int_0^t b(s, \xi_s^{i,N}, u_s^{N,\varepsilon}(\xi_s^{i,N})) ds \\ \square & \xi_0^{i,N} \text{ i.i.d. } \sim \mathbf{u}_0 \\ \square & u_t^{N,\varepsilon} = \frac{1}{N} \sum_{j=1}^N K^\varepsilon(\cdot, \xi_t^{j,N}). \end{aligned} \tag{2.8}$$

Under the smooth assumptions on  $b, \sigma, u_0$  mentioned before and non-degeneracy of  $\sigma$ , [66] proved the convergence of the regularized particle approximation  $u_t^{N,\varepsilon}$  to the solution  $u$  of the pointwise non-linear PDE (2.7) as soon as  $\varepsilon(N) \xrightarrow{N \rightarrow \infty} 0$  slowly enough. According to [89], the system (2.8) defines a so-called *moderately interacting* particle system with  $u_t^{N,\varepsilon}(x) = \frac{1}{N\varepsilon^d} \sum_{j=1}^N \varphi\left(\frac{x-\xi_t^{j,N}}{\varepsilon}\right)$ . Indeed as the window width of the kernel,  $\varepsilon$ , goes to zero, the number of particles that significantly impact a single one is of order  $N\varepsilon^d$  with a strength of interaction of order  $\frac{1}{N\varepsilon^d}$ . In contrast, when  $\varepsilon$  is fixed, we recover the weakly interacting situation in which case the strength of interaction of each particle is of order  $\frac{1}{N}$  which is smaller than  $\frac{1}{N\varepsilon^d}$ . In this case of moderate interaction, the propagation of chaos occurs with a slower rate than  $C/N$  and depends exponentially on the space dimension. [66] constitutes an extension of the weak propagation of chaos of moderately interacting particles proved in [89] for the limited case of identity diffusion matrix.

The peculiar case where the drift vanishes and the diffusion coefficient  $\sigma(u(t, Y_t))$  has a pointwise dependence on the law density  $u(t, \cdot)$  of  $Y_t$  has been more particularly studied in [17] for classical porous media type equations and [22, 10, 16, 15, 9] who obtain well-posedness results for measurable and possibly singular functions  $\sigma$ . In that case the solution  $u$  of the associated PDE (1.1), is understood in the sense of distributions.

### 3.3 McKean Feynman-Kac representations for non-conservative and non-linear PDEs

The idea of generalizing MSDEs to MFKEs (1.2) was originally introduced in the sequence of papers [76, 75, 77], with an earlier contribution in [12], where  $\Lambda(t, x, u, \nabla u) = \xi_t(x)$ ,  $\xi$  being the sample of a Gaussian noise random field, white in time and regular in space, see Section 3.6. The goal was

to provide some probabilistic representation for non-conservative non-linear PDEs (1.1) by introducing some exponential weights defining Feynman-Kac measures instead of probability measures. An interesting aspect of this strategy is that it is potentially able to represent an extended class of second order non-linear PDEs. One particularity of MFKE equations is that the probabilistic representation involves the past of the process (via the exponential weights). In this context, it is worth to quote the recent paper [63] which proposes a probabilistic representation, which also includes a dependence on the past, in relation with Keller-Segel model with application to chemotaxis.

It is important to consider carefully the two major features differentiating the MFKE (1.2) from the traditional setting of MSDEs. To recover the traditional setting one has to do the following.

1. First, one has to put  $\Lambda = 0$  in the third line equation of (1.2) Then  $u(t, \cdot)$  is explicitly given by the third line equation of (1.2) and reduces to the density of the marginal distribution,  $L(Y_t)$ . When  $\Lambda \neq 0$ , the relation between  $u(t, \cdot)$  and the process  $Y$  is more complex. Indeed, not only does  $\Lambda$  embed an additional non-linearity with respect to  $u$ , but it also involves the whole past trajectory  $(Y_s)_{0 \leq s \leq t}$  of the process  $Y$ .
2. Secondly, one has to replace the pointwise dependence  $b(s, Y_s, u(s, Y_s))$  in equation (1.2) with a mollified dependence  $b(s, Y_s, \int_{\mathbb{R}^d} K(Y_s - y)u(s, y)dy)$ , where the dependence with respect to  $u(s, \cdot)$  is Wasserstein continuous. Here  $K : \mathbb{R} \rightarrow \mathbb{R}$  is a convolution kernel.

One interesting aspect of probabilistic representation (1.2) is that it naturally yields numerical approximation schemes involving weighted interacting particle systems. More precisely, we consider  $N$  interacting particles  $(\xi^{i,N})_{i=1, \dots, N}$  with  $N$  i.i.d. Brownian motions  $(W^i)_{i=1, \dots, N}$ , i.e.

$$\begin{aligned}
 \square \quad & \xi_t^{j,N} = \xi_0^{j,N} + \int_0^t \sigma(s, \xi_s^{j,N}, u^{N,\varepsilon}(\xi_s^{j,N})) dW_s^j + \int_0^t b(s, \xi_s^{j,N}, u^{N,\varepsilon}(\xi_s^{j,N})) ds \\
 \square \quad & \xi_0^{i,N} \text{ i.i.d. } \sim \mathbf{u}_0 \\
 \square \quad & u_t^{N,\varepsilon}(\xi^i) = \frac{1}{N} \sum_{j=1}^N \omega_t^{j,N} K^\varepsilon(\xi_t^{i,N} - \xi_t^{j,N}),
 \end{aligned} \tag{3.1}$$

where the mollifier  $K^\varepsilon$  is such that  $K^\varepsilon(x) = \frac{1}{\varepsilon^d} \varphi(\frac{x}{\varepsilon}) \xrightarrow{\varepsilon \rightarrow 0} \delta_0$  and the weights  $\omega_t^{j,N}$  for  $j = 1, \dots, N$  verify

$$\begin{aligned}
 \omega_t^{j,N} &:= \exp \int_0^t \Lambda(r, \xi_r^{j,N}, u_r^{\varepsilon,N}(\xi_r^{j,N}), \nabla u_r^{\varepsilon,N}(\xi_r^{j,N})) dr \\
 &= \omega_s^{j,N} \exp \int_s^t \Lambda(r, \xi_r^{j,N}, u_r^{\varepsilon,N}(\xi_r^{j,N}), \nabla u_r^{\varepsilon,N}(\xi_r^{j,N})) dr.
 \end{aligned}$$

[78, 77] consider the case of pointwise semilinear PDEs of the form

$$\begin{aligned}
 \partial_t u &= \frac{1}{2} \partial_{xx}^2 (\sigma^2(t, x)u) - \partial_x b(t, x)u + \Lambda(t, x, u, \nabla u)u \\
 u(0, x) &= u_0(x),
 \end{aligned} \tag{3.2}$$

for which the target probabilistic representation is

$$\begin{aligned} \square & Y_t = Y_0 + \int_0^t \sigma(s, Y_s) dW_s + \int_0^t b(s, Y_s) ds \\ \square & Y_0 \sim \mathbf{u}_0 \\ \square & \int_{\mathbf{R}^d} \phi(x) u_t(x) dx := \mathbf{E} \left[ \phi(Y_t) \exp \left( - \int_0^t \Lambda(s, Y_s, u_s(Y_s), \nabla u_s(Y_s)) ds \right) \right]. \end{aligned} \tag{3.3}$$

We set

$$L_t f := \frac{1}{2} \sigma^2(t, x) f''(x) + b(t, x) f'(x), t \in ]0, T[, \text{ for any } f \in C^2(\mathbf{R}). \tag{3.4}$$

Let us consider the family of Markov transition functions  $P(s, x_0, t, \cdot)$  associated with  $(L_t)$ , see [78]. We recall that if  $X$  is a process solving the first line of (3.1) with  $X_s \equiv x_0 \in \mathbf{R}$ , then  $\int_{\mathbf{R}} P(s, x_0, t, x) f(x) dx = \mathbf{E}(f(X_t))$ ,  $t \geq s$ , for every bounded Borel function  $f: \mathbf{R} \rightarrow \mathbf{R}$ .  $u: [0, T] \times \mathbf{R} \rightarrow \mathbf{R}$  will be called **mild solution** of (3.2) (related to  $(L_t)$ ) if for all  $\phi \in C^\infty(\mathbf{R})$ ,  $t \in [0, T]$ ,

$$\begin{aligned} \int_{\mathbf{R}^d} \phi(x) u(t, x) dx &= \int_{\mathbf{R}^d} \phi(x) \int_{\mathbf{R}^d} \mathbf{u}_0(dx_0) P(0, x_0, t, dx) \\ &+ \int_{[0, t] \times \mathbf{R}^d} \int_{\mathbf{R}^d} \phi(x) P(s, x_0, t, dx) \Lambda(s, x_0, u(s, x_0), \nabla u(s, x_0)) u(s, x_0) dx_0 ds. \end{aligned}$$

The following theorem states conditions ensuring equivalence between (3.3) and (3.2) together with the convergence of the related particle approximation (3.1).

**Theorem 3.3.1.** *We suppose that  $\sigma$  and  $b$  are Lipschitz with linear growth and  $\Lambda$  is bounded.*

1. *Let  $u: [0, T] \times \mathbf{R} \rightarrow \mathbf{R} \in L^1([0, T]; W^{1,1}(\mathbf{R}^d))$ .  $u$  is a mild solution of PDE (3.2) if and only if  $u$  verifies (3.3).*
2. *Suppose that  $\sigma \geq c > 0$  and  $\Lambda$  is uniformly Lipschitz w.r.t. to  $u$  and  $\nabla u$ . There is a unique mild solution in  $L^1([0, T]; W^{1,1}(\mathbf{R}) \cap L^\infty([0, T] \times \mathbf{R}))$  of (3.2), therefore also of (3.3).*
3. *Under the same assumption of item 2., the particle approximation  $u^{N,\varepsilon}$  (3.1) converges in  $L^1([0, T]; W^{1,1}(\mathbf{R}))$  to the solution of (3.2) as  $N \rightarrow \infty$  and  $\varepsilon(N) \rightarrow 0$  slowly enough.*

Item 1. was the object of Theorem 3.5 in [78]. Item 2. (resp. item 3.) was treated in Theorem 3.6 (resp. Corollary 3.5) in [78].

**Remark 3.3.2.** *The error induced by the discrete time approximation of the particle system was evaluated in [77].*

[79] considers the case where  $b$  is replaced by  $b + b_1$  where  $b$  is only supposed bounded Borel, without regularity assumption on the space variable. In particular they treat the pointwise semilinear PDEs of the form

$$\begin{aligned} \square & \partial_t u = \frac{1}{2} \partial_{xx}^2 (\sigma^2(t, x) u) - \partial_x (b(t, x) + b_1(t, x, u)) u + \Lambda(t, x, u) u \\ \square & u(0, x) = u_0(x), \end{aligned} \tag{3.5}$$

for which the target probabilistic representation is

$$\begin{aligned} \square \quad & Y_t = Y_0 + \int_0^t \sigma(s, Y_s) dW_s + \int_0^t [b(s, Y_s) + b_1(s, Y_s, u(s, Y_s))] ds \\ \square \quad & Y_0 \sim \mathbf{u}_0 \\ \square \quad & \int \phi(x) u_t(x) dx := \mathbf{E} \left[ \phi(Y_t) \exp \left\{ - \int_0^t \Lambda(s, Y_s, u_s(Y_s)) ds \right\} \right]. \end{aligned} \tag{3.6}$$

The following theorem states conditions ensuring equivalence between (3.6) and (3.5) together with well-posedness conditions for both equations.

**Theorem 3.3.3.** *We formulate the following assumptions.*

1. The PDE in the sense of distributions  $\partial_t u = L_t^* u_t$  admits as unique solution  $u \equiv 0$ , where  $L_t$  was defined in (3.4).
2.  $b$  is bounded measurable and  $\sigma$  is continuous  $\sigma \geq c > 0$  for some constant  $c > 0$ .
3.  $b, \Lambda : [0, T] \times \mathbf{R} \times \mathbf{R} \rightarrow \mathbf{R}$  is uniformly bounded, Lipschitz with respect to the third argument.
4. The family of Markov transition functions associated with  $(L_t)$ , are of the form  $P(s, x_0, t, dx) = p(s, x_0, t, x) dx$ , i.e. they admit measurable densities  $p$ .
5. The first order partial derivatives of the map  $x_0 \mapsto p(s, x_0, t, x)$  exist in the distributional sense.
6. For almost all  $0 \leq s < t \leq T$  and  $x_0, x \in \mathbf{R}$  there are constants  $C_u, c_u > 0$  such that

$$p(s, x_0, t, x) \leq C_u q(s, x_0, t, x) \quad \text{and} \quad |\partial_{x_0} p(s, x_0, t, x)| \leq C_u \sqrt{\frac{1}{t-s}} q(s, x_0, t, x), \tag{3.7}$$

where  $q(s, x_0, t, x) := \frac{c_u(t-s)}{\pi} \frac{1}{2} e^{-\frac{|x-x_0|^2}{t-s}}$  is a Gaussian probability density.

The following results hold.

1. Let  $u \in (L^1 \cap L^\infty)([0, T] \times \mathbf{R})$ .  $u$  is a solution of PDE (3.5) in the sense of distributions if and only if  $u$  verifies (3.6) for a solution  $Y$  in the sense of probability laws.
2. There is a unique solution  $u \in (L^1 \cap L^\infty)([0, T] \times \mathbf{R})$  in the sense of distributions of PDE (3.5) (and therefore of (3.6)).

The result 1. (resp. result 2.) was the object of Theorem 12. (resp. Proposition 16., Theorems 13., 22.) of [79].

**Remark 3.3.4.** Under more restrictive assumptions on  $b$ , item 3. of Theorem 13. in [79] states the well-posedness of (3.6) in the sense of strong existence and pathwise uniqueness.

[76] and [75] studied a mollified version of (1.1), whose probabilistic representation falls into the Wasserstein continuous traditional setting mentioned above. Following the spirit of [105], a fixed point argument was carried out in the general case in [76] to prove well-posedness of

$$\begin{aligned} \square \quad & Y_t = Y_0 + \int_0^t \sigma(s, Y_s, K * u_s(Y_s)) dW_s + \int_0^t b(s, Y_s, K * u_s(Y_s)) ds \\ & Y_0 \sim \mathbf{u}_0 \\ \square \quad & (K * u_t)(x) := \mathbf{E} \left[ K(x - Y_t) \exp \left( - \int_0^t \Lambda(s, Y_s, K * u_s(Y_s)) ds \right) \right], \end{aligned} \tag{3.8}$$

where  $K : \mathbb{R} \rightarrow \mathbb{R}$  is a mollified kernel. We remark that if  $(Y, u)$  is a solution of (3.8), then  $u$  is a solution (in the sense of distribution) of

$$\begin{aligned} \partial_t u &= \frac{1}{2} \partial_{xx}^2 (\sigma^2(t, x, K * u)u) - \partial_x (b(t, x, K * u)u) + \Lambda(t, x, K * u)u \\ u(0, x) &= u_0(x). \end{aligned} \tag{3.9}$$

- Remark 3.3.5.**
1. Existence and uniqueness results (in the strong sense and in the sense of probability laws) for the MFKE (3.8) are established under various technical assumptions, see [75].
  2. Chaos propagation for the interacting particle system (3.1) providing an approximation to the regularized PDE (3.9), as  $N \rightarrow \infty$  (for fixed  $K$ ), [76].

### 3.4 McKean representation of a Fokker-Planck equation with terminal condition

Let us consider the PDE with terminal condition (1.5) and  $\Lambda = 0$ .

$$\begin{aligned} \square \quad \partial_t u &= \frac{1}{2} \sum_{i,j=1}^d \partial_{ij}^2 (\sigma \sigma^t)_{ij}(t, x)u - \operatorname{div} (b(t, x)u) \\ \square \quad u(T, dx) &= \mathbf{u}_T(dx), \end{aligned} \tag{4.1}$$

where  $\mathbf{u}_T$  is a given Borel probability measure. In the present section we assume the following.

**Assumption 1.** Suppose that (4.1) admits uniqueness, i.e. that there is at most one solution of (4.1).

**Remark 3.4.1.** Different classes of sufficient conditions for the validity of Assumption 1 are provided in [62]. In particular one significant result is Theorem 4.14 of [62] which states that previous Assumption 1 holds if  $\sigma, b$  are time-homogeneous and the following holds.

- Assumption 2.**
1.  $\Sigma = \sigma \sigma^*$  is strictly non-degenerate.
  2. The functions  $\sigma$  is Lipschitz in space.
  3. The functions  $\sigma, b, (\nabla_x b_i)_{i \in \llbracket 1, d \rrbracket}, (\nabla_x \Sigma_{ij})_{i,j \in \llbracket 1, d \rrbracket}$  are continuous bounded and  $\nabla_x^2 \Sigma$  is Hölder continuous with exponent  $\alpha \in ]0, 1[$ .



Generalizations to the time-inhomogeneous setup are also available.

A natural representation of (4.1) is the following MSDE, where  $\beta$  is a Brownian motion.

$$\begin{aligned} \square \quad Y_t &= \xi - \int_0^t \tilde{b}(s, Y_s; v_s) ds + \int_0^t \sigma(T-s, Y_s) d\beta_s, \quad t \in [0, T] \\ \square \quad \int_{\mathbb{R}^d} v_t(x) \phi(x) dx &= \mathbf{E}(\phi(Y_t)), \quad t \in [0, T] \\ \square \quad \xi &\sim \mathbf{u}_T, \end{aligned} \quad (4.2)$$

where  $\tilde{b}(s, y; v_s) = (\tilde{b}^1(s, y; v_s), \dots, \tilde{b}^d(s, y; v_s))$  is defined as

$$\tilde{b}(s, y; v_s) := \frac{\operatorname{div}_y \sigma \sigma_j^t(T-s, y) v_s(y)}{v_s(y)} - b(T-s, y). \quad (4.3)$$

$j \in \{1, d\}$

For  $d = 1$  previous expression gives

$$\tilde{b}(s, y; v_s) := \frac{\sigma^2(T-s, \cdot) v_s}{v_s}(y) - b(T-s, y). \quad (4.4)$$

**Remark 3.4.2.** (4.2) is in particular fulfilled if  $Y$  is the time reversal process  $\hat{X}_t := X_{T-t}$  of a diffusion satisfying the SDE

$$\begin{aligned} \square \quad X_t &= X_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s, \quad t \in [0, T] \\ \square \quad X_0 &\sim \mathbf{u}_0 \in \mathbf{P}(\mathbb{R}). \end{aligned} \quad (4.5)$$

This happens under locally Lipschitz conditions on  $\sigma$  and  $b$  and minimal regularity conditions on the law density  $p_t$  of  $X_t$ . Indeed in [55], the authors prove that

$$\hat{X}_t = X_T + \int_0^t \tilde{b}(s, \hat{X}_s; p_{T-s}) ds + \int_0^t \sigma(T-s, \hat{X}_s) d\beta_s, \quad t \in [0, T], \quad (4.6)$$

where  $\tilde{b}$  is defined in (4.3) and  $p_t$  is the density of  $X_t$ . We emphasize that the main difference between (4.2) and (4.6) is that in the first equation the solution is a couple  $(Y, v)$ , in the second one, a solution is just  $Y$ ,  $p$  being exogeneously defined by (4.5).

We observe now that a solution  $(Y, v)$  of (4.2) provides a solution  $u$  of (4.1). This justifies indeed the terminology of probabilistic representation.

**Proposition 3.4.3.** 1. Let  $(Y, v)$  be a solution of (4.2). Then  $u(t, \cdot) := v(T-t, \cdot)$ ,  $t \in [0, T]$ , is a solution of (4.1) with terminal value  $\mathbf{u}_T$ .

2. If (4.1) admits at most one solution, then there is at most one  $v$  such that  $(Y, v)$  solves (4.2).

*Proof.* In order to avoid technicalities which complicate the task of the reader we express the proof for  $d = 1$ . We prove 1. since 2. is an immediate consequence of 1.

Let  $\varphi \in C^\infty(\mathbb{R})$  with compact support and  $t \in [0, T]$ . Itô formula gives

$$\mathbf{E}[\varphi(Y_{T-t})] - \int_{\mathbb{R}^d} \varphi(y) \mathbf{u}_T(dy) = \int_0^{T-t} \mathbf{E} \left[ \tilde{b}(s, Y_s; v_s) \varphi(Y_s) + \frac{1}{2} \sigma^2(T-s, Y_s) \varphi''(Y_s) \right] ds. \quad \blacksquare$$

Fixing  $s \in [0, T]$ , we have

$$\begin{aligned} \mathbb{E} \tilde{b}(s, Y_s; v_s) \varphi(Y_s) &= \int_{\mathbb{R}} (\sigma^2(T-s, \cdot) v_s)(y) \varphi(y) dy - \int_{\mathbb{R}} b(T-s, y) \varphi(y) v_s(y) dy \\ &= - \int_{\mathbb{R}} (\sigma^2)(T-s, y) \varphi(y) v_s(y) dy - \int_{\mathbb{R}} b(T-s, y), \varphi(y) v_s(y) dy. \end{aligned}$$

Hence, we have the identity

$$\mathbb{E}[\varphi(Y_{T-t})] = \int_{\mathbb{R}} \varphi(y) \mathbf{u}_T(dy) - \int_0^{T-t} \int_{\mathbb{R}} L_{T-s} \varphi(y) v_s(y) dy ds.$$

Applying the change of variable  $t \mapsto T - t$ , we finally obtain the identity

$$\int_{\mathbb{R}} \varphi(y) v_{T-t}(y) dy = \int_{\mathbb{R}} \varphi(y) \mathbf{u}_T(dy) - \int_t^T \int_{\mathbb{R}^d} L_s \varphi(y) v_{T-s}(y) dy ds.$$

This means that  $t \mapsto u_t$  is a solution of (3.1.2) with terminal value  $\mathbf{u}_T$ .

□

**Remark 3.4.4.** *Precise discussions on existence and uniqueness of (4.2) are provided in [62]. In particular we have the following.*

1. *There is at most one solution (in law)  $(Y, v)$  of (4.2) such that  $v$  is locally bounded in  $[0, T] \times \mathbb{R}^d$ .*
2. *There is at most one strong solution  $(Y, v)$  of (4.2) such that  $v$  is locally Lipschitz in  $[0, T] \times \mathbb{R}^d$ .*

*Item 1. is a consequence of Theorem 10.1.3 of [104]. Item 2. is a consequence of usual pathwise uniqueness arguments for SDEs.*

### 3.5 Probabilistic representation with jumps for non-conservative PDEs

In this section, we outline the link between non-conservative PDEs and non-linear jump diffusions. This kind of representation was emphasized in [37, 34] to design interacting jump particles systems to approximate time-dependent Feynman-Kac measures. For simplicity, we present this correspondence in the simple case of the non-conservative linear PDE (1.1) when the coefficients do not depend on the solution, see (1.1). However, the same ideas could be extended to the non-linear case where the coefficients  $\sigma, b, \Lambda$  may depend on the PDE solution.

Let us consider the SDE

$$\begin{aligned} dX_t &= b(t, X_t)dt + \sigma(t, X_t)dW_t \\ X_0 &\sim \mathbf{u}_0, \end{aligned} \tag{5.1}$$

where  $W$  is a one-dimensional Brownian motion. Assume that (5.1) admits a (weak) solution. Let  $\Lambda$  be a bounded and negative function defined on  $[0, T] \times \mathbb{R}$ . For any  $t \in [0, T]$ , we define the measure,  $\gamma(t, \cdot)$  such that for any real-valued Borel measurable test function  $\phi$

$$\gamma(t, dx) \phi(x) = \mathbb{E} \left[ \phi(X_t) \exp \left( - \int_0^t \Lambda(s, X_s) ds \right) \right]. \tag{5.2}$$

We recall that by Section 3.3 we know that  $\gamma$  is a solution (in the distributional sense) of the linear and non-conservative PDE

$$\begin{aligned} \partial_t \gamma &= \frac{1}{2} \partial_{xx}^2 (\sigma^2(t, x) \gamma) - \partial_x (b(t, x) \gamma) + \Lambda(t, x) \gamma \\ \gamma(0, \cdot) &= \mathbf{u}_0. \end{aligned} \tag{5.3}$$

**Remark 3.5.1.** *If uniqueness of distributional solutions of (5.3) holds, then  $\gamma$  defined by (5.1,5.2) is the unique solution of (5.3).*

Let  $\gamma(t, \cdot)$  be a solution of (5.2) which for each  $t$  is a positive measure. We introduce the family of probability measures  $(\eta(t, \cdot))_{t \in [0, T]}$ , obtained by normalizing  $\gamma(t, \cdot)$ , such that for any real valued bounded and measurable test function  $\phi$  we have

$$\eta(t, dx) \phi(x) := \frac{\int \gamma(t, dx) \phi(x)}{\int \gamma(t, dx)}. \tag{5.4}$$

By simple differentiation of the above ratio and using the fact that  $\gamma$  satisfies (5.3), we obtain that  $\eta$  is a solution in the distributional sense of the integro-differential PDE

$$\begin{aligned} \square \quad \partial_t \eta &= \frac{1}{2} \partial_{xx}^2 (\sigma^2(t, x) \eta) - \partial_x (b(t, x) \eta) + \int \eta(t, dx) \Lambda(t, x) \eta \\ \square \quad \eta_0 &= \mathbf{u}_0. \end{aligned} \tag{5.5}$$

Besides one can express  $\gamma(t, \cdot)$  as a function of  $(\eta(s, \cdot))_{s \in [0, t]}$ . Indeed, since  $\gamma$  solves the linear PDE (5.3) then in particular approaching the constant test function 1, yields

$$\partial_t \int \gamma(t, dx) = \int \gamma(t, dx) \Lambda(t, x) = \int \gamma(t, dx) \int \eta(t, dx) \Lambda(t, x),$$

which gives  $\int \gamma(t, dx) = \exp \int_0^t \int \eta(s, dx) \Lambda(s, x) ds$ . Then by definition (5.4) of  $\eta$ ,

$$\gamma(t, \cdot) = \int \gamma(t, dx) \eta(t, \cdot) = \exp \int_0^t \int \eta(s, dx) \Lambda(s, x) ds \int \eta(t, \cdot). \tag{5.6}$$

We already know that for any solution  $\gamma$  of (5.3) one can build a solution  $\eta$  to (5.5) according to relation (5.4). Conversely, for any solution  $\eta$  of (5.5), by similar manipulations one can build a solution  $\gamma$  of (5.3) according to (5.6). Hence well-posedness of (5.3) is equivalent to well-posedness of (5.5).

We propose now an alternative probabilistic representation to (5.1) and (5.2) of (5.3). Let us introduce the non-linear jump diffusion  $Y$  (if it exists), which evolves between two jumps according to the diffusion dynamics (5.1) and jumps at exponential times with intensity  $-\Lambda_t(Y_t) \geq 0$  to a new point independent of the current position and distributed according to the current law,  $\mathbf{L}(Y_t)$ . More specifically, we consider a process  $Y$  solution of the following non-linear (in the sense of McKean) SDE with jumps

$$\begin{aligned} \square \quad dY_t &= b(t, Y_{t-}) dt + \sigma(t, Y_{t-}) dW_t + \int_{\mathbb{R}} x \mathbf{1}_{|x| > 1} J_t(\mu_{t-}, Y_{t-}, dx) dt + \int_{\mathbb{R}} x \mathbf{1}_{|x| \leq 1} (J_t - \bar{J}_t)(\mu_{t-}, Y_{t-}, dx) dt \\ Y_0 &\sim \mathbf{u}_0 \\ \square \quad \mu_{t-} &= \mathbf{L}(Y_{t-}), \end{aligned} \tag{5.7}$$

where  $J$  denotes the jump measure and  $\bar{J}$  is the associated predictable compensator such that for any probability measure  $\nu$  on  $\mathbb{R}$

$$\bar{J}_t(\nu, y, d(y - y)) = -\Lambda_t(y)\nu(dy), \quad \text{for any } y, y \in \mathbb{R}.$$

Note that well-posedness analysis of the above equation constitutes a difficult task. In particular, [67] analyzes well-posedness and particle approximations of some types of non-linear jump diffusions. However, contrarily to (5.7), the nonlinearity considered in [67] is concentrated on the diffusion matrix (assumed to be Lipschitz in the time-marginals of the process w.r.t. Wasserstein metric) and does not involve the jump measure which is assumed to be given.

Assume that MSKE (5.7) admits a weak solution. By application of Itô formula, we observe that the marginals of  $Y$  are distributional solutions of (5.5). Indeed, for any real valued test function in  $C^\infty(\mathbb{R})$

$$\begin{aligned} \mathbb{E}[\phi(Y_t)] &= \mathbb{E}[\phi(Y_0)] \\ &+ \int_0^t \mathbb{E} \left[ b(s, Y_{s-})\phi(Y_{s-}) + \frac{1}{2}\sigma^2(s, Y_{s-})\phi''(Y_{s-}) \right] ds \\ &+ \int_0^t \mathbb{E} \left[ \phi(Y_{s-} + x)\bar{J}_s(\mu_{s-}, Y_{s-}, dx) \right] ds \\ &- \int_0^t \mathbb{E} \left[ \phi(Y_{s-})\bar{J}_s(\mu_{s-}, Y_{s-}, \mathbb{R}) \right] ds. \end{aligned} \tag{5.8}$$

**Conclusion.**

Suppose that (5.3) admits a unique distributional solution  $\gamma$ ; let  $\eta$  defined by (5.4). Suppose the existence of a (weak) solution  $X$  (resp.  $Y$ ) of (5.1) (resp. (5.7)).

1.  $\eta$  is the unique solution (in the sense of distributions) of (5.5). Moreover  $\int_{\mathbb{R}} \phi(x)\eta(t, dx) = \mathbb{E}[\phi(Y_t)], t \geq 0$ .
2. We obtain the following identities for  $\gamma$  and  $\eta$ :

$$\begin{aligned} \gamma(t, dx)\phi(x) &= \mathbb{E} \left[ \phi(X_t) \exp \left( \int_0^t \Lambda(s, X_s) ds \right) \right] \\ &= \exp \left( \int_0^t \eta(s, dx)\Lambda(s, x) ds \right) \eta_t(\phi) \\ &= \exp \left( \int_0^t \mathbb{E}[\Lambda(s, Y_s)] ds \right) \mathbb{E}[\phi(Y_t)]. \end{aligned} \tag{5.9}$$

Using the above identities allows to design discrete time interacting particles systems with geometric interacting jump processes. In particular, in [36] the authors provide non asymptotic bias and variance theorems w.r.t. the time step and the size of the system, allowing to numerically approximate the time-dependent family of Feynman-Kac measures  $\gamma$ .

### 3.6 McKean SDEs in random environment

#### 3.6.1 The (S)PDE and the basic idea

Let  $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$  be a filtered probability space. We consider a progressively measurable random field  $(\xi(t, x))$ . We want to discuss probabilistic representations of

$$\begin{aligned} \partial_t u &= \frac{1}{2} \Delta(\beta(u)) + \partial_t \xi(t, x) u(t, x), \quad \text{with } \beta(u) = \sigma^2(u)u. \\ u(0, \cdot) &= \mathbf{u}_0. \end{aligned} \tag{6.1}$$

Suppose for a moment that  $\xi$  has random realizations which are smooth in time so that

$$\partial_t \xi(t, x) = \Lambda(t, x; \omega). \tag{6.2}$$

Under some regularity assumptions on  $\Lambda$ , (6.1) can be observed as a randomization of a particular case of the PDE (1.1). For each random realization  $\omega \in \Omega$ , the natural (double) probabilistic representation is

$$\begin{aligned} \square \quad & Y_t = Y_0 + \int_0^t \sigma(u(s, Y_s)) dW_s \\ \square \quad & Y_0 \sim \mathbf{u}_0 \\ \square \quad & \int_{\mathbb{R}} \phi(x) u(t, x) dx = \mathbf{E}^\omega \left[ \int_{\mathbb{R}} \phi(Y_t) \exp \left\{ \int_0^t \Lambda(s, Y_s; \omega) ds \right\} dx \right], \quad \text{for } t \in [0, T], \end{aligned} \tag{6.3}$$

where  $\mathbf{E}^\omega$  denotes the expectation with frozen  $\omega$ . However the assumption (6.2) is not realistic and we are interested in  $\partial_t \xi$  being a white noise in time. Let  $N \in \mathbb{N}^*$ . Let  $B^1, \dots, B^N$  be  $N$  independent  $(\mathcal{F}_t)$ -Brownian motions,  $e^1, \dots, e^N$  be functions in  $C^2(\mathbb{R})$ . In particular they are  $H^{-1}$ -multiplier, i.e. the maps  $\phi \rightarrow \phi e^i$  are continuous in  $H^{-1}$ .

We define the random field  $\xi(t, x) = \sum_{i=1}^N e^i(x) B^i_t$  where  $B^i_0 = 0$  and we consider the SPDE (6.1) in the sense of distributions, i.e.

$$\int_{\mathbb{R}} \phi(x) u(t, x) dx = \int_{\mathbb{R}} \phi(x) \mathbf{u}_0(dx) + \frac{1}{2} \int_0^t \int_{\mathbb{R}} \phi(x) \sigma^2(u(s, x)) ds dx + \int_0^t \int_{\mathbb{R}} \phi(x) u(s, x) \xi(ds, x) dx, \tag{6.4}$$

where the latter stochastic integral is intended in the Itô sense.

#### 3.6.2 Well-posedness of the SPDE

The theorem below contains results taken from [11, 98].

**Theorem 3.6.1.** *Suppose that  $\beta$  is Lipschitz.*

- *Suppose that  $u_0 \in L^2(\mathbb{R})$ . There is a solution to equation (6.1).*
- *Assume further that  $\beta$  is non-degenerate, i.e.  $\beta(r) \geq ar^2$ ,  $r \in \mathbb{R}$ , where  $a > 0$ . Then, there is a solution  $u$  to (6.1) for any probability  $\mathbf{u}_0(dx)$  (even in  $H^{-1}(\mathbb{R})$ ).*

- There is at most one solution in the class of random fields  $u$  such that  $\int_{[0,T] \times \mathbb{R}} u^2(t, x) dt dx < \infty$  a.s.

**Remark 3.6.2.** • Previous result extends to the case of an infinite number of modes  $e^i$  and for  $d \geq 1$ .

- We remark that the  $\partial_t \xi(t, x)$  is a colored noise (in space). The case of space-time white noise seems very difficult to treat.

### 3.6.3 McKean equation in random environment

Given a local martingale  $M$ ,  $E(M)$  denotes the Doléans exponential of  $M$  i.e.  $\exp(M_t - \frac{1}{2} \int_0^t M^2 ds)$ ,  $t \geq 0$ . We say that a filtered probability space  $(\Omega, \mathbf{G}, (\mathbf{G}_t), Q)$  is a **suitable enlarged space** of  $(\Omega, \mathbf{F}, (\mathbf{F}_t), P)$ , if the following holds.

1. There is a measurable space  $(\Omega_1, \mathbf{H})$  with  $\Omega_0 = \Omega \times \Omega_1$ ,  $\mathbf{G} = \mathbf{F} \otimes \mathbf{H}$  and a random kernel  $(\omega, H) \mapsto Q^\omega(H)$  defined on  $\Omega \times H \rightarrow [0, 1]$  such that the probability  $Q$  on  $(\Omega_0, \mathbf{G})$  is defined by  $dQ(\omega, \omega_1) = dP(\omega)Q^\omega(\omega_1)$ .
2. The processes  $B^1, \dots, B^N$  are  $(\mathbf{G}_t)$ -Brownian motions where  $\mathbf{G}_t = \mathbf{F}_t \vee \mathbf{H}$ .

**Definition 3.6.3.** We say that the non-linear doubly-stochastic diffusion

$$\begin{aligned} \square \quad Y_t &= Y_0 + \int_0^t \Phi(u(s, Y_s)) dW_s, \\ \int \phi(x) u(t, x) dx &= E^{Q^\omega} \left[ \phi(Y_t(\omega, \cdot)) E_t \int_0^t \xi(ds, Y_s)(\omega, \cdot) \right], \\ \square \quad \xi - \text{Law}(Y_0) &= \mathbf{u}_0(dx), \end{aligned} \quad (6.5)$$

admits **weak existence** on  $(\Omega, \mathbf{F}, (\mathbf{F}_t), P)$  if there is a suitably enlarged probability space  $(\Omega_0, \mathbf{G}, (\mathbf{G}_t), Q)$  an  $(\mathbf{G}_t)$ -Brownian motion  $W$  such that (6.5) is verified. The couple  $(Y, u)$  will be called **weak solution** of (6.5).

**Remark 3.6.4.** • We remark that the second line in (6.5) represents a sort of  $\xi$ -marginal weighted law of  $Y_t$ .

- Let  $(Y, u)$  be a solution to (6.5). Then  $u$  is a solution to (6.1).
- Such representation allows to show that  $u(t, x) \geq 0$ ,  $dP dt dx$  a.e. and, at least if  $e^0 = 0$ ,  $E \int_{\mathbb{R}} u(t, x) dx = 1$ , so that the conservativity is maintained at the expectation level.

**Definition 3.6.5.** Let two measurable random fields  $u^i : \Omega \times [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ ,  $i = 1, 2$  on  $(\Omega, \mathbf{F}, P, (\mathbf{F}_t))$ , and  $Y^i$ , on a suitable extended probability space  $(\Omega^i, \mathbf{G}^i, (\mathbf{G}_t^i), Q^i)$ ,  $i = 1, 2$ , such that  $(Y^i, u^i)$  are (weak) solutions of (6.5) on  $(\Omega, \mathbf{F}, (\mathbf{F}_t), P)$ . If we always have that  $(Y^1, B^1, \dots, B^N)$  and  $(Y^2, B^1, \dots, B^N)$  have the same law, then we say that (6.5) admits **weak uniqueness** (on  $(\Omega, \mathbf{F}, (\mathbf{F}_t), P)$ ).

**Theorem 3.6.6.** Under the assumption of Theorem 3.6.1 equation (6.5) admits (weak) existence and uniqueness on  $(\Omega, \mathbf{F}, (\mathbf{F}_t), P)$ .

### 3.7 McKean representation of stochastic control problems

#### 3.7.1 Stochastic control problems and non-linear Partial Differential Equations

There are several connections between stochastic control and McKean type SDEs, see e.g. [1]. Here, we propose an original (and maybe unexpected) point of view. Let us briefly recall the link between stochastic control and non-linear PDEs given by the Hamilton-Jacobi-Bellman (HJB) equation. We refer for instance to [108, 94, 42] for more details. Consider a state process  $(X_s^{t_0, x, \alpha})_{t_0 \leq s \leq T}$  on  $[t_0, T] \times \mathbb{R}^d$  solution to the controlled SDE

$$\begin{aligned} dX_s^{t_0, x, \alpha} &= b(s, X_s^{t_0, x, \alpha}, \alpha(s, X_s^{t_0, x, \alpha})) ds + \sigma(s, X_s^{t_0, x, \alpha}, \alpha(s, X_s^{t_0, x, \alpha})) dW_s \\ X_{t_0}^{t_0, x, \alpha} &= x, \end{aligned} \tag{7.1}$$

where  $W$  denotes the Brownian motion on  $[t_0, T] \times \mathbb{R}^d$ , and  $\alpha(s, X_s^{t_0, x, \alpha})$  represents Markovian control in the sense that the control at time  $t$  is supposed here to depend on  $t$  and on the current value of the state process:

$$\alpha \in \mathbf{A}_{t_0, T} := \left\{ \alpha : (t, x) \in [t_0, T] \times \mathbb{R}^d \mapsto \alpha(t, x) \in A \subset \mathbb{R}^k \right\}, \tag{7.2}$$

$A$  being a subset of  $\mathbb{R}^k$ . For a given initial time and state  $(t_0, x) \in [0, T] \times \mathbb{R}^d$ , we are interested in maximizing, over the Markovian controls  $\alpha \in \mathbf{A}_{t_0, T}$ , the criteria

$$J(t_0, x, \alpha) := \mathbf{E} \left[ g(X_T^{t_0, x, \alpha}) + \int_{t_0}^T f(s, X_s^{t_0, x, \alpha}, \alpha(s, X_s^{t_0, x, \alpha})) ds \right]. \tag{7.3}$$

In the above criteria, the function  $f$  is called the *running gain* whereas  $g$  is called the *terminal gain*.

**Remark 3.7.1.** At first glance, the set of control processes of the form  $\alpha_t = \alpha(t, X_t)$  defined in (7.2) may appear too restrictive compared to a larger set of non-anticipative controls  $(\alpha_t)$  which may depend on all the past history of the state process  $(X_t)$ . However, in the framework of Markov control problems (for which the state process  $(X_t^{t_0, x, \alpha})$  is Markov, as soon as the control is fixed to a deterministic value  $\alpha_t = a \in A$ , for all  $t \in [t_0, T]$ ), it is well-known that the optimal control process  $(\alpha_t)$  lies in the set of Markovian controls verifying  $\alpha_t = \alpha(t, X_t)$ . Hence, considering controls of the particular form (7.2) is done here without loss of generality.

To tackle this finite horizon stochastic control problem, the usual approach consists in introducing the associated value (or Bellman) function  $v : [t_0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  representing the maximum gain one can expect, starting from time  $t$  at state  $x$ , i.e.

$$v(t, x) := \sup_{\alpha \in \mathbf{A}_{t, T}} J(t, x, \alpha), \quad \text{for } t \in [t_0, T]. \tag{7.4}$$

Note that the terminal condition is known, which fixes  $v(T, x) = g(x)$ , whereas the initial condition  $v(t_0, x)$  corresponds to the solution of the original minimization problem. The value function is then proved to verify the *Dynamic Programming Principle (DPP)* which consists in the backward induction

$$v(t, x) = \sup_{\alpha \in \mathbf{A}_{t, \tau}} \mathbf{E} \left[ \int_t^\tau f(s, X_s^{t, x, \alpha}, \alpha(s, X_s^{t, x, \alpha})) ds + v(\tau, X_\tau^{t, x, \alpha}) \right], \quad \text{for any stopping time } \tau \in ]t, T]. \tag{7.5}$$

Under continuity assumptions on  $b, \sigma, f, g$ , using DPP together with Itô formula allows to characterize  $v$  as a viscosity solution of the HJB equation

$$\begin{aligned} v(T, x) &= g(x) \\ \partial_t v(t, x) + H(t, x, \nabla v(t, x), \nabla^2 v(t, x)) &= 0, \end{aligned} \quad (7.6)$$

where  $\nabla$  and  $\nabla^2$  denote the gradient and the Hessian operators and the so-called, *Hamiltonian*,  $H$  denotes the real valued function defined on  $[0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{S}^d$  ( $\mathbb{S}^d$  denoting the set of symmetric matrices in  $\mathbb{R}^{d \times d}$ ), such that

$$H(t, x, \delta, \gamma) := \sup_{a \in A} f(t, x, a) + b(t, x, a)^\top \delta(t, x) + \frac{1}{2} \text{Tr}[\sigma \sigma^\top(t, x, a) \gamma(t, x)] \quad . \quad (7.7)$$

Note that (7.6) is a non-linear PDE because of the nonlinearity in the Hamiltonian induced by the supremum operator. Besides, assuming that, for all  $(t, x) \in [t_0, T] \times \mathbb{R}^d$ , the supremum in (7.7) is attained at a unique maximizer, then the optimal control  $\alpha^*$  is directly obtained as a function of the Bellman function and its derivatives, i.e.

$$\alpha^*(t, x) = \arg \max_{a \in A} f(t, x, a) + b(t, x, a)^\top \nabla v(t, x) + \frac{1}{2} \text{Tr}[\sigma \sigma^\top(t, x, a) \nabla^2 v(t, x)] \quad . \quad (7.8)$$

Except in some very concrete cases such as the Linear Quadratic Gaussian (LQG) setting (where the states dynamics involve an affine drift with Gaussian noise and the cost is quadratic both w.r.t. the control and the state), there is no explicit solution to stochastic control problems. To numerically approximate the solution of equation (7.6), several approaches have been proposed, mainly differing in the way the value function  $v$  is interpreted. Indeed, as pointed out,  $v$  can be viewed either as the solution to the control problem (7.4), or as a (viscosity) solution of the non-linear PDE (7.6).

1. When  $v$  is defined as the solution to the control problem (7.4), a natural approach consists in discretizing the time continuous control problem and apply the time discrete Dynamic Programming Principle [21]. Then the problem consists in maximizing over the controls, backwardly in time, the conditional expectation of the value function related to (7.5). The maximization at time step  $t_k$  can be done via a parametrization of the control  $x \mapsto \alpha_{t_k}^\theta(x)$  via a parameter  $\theta$  so that parametric optimization methods such as the stochastic gradient algorithm could be applied to maximize the expectation over  $\theta$ . It remains to approximate the conditional expectations by numerical methods such as PDE, Fourier, Monte Carlo, Quantization or lattice methods. . . A great variety of numerical approximation schemes has been developed in the specific Bermudan option valuation test-bed [27, 81, 8, 109, 35, 26]. Alternatively, one can use Markov chain approximation method [72] which consists in a time-space discretization designed to obtain a proper Markov chain.
2. In the second approach we recall that  $v$  is viewed as the solution of (7.6). The problem amounts then to discretize a non-linear PDE. Then one can rely on numerical analysis methods (e.g. finite differences, or finite elements) and use monotone approximation schemes in the sense



of Barles and Souganidis [13] to build converging approximation schemes, e.g. [23, 46]. This type of approach is in general limited to state space dimension lower than 4. To tackle higher dimensional problems, one approach consists in converting the PDE into a probabilistic setting in order to apply Monte Carlo types algorithms. To that end, various kinds of probabilistic representations of non-linear PDEs are available. Forward Backward Stochastic Differential Equations (FBSDE) were introduced in [92] as probabilistic representations of semi-linear PDEs. Then various types of numerical schemes for FBSDE have been developed. They mainly differ in the approach of evaluating conditional expectations: [25] (resp. [50], [38, 90]) use kernel (resp. regression, quantization) methods. Recently, important progresses have been done performing machine learning techniques, see e.g. [40, 57]. Branching processes [85, 56] can also provide probabilistic representations of semi-linear PDEs via Feynman-Kac formula. Non-linear SDEs in the sense of McKean [84] are another approach that constitutes the subject of the present paper.

3. Other approaches take advantage of both interpretations see for instance [43] and in [106].

### 3.7.2 McKean type representation in a toy control problem example

In order to illustrate the application of MFKEs to control problems, we consider a simple example corresponding to an inventory problem, for which the Hamiltonian maximization (7.7),(7.8) is explicit. The state  $(X_t)_{t \in [t_0, T]}$  denotes the stock level evolving randomly with a control of the drift  $\alpha$ :

$$\begin{aligned} \square \quad dX_t^{t_0, x, \alpha} &= -\alpha(t, X_t^{t_0, x, \alpha})dt + \sigma dW_t \\ \square \quad J(t_0, x, \alpha) &= \sup_{\alpha \in A_{t_0, T}} \mathbf{E} \int_{t_0}^T g(X_t^{t_0, x, \alpha}) - \int_{t_0}^T \left[ \alpha(t, X_t^{t_0, x, \alpha}) - D_t^2 + h(X_t^{t_0, x, \alpha}) \right] dt . \end{aligned}$$

Bound constraints on the storage level are implicitly forced by the penalization  $h$ . A target terminal level is indicated by the terminal gain  $g$ , supposed here to be Lebesgue integrable. The objective is then to follow a deterministic target profile  $(D_t)_{t \in [0, T]}$ , on a given finite horizon  $[t_0, T]$ . When the admissible set in which the controls take their values  $A = \mathbf{R}$ , one can explicitly derive the optimal control as a function of the value function derivative

$$\alpha^*(t, \cdot) = D_t + \frac{1}{2}(\partial_x v)(t, \cdot),$$

which yields the following HJB equation

$$\partial_t v + \frac{1}{4}(\partial_x v)^2 + D_t \partial_x v + \frac{\sigma^2}{2} \partial_{xx} v - h = 0 .$$

Reversing the time, (with  $t_0 = 0$ ) gives  $(t, x) \mapsto u(t, x) := v(T - t, x)$  solution of

$$\begin{aligned} \partial_t u &= \frac{1}{4}(\partial_x u)^2 + \frac{\sigma^2}{2} \partial_{xx} u + D_t \partial_x u - h, \\ u(0, x) &= g(x). \end{aligned} \tag{7.9}$$

We recover the framework of (1.1), with  $\Lambda(t, x, y, z) = \frac{1}{4} \frac{|z|^2}{y} - \frac{h(x)}{y}$  and  $b(t, x, y) = -D_t$ . Consequently

the Bellman function  $v$  can be represented via

$$\begin{aligned}
 & \square Y_t = Y_0 + \sigma W_t - \int_0^t D_s ds \\
 & \square Y_0 \sim \frac{\int g(x) dx}{\int g(y) dy} \\
 & \int \phi(x) v(t, x) dx = \int_{\mathbb{R}} g(y) dy \mathbf{E} \left\{ \phi(Y_{T-t}) \exp \int_t^T \Lambda(s, Y_{T-s}, v(s, Y_{T-s}), \nabla v(s, Y_{T-s})) ds \right\}, \\
 & \square \text{ for } t \in [0, T].
 \end{aligned} \tag{7.10}$$



# Bibliography

- [1] S. Albeverio, F.C. De Vecchi, A. Romano, and S. Ugolini. Mean-field limit for a class of stochastic ergodic control problems. *ArXiv:2003.06469*, 2020.
- [2] S. Albeverio and S. Ugolini. A Doob h-transform of the Gross-Pitaevskii Hamiltonian. *J. Stat. Phys.*, 161(2):486–508, 2015.
- [3] D. G. Aronson. Non-negative solutions of linear parabolic equations. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (3)*, 22:607–694, 1968.
- [4] D. G. Aronson and H. F. Weinberger. Multidimensional nonlinear diffusion arising in population genetics. *Adv. Math.*, 30:33–76, 1978.
- [5] A. C. Bagtzoglou and J. Atmadja. Marching-jury backward beam equation and quasi-reversibility methods for hydrologic inversion: Application to contaminant plume spatial distribution recovery. *Water Resources Research*, 39(2), 2003.
- [6] P. Bak. *How nature works*. Copernicus, New York, 1996. The science of self-organized criticality.
- [7] P. Bak, C. Tang, and K. Wiesenfeld. Self-organized criticality. *Phys. Rev. A (3)*, 38(1):364–374, 1988.
- [8] V. Bally, G. Pagès, and J. Printems. A quantization tree method for pricing and hedging multi-dimensional American options. *Math. Finance*, 15(1):119–168, 2005.
- [9] V. Barbu and M. Röckner. Probabilistic representation for solutions to nonlinear Fokker-Planck equations. *SIAM J. Math. Anal.*, 50(4):4246–4260, 2018.
- [10] V. Barbu, M. Röckner, and F. Russo. Probabilistic representation for solutions of an irregular porous media type equation: the irregular degenerate case. *Probab. Theory Related Fields*, 151(1-2):1–43, 2011.
- [11] V. Barbu, M. Röckner, and F. Russo. Stochastic porous media equations in  $\mathbb{R}^d$ . *J. Math. Pures Appl. (9)*, 103(4):1024–1052, 2015.
- [12] V. Barbu, M. Röckner, and F. Russo. Doubly probabilistic representation for the stochastic porous media type equation. *Ann. Inst. Henri Poincaré Probab. Stat.*, 53(4):2043–2073, 2017.

- [13] G. Barles and P. E. Souganidis. Convergence of approximation schemes for fully nonlinear second order equations. *Asymptotic Anal.*, 4(3):271–283, 1991.
- [14] J. V. Beck, B. Blackwell, and Ch. R. St. Clair Jr. *Inverse heat conduction: Ill-posed problems*. James Beck, 1985.
- [15] N. Belaribi, F. Cuvelier, and F. Russo. Probabilistic and deterministic algorithms for space multidimensional irregular porous media equation. *SPDEs: Analysis and Computations*, 1(1):3–62, 2013.
- [16] N. Belaribi and F. Russo. Uniqueness for Fokker-Planck equations with measurable coefficients and applications to the fast diffusion equation. *Electron. J. Probab.*, 17:no. 84, 28, 2012.
- [17] S. Benachour, P. Chassaing, B. Roynette, and P. Vallois. Processus associés à l'équation des milieux poreux. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 23(4):793–832, 1996.
- [18] Ch. Bender and R. Denk. A forward scheme for backward SDEs. *Stochastic Processes Appl.*, 117(12):1793–1812, 2007.
- [19] Ch. Bender and T. Moseler. Importance sampling for backward SDEs. *Stochastic Anal. Appl.*, 28(2):226–253, 2010.
- [20] Ch. Bender and J. Steiner. Least-squares Monte Carlo for backward SDEs. In *Numerical methods in finance. Selected papers based on the presentations at the workshop, Bordeaux, France, June 2010*, pages 257–289. Berlin: Springer, 2012.
- [21] D. P. Bertsekas and S. E. Shreve. *Stochastic optimal control*, volume 139 of *Mathematics in Science and Engineering*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1978. The discrete time case.
- [22] P. Blanchard, M. Röckner, and F. Russo. Probabilistic representation for solutions of an irregular porous media type equation. *Ann. Probab.*, 38(5):1870–1900, 2010.
- [23] J.F. Bonnans and H. Zidani. Consistency of generalized finite difference schemes for the stochastic HJB equation. *SIAM J. Numer. Anal.*, 41(3):1008–1021, 2003.
- [24] M. Bossy and D. Talay. Convergence rate for the approximation of the limit law of weakly interacting particles: application to the Burgers equation. *Ann. Appl. Probab.*, 6(3):818–861, 1996.
- [25] B. Bouchard and N. Touzi. Discrete-time approximation and Monte Carlo simulation of backward stochastic differential equations. *Stochastic Process. Appl.*, 111:175–206, 2004.
- [26] B. Bouchard and X. Warin. Monte-Carlo valorisation of American options: facts and new algorithms to improve existing methods. In *Numerical Methods in finance*. Springer, 2012.

- [27] M. Broadie, P. Glasserman, et al. A stochastic mesh method for pricing high-dimensional American options. *Journal of Computational Finance*, 7:35–72, 2004.
- [28] R. Bronson and G. B. Costa. *Matrix Methods: Applied Linear Algebra*. Academic Press, 2008.
- [29] R. Cafiero, V. Loreto, L. Pietronero, A. Vespignani, and S. Zapperi. Local rigidity and self-organized criticality for avalanches. *Europhysics Letters*, 29:111–116, 1995.
- [30] D. S. Callaway and I. A. Hiskens. Achieving controllability of electric loads. *Proceedings of the IEEE*, 99(1):184–199, 2010.
- [31] P. Cattiaux, G. Conforti, Gentil I., and Ch. Léonard. Time reversal of diffusion processes under a finite entropy condition, Preprint Arxiv, 2021.
- [32] G. Conforti and Ch. Léonard. Time reversal of Markov processes with jumps under a finite entropy condition, Preprint Arxiv, 2021.
- [33] M. G. Crandall, H. Ishii, and P.L. Lions. User’s guide to viscosity solutions of second order partial differential equations. *Bulletin of the American Mathematical Society*, 27(1):1–67, 1992.
- [34] P. Del Moral. *Feynman-Kac formulae*. Probability and its Applications (New York). Springer-Verlag, New York, 2004. Genealogical and interacting particle systems with applications.
- [35] P. Del Moral, P. Hu, N. Oudjane, and B. Rémillard. On the robustness of the Snell envelope. *SIAM J. Financial Math.*, 2(1):587–626, 2011.
- [36] P. Del Moral, P. E. Jacob, A. Lee, L. Murray, and G. W. Peters. Feynman-Kac particle integration with geometric interacting jumps. *Stoch. Anal. Appl.*, 31(5):830–871, 2013.
- [37] P. Del Moral and L. Miclo. Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering. In *Séminaire de Probabilités, XXXIV*, volume 1729 of *Lecture Notes in Math.*, pages 1–145. Springer, Berlin, 2000.
- [38] F. Delarue and S. Menozzi. An interpolated stochastic algorithm for quasi-linear PDEs. *Math. Comp.*, 77(261):125–158 (electronic), 2008.
- [39] C. Di Girolami and F. Russo. About classical solutions of the path-dependent heat equation. *Random Oper. Stoch. Equ.*, 28(1):35–62, 2020.
- [40] W. E, M. Hutzenthaler, A. Jentzen, and Th. Kruse. On multilevel Picard numerical approximations for high-dimensional nonlinear parabolic partial differential equations and high-dimensional nonlinear backward stochastic differential equations. *J. Sci. Comput.*, 79(3):1534–1571, 2019.
- [41] I. Exarchos and E. A. Theodorou. Stochastic optimal control via forward and backward stochastic differential equations and importance sampling. *Automatica*, 87:159–165, 2018.

- [42] G. Fabbri, F. Gozzi, and A. Świech. *Stochastic optimal control in infinite dimension*, volume 82 of *Probability Theory and Stochastic Modelling*. Springer, Cham, 2017. Dynamic programming and HJB equations, With a contribution by Marco Fuhrman and Gianmario Tessitore.
- [43] A. Fahim, N. Touzi, and X. Warin. A probabilistic numerical method for fully nonlinear parabolic PDEs. *Ann. Appl. Probab.*, 21(4):1322–1364, 2011.
- [44] A. Figalli. Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients. *J. Funct. Anal.*, 254(1):109–153, 2008.
- [45] H. Föllmer and A. Wakolbinger. Time reversal of infinite-dimensional diffusions. *Stochastic Process. Appl.*, 22(1):59–77, 1986.
- [46] P.A. Forsyth and K. R. Vetzal. Numerical methods for nonlinear pdes in finance. In *Handbook of Computational Finance*, pages 503–528. Springer, 2012.
- [47] A. Friedman. *Partial differential equations of parabolic type*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1964.
- [48] E. Gobet and C. Labart. Error expansion for the discretization of backward stochastic differential equations. *Stochastic Process. Appl.*, 117(7):803–829, 2007.
- [49] E. Gobet and C. Labart. Solving BSDE with adaptive control variate. *SIAM J. Numer. Anal.*, 48(1):257–277, 2010.
- [50] E. Gobet, J-P. Lemor, and X. Warin. A regression-based Monte Carlo method to solve backward stochastic differential equations. *Ann. Appl. Probab.*, 15(3):2172–2202, 2005.
- [51] E. Gobet and P. Turkedjiev. Linear regression MDP scheme for discrete backward stochastic differential equations under general conditions. *Math. Comput.*, 85(299):1359–1391, 2016.
- [52] E. Gobet and P. Turkedjiev. Adaptive importance sampling in least-squares Monte Carlo algorithms for backward stochastic differential equations. *Stochastic Processes Appl.*, 127(4):1171–1203, 2017.
- [53] F. Gozzi and F. Russo. Verification theorems for stochastic optimal control problems via a time dependent Fukushima-Dirichlet decomposition. *Stochastic Process. Appl.*, 116(11):1530–1562, 2006.
- [54] F. Gozzi and F. Russo. Weak Dirichlet processes with a stochastic control perspective. *Stochastic Process. Appl.*, 116(11):1563–1583, 2006.
- [55] U. G. Haussmann and É. Pardoux. Time reversal of diffusions. *Ann. Probab.*, 14(4):1188–1205, 1986.

- [56] P. Henry-Labordère, N. Oudjane, X. Tan, N. Touzi, and X. Warin. Branching diffusion representation of semilinear pdes and Monte Carlo approximations. *Available at <http://arxiv.org/pdf/1603.01727v1.pdf>*, 2016.
- [57] C. Huré, H. Pham, and X. Warin. Deep backward schemes for high-dimensional nonlinear PDEs. *Math. Comp.*, 89(324):1547–1579, 2020.
- [58] H. Ishii. On uniqueness and existence of viscosity solutions of fully nonlinear second-order elliptic pde's. *Communications on Pure and Applied Mathematics*, 42(1):15–45, 1989.
- [59] H. Ishii and K. Kobayasi. On the uniqueness and existence of solutions of fully nonlinear parabolic pdes under the osgood type condition. *Differential Integral Equations*, 7(3-4):909–920, 1994.
- [60] L. Izydorczyk, N. Oudjane, and F. Russo. A fully backward representation of semilinear pdes applied to the control of thermostatic loads in power systems. *Preprint hal-03210302*, 2021.
- [61] L. Izydorczyk, N. Oudjane, and F. Russo. McKean Feynman-Kac probabilistic representations of nonlinear partial differential equations. *Geometry and Invariance in Stochastic Dynamics*. Eds. S. Ugolini et al., to appear.
- [62] L. Izydorczyk, N. Oudjane, F. Russo, and G. Tessitore. Fokker-Planck equations with terminal condition and related McKean probabilistic representation. *Preprint hal-02902615*, 2020.
- [63] J-F Jabir, D. Talay, and M. Tomašević. Mean-field limit of a particle approximation of the one-dimensional parabolic-parabolic Keller-Segel model without smoothing. *Electron. Commun. Probab.*, 23:Paper No. 84, 14, 2018.
- [64] J. Jacod and Ph. Protter. Time reversal on Lévy processes. *Ann. Probab.*, 16(2):620–641, 1988.
- [65] R. Jensen, P.-L. Lions, and P. E. Souganidis. A uniqueness result for viscosity solutions of second order fully nonlinear partial differential equations. *Proceedings of the American Mathematical Society*, page 975?975, 1988.
- [66] B. Jourdain and S. Méléard. Propagation of chaos and fluctuations for a moderate model with smooth initial data. *Ann. Inst. H. Poincaré Probab. Statist.*, 34(6):727–766, 1998.
- [67] B. Jourdain, S. Méléard, and W. A. Woyczynski. Nonlinear SDEs driven by Lévy processes and related PDEs. *ALEA Lat. Am. J. Probab. Math. Stat.*, 4:1–29, 2008.
- [68] M. Kac. *Probability and related topics in physical sciences*, volume 1957 of *With special lectures by G. E. Uhlenbeck, A. R. Hibbs, and B. van der Pol. Lectures in Applied Mathematics. Proceedings of the Summer Seminar, Boulder, Colo.* Interscience Publishers, London-New York, 1959.



- [69] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.
- [70] J. P. Keener and J. Sneyd. *Mathematical Physiology II: Systems Physiology*. Springer, New York, 2008.
- [71] N. V. Krylov. *Controlled diffusion processes*, volume 14 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2009. Translated from the 1977 Russian original by A. B. Aries, Reprint of the 1980 edition.
- [72] H. J. Kushner and P. G. Dupuis. *Numerical methods for stochastic control problems in continuous time*, volume 24 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1992.
- [73] C. Labart and J. Lelong. A parallel algorithm for solving BSDEs. *Monte Carlo Methods Appl.*, 19(1):11–39, 2013.
- [74] R. Lattès and J.-L. Lions. *The method of quasi-reversibility. Applications to partial differential equations*. Translated from the French edition and edited by Richard Bellman. Modern Analytic and Computational Methods in Science and Mathematics, No. 18. American Elsevier Publishing Co., Inc., New York, 1969.
- [75] A. Le Cavil, N. Oudjane, and F. Russo. Particle system algorithm and chaos propagation related to a non-conservative McKean type stochastic differential equations. *Stochastics and Partial Differential Equations: Analysis and Computation.*, pages 1–37, 2016.
- [76] A. Le Cavil, N. Oudjane, and F. Russo. Probabilistic representation of a class of non-conservative nonlinear partial differential equations. *ALEA Lat. Am. J. Probab. Math. Stat*, 13(2):1189–1233, 2016.
- [77] A. Le Cavil, N. Oudjane, and F. Russo. Monte-Carlo algorithms for a forward Feynman–Kac-type representation for semilinear nonconservative partial differential equations. *Monte Carlo Methods Appl.*, 24(1):55–70, 2018.
- [78] A. Le Cavil, N. Oudjane, and F. Russo. Forward Feynman-Kac type representation for semilinear nonconservative partial differential equations. *Stochastics: an International Journal of Probability and Stochastic Processes*, to appear. First version 2016, Preprint hal-01353757.
- [79] J. Lieber, N. Oudjane, and F. Russo. On the well-posedness of a class of McKean Feynman-Kac equations. *Markov Process. Relat. Fields*, 25(5):821–862, 2019.
- [80] P. L. Lions. Optimal control of diffusion processes and hamilton jacobi bellman equations part 2 : viscosity solutions and uniqueness. *Communications in Partial Differential Equations*, 8(11):1229–1276, 1983.

- [81] F. A. Longstaff and E. S. Schwartz. Valuing American options by simulation: a simple least-squares approach. *Review of Fin. Stud.*, 14:113–147, 2001.
- [82] A. Lunardi. *Analytic semigroups and optimal regularity in parabolic problems*, volume 16 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser Verlag, Basel, 1995.
- [83] H. P. Jr. McKean. A class of Markov processes associated with nonlinear parabolic equations. In *Proc. Nat. Acad. Sci. U.S.A.*, 1966), pages 1907–1911. 1966.
- [84] H. P. Jr. McKean. Propagation of chaos for a class of non-linear parabolic equations. In *Stochastic Differential Equations (Lecture Series in Differential Equations, Session 7, Catholic Univ., 1967)*, pages 41–57. Air Force Office Sci. Res., Arlington, Va., 1967.
- [85] H. P. Jr. McKean. Application of Brownian motion to the equation of Kolmogorov-Petrovskii-Piskunov. *Comm. Pure Appl. Math.*, 28(3):323–331, 1975.
- [86] S. Méléard and S. Roelly-Coppoletta. A propagation of chaos result for a system of particles with moderate interaction. *Stochastic Process. Appl.*, 26(2):317–332, 1987.
- [87] J. D. Murray. *Mathematical biology. I*, volume 17 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, New York, third edition, 2002. An introduction.
- [88] D. Nunziante. Existence and uniqueness of unbounded viscosity solutions of parabolic equations with discontinuous time-dependence. *Nonlinear Analysis: Theory, Methods & Applications*, 18(11):1033 – 1062, 1992.
- [89] K. Oelschläger. A martingale approach to the law of large numbers for weakly interacting stochastic processes. *Ann. Probab.*, 12(2):458–479, 1984.
- [90] G. Pagès and A. Sagna. Improved error bounds for quantization based numerical schemes for BSDE and nonlinear filtering. *Stochastic Process. Appl.*, 128(3):847–883, 2018.
- [91] E. Pardoux. Backward stochastic differential equations and viscosity solutions of systems of semilinear parabolic and elliptic PDEs of second order. In *Stochastic analysis and related topics, VI (Geilo, 1996)*, volume 42 of *Progr. Probab.*, pages 79–127. Birkhäuser Boston, Boston, MA, 1998.
- [92] É. Pardoux and S. Peng. Backward stochastic differential equations and quasilinear parabolic partial differential equations. In *Stochastic partial differential equations and their applications (Charlotte, NC, 1991)*, volume 176 of *Lect. Notes Control Inf. Sci.*, pages 200–217. Springer, Berlin, 1992.
- [93] É. Pardoux, F. Pradeilles, and Z. Rao. Probabilistic interpretation of a system of semi-linear parabolic partial differential equations. *Annales de l'I.H.P. Probabilités et statistiques*, 33(4):467–490, 1997.

- [94] H. Pham. *Continuous-time stochastic control and optimization with financial applications*, volume 61 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2009.
- [95] M. Renardy, W. J. Hrusa, and J. A. Nohel. *Mathematical problems in viscoelasticity*, volume 35 of *Pitman Monographs and Surveys in Pure and Applied Mathematics*. Longman Scientific & Technical, Harlow; John Wiley & Sons, Inc., New York, 1987.
- [96] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.
- [97] C. Ribeiro and N. Webber. Valuing path-dependent options in the variance-gamma model by Monte Carlo with a gamma bridge. *Journal of Computational Finance*, 7(2):81–100, 2004.
- [98] M. Röckner and F. Russo. Uniqueness for a class of stochastic Fokker-Planck and porous media equations. *J. Evol. Equ.*, 17(3):1049–1062, 2017.
- [99] L. C. G. Rogers and D. Williams. *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Itô calculus, Reprint of the second (1994) edition.
- [100] W. Rudin. *Functional analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, second edition, 1991.
- [101] P. Sabino. Forward or backward simulation? A comparative study. *Quantitative Finance*, 20(7):1213–1226, 2020.
- [102] A. Seguret, C. Alasseur, J. F. Bonnans, A. De Paola, N. Oudjane, and V. Trovato. Decomposition of high dimensional aggregative stochastic control problems. *Preprint arXiv:2008.09827*, 2020.
- [103] B. W. Silverman. *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1986.
- [104] D. W. Stroock and S. R. S. Varadhan. *Multidimensional diffusion processes*. Classics in Mathematics. Springer-Verlag, Berlin, 2006. Reprint of the 1997 edition.
- [105] A-S. Sznitman. Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*, volume 1464 of *Lecture Notes in Math.*, pages 165–251. Springer, Berlin, 1991.
- [106] X. Tan. Probabilistic numerical approximation for stochastic control problems. *preprint*, 2012.
- [107] A. N. Tikhonov and V. Y. Arsenin. *Solutions of ill-posed problems*. V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York-Toronto, Ont.-London, 1977. Translated from the Russian, Preface by translation editor Fritz John, Scripta Series in Mathematics.

- [108] N. Touzi. *Optimal stochastic control, stochastic target problems, and backward SDE*, volume 29 of *Fields Institute Monographs*. Springer, New York; Fields Institute for Research in Mathematical Sciences, Toronto, ON, 2013. With Chapter 13 by Agnès Tourin.
- [109] J. N. Tsitsiklis and B. Van Roy. Regression methods for pricing complex American-style options. *IEEE Transactions on Neural Networks*, 12(4):694–703, 2001.
- [110] A. Yu. Veretennikov. Parabolic equations and Itô's stochastic equations with coefficients discontinuous in the time variable. *Mathematical Notes of the Academy of Sciences of the USSR*, 31(4):278–283, April 1982.
- [111] X. Y. Wang, Z. S. Zhu, and Y. K. Lu. Solitary wave solutions of the generalized Burgers-Huxley equation. *J. Phys. A: Math. Gen.*, 23:271–274, 1990.
- [112] A. K. Zvonkin. A transformation of the phase space of a diffusion process that will remove the drift. *Mat. Sb. (N.S.)*, 93(135):129–149, 152, 1974.





# Introduction

Une des visées de l'analyse stochastique est d'établir des connexions entre problèmes déterministes d'évolution potentiellement non linéaires et processus stochastiques. L'avantage que l'on peut tirer de ces connexions réside dans la possibilité de montrer qu'une équation aux dérivées partielles (EDP) est bien posée ou de discuter la régularité de ses solutions, en utilisant des méthodes probabilistes. De plus, une telle correspondance peut apporter un éclairage nouveau sur l'approximation numérique des solutions, en particulier lorsque ces équations sont posées sur un espace de grande dimension.

Dans le cas particulier des problèmes de Cauchy paraboliques du second ordre posés sur  $[0, T] \times \mathbb{R}^d$ , étant donné un horizon de temps  $T > 0$  et une dimension spatiale  $d$ , les objets probabilistes qui émergent sont solutions d'équations différentielles stochastiques markoviennes (EDS). Étant données deux fonctions  $b, \sigma$  définies sur  $[0, T] \times \mathbb{R}^d$  à valeurs respectivement  $\mathbb{R}^d$  et matrices, nous nous concentrons sur deux types d'EDP: L'équation de *Kolmogorov progressive* (ou équation de *Fokker-Planck*)

$$\begin{aligned} \square & \partial_t \mathbf{u}(t) = \frac{1}{2} \sum_{i,j=1}^d \partial_{ij} \sigma \sigma^T(t, \cdot)_{ij} \mathbf{u}(t) - \sum_{i=1}^d \partial_i (b(t, \cdot)_i \mathbf{u}(t)), \quad t \in ]0, T[ \\ \square & \mathbf{u}(0) = \nu, \end{aligned} \tag{1.2}$$

où  $\nu$  est une mesure donnée et l'équation de *Kolmogorov rétrograde*

$$\begin{aligned} \square & \partial_t v(t, \cdot) + \frac{1}{2} \sum_{i,j=1}^d \sigma \sigma^T(t, \cdot)_{ij} \partial_{ij} v(t, \cdot) + \sum_{i=1}^d b(t, \cdot)_i \partial_i v(t, \cdot) = 0, \quad t \in [0, T[ \\ \square & v(T, \cdot) = g, \end{aligned} \tag{1.3}$$

où  $g$  est une fonction définie sur  $\mathbb{R}^d$  à valeurs dans  $\mathbb{R}$ . Il est bien connu que les EDP (1.2) et (1.3) peuvent être représentées en utilisant les solutions  $X^{s,x}$ , lorsqu'elles existent, de l'EDS

$$X_t^{s,x} = x + \int_s^t b(r, X_r^{s,x}) ds + \int_s^t \sigma(r, X_r^{s,x}) dW_r, \quad t \in [0, T], \tag{1.4}$$

pour tout  $(s, x) \in [0, T] \times \mathbb{R}^d$ , où  $W$  est un mouvement brownien.

Plus précisément, la fonction  $\mathbf{u} : t \mapsto \int_{\mathbb{R}^d} \mathbb{E}_{t,0,x} \nu(dx)$  fournit une solution à valeurs

mesures de (1.2). Le cas  $b \equiv 0, \sigma \equiv I_d$  donnant le lien classique entre mouvement brownien et équation de la *chaleur*. Les équations du type (1.2) sont des EDP *conservatives*, dans le sens où la masse de la fonction  $t \mapsto \int_{\mathbb{R}^d} \mathbf{u}(t)(dx)$  est constante.

De plus, les EDP du type (1.3) sont généralement représentées par une formule de type *Feynman-Kac*. En particulier, si  $v$  est une solution régulière de (1.3) alors

$$v(s, x) = \mathbf{E} \left[ g \left( X_T^{s,x} \right) \right], (s, x) \in [0, T] \times \mathbb{R}^d.$$

Il est possible d'obtenir des représentations probabilistes de versions non-linéaires des précédentes EDP. Dans ce cas, d'autres objets probabilistes émergent. Parmi beaucoup d'autres, deux classes de processus de représentation sont d'une importance particulière.

1. Les solutions d'EDS de *McKean*, à savoir des EDS dans lesquelles les coefficients ne dépendent pas uniquement du temps et de la position des *particules*, mais également de leurs lois (marginales). De tels objets fournissent des solutions de versions non-linéaires de l'EDP (1.2), voir [85]. Dans le Chapitre 3, nous étudierons le fonctionnement de telles représentations, avec d'importantes extensions à une grande classe d'EDP non conservatives.
2. Les solutions d'*Equations différentielles stochastiques progressives rétrogrades* (EDSPR). Une EDSPR est un système d'équations composé d'une EDS progressive et d'une équation différentielle stochastique avec condition terminale, i.e. une EDS rétrograde (EDSR), dans le sens de Pardoux et Peng, voir par exemple [91]. Une solution d'EDSR est en général un couple de processus progressivement mesurables. Dans cette introduction nous ne mentionnerons que le cas où l'EDS progressive (markovienne) est du type (0.3), i.e. le cas où EDS et EDSR sont découplées. En général, l'EDSR est définie par un *driver*  $F : [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}$  et une fonction *terminale*  $g$ . Plus précisément, résoudre la BSDE (markovienne) avec driver  $F$  et fonction terminale  $g$  se réduit à trouver un couple de processus progressivement mesurables  $(Y^{s,x}, Z^{s,x})$  tels que

$$Y_t^{s,x} = g \left( X_T^{s,x} \right) + \int_t^T F \left( r, X_r^{s,x}, Y_r^{s,x}, Z_r^{s,x} \right) dr - \int_t^T Z_r^{s,x} dW_r, t \in [s, T], \quad (1.5)$$

pour tout  $(s, x) \in [0, T] \times \mathbb{R}^d$ .

Ces équations sont des outils probabilistes permettant de représenter les solutions de *viscosité* de généralisations semi-linéaires de (0.2), i.e.

$$\begin{aligned} & \square \quad \partial_t v(t, \cdot) + \frac{1}{2} \sum_{i,j=1}^d \sigma \sigma^T(t, \cdot)_{ij} \partial_{ij} v(t, \cdot) + \sum_{i=1}^d b(t, \cdot)_i \partial_i v(t, \cdot) + F \left( t, \cdot, v(t, \cdot), \sigma(t, \cdot)^T \nabla_x v(t, \cdot) \right) = 0 \\ & \square \quad v(T, \cdot) = g \end{aligned} \quad (1.6)$$

pour tout  $t \in [0, T[$ . Sous des hypothèses générales, la fonction  $v := (s, x) \mapsto Y_s^{s,x}$  fournit une solution de viscosité de (1.6), voir par exemple [92].

Si  $\sigma$  est non-dégénérée, une importante particularité des EDSR de type (1.4)-(1.5) réside dans la possibilité de changer le processus progressif  $X$  (devenant solution d'une EDS avec une autre dérive  $b$ ) sans modifier la fonction  $v := (s, x) \mapsto Y_s^{s,x}$ , à condition de remplacer le driver  $F$  par

$$\tilde{F} := (t, x, v, z) \mapsto F(t, x, v, z) - \sigma^{-1}(t, x) \cdot (b - b(t, x), z) \cdot.$$



Ce qui précède signifie qu'une EDS semi-linéaire donnée peut être représentée par différentes EDSR. Sur le plan théorique, ces représentations sont équivalentes. Sur le plan numérique, le choix d'une dérive ad hoc est cruciale. En effet, les méthodes numériques classiques permettant de résoudre des EDSR reposent sur des schémas *progressifs rétrogrades*. Dans ces schémas sont simulées des trajectoires du processus progressif choisi, ces trajectoires étant utilisées pour estimer, de manière rétrograde en temps, des espérances conditionnelles par moindres carrés Monte-Carlo, voir e.g. [50].

La solution de l'EDSR est donnée par ces espérances conditionnelles, estimées par régressions Monte-Carlo. Les trajectoires du processus progressif constituant une *grille d'estimation*. Lorsque la dimension  $d$  de l'espace d'état augmente, il est nécessaire de simuler et de garder en mémoire un grand nombre de trajectoires du processus progressif de manière à obtenir une bonne approximation de la solution. Pour contourner le problème de la dimension, il est nécessaire de guider le processus progressif de manière à explorer efficacement l'espace d'état.

Une EDP pour laquelle le choix de la dérive guidant la grille d'estimation est naturel, est l'équation d'*Hamilton-Jacobi-Bellman* (HJB), lorsqu'elle est semi-linéaire. En toute généralité, l'équation de HJB est une EDP potentiellement non-linéaire utilisée pour étudier des *problèmes de contrôle stochastique*.

L'équation s'écrit

$$\begin{aligned} \square \partial_t v(t, \cdot) + H(t, \cdot, v(t, \cdot), \nabla_x v(t, \cdot), \nabla_x^2 v(t, \cdot)) &= 0 \\ \square v(T, \cdot) &= g, \end{aligned} \quad (1.7)$$

où le hamiltonien  $H$  s'écrit  $\square$

$$H : (t, x, \delta, \gamma) \mapsto \inf_{a \in A} \square \left[ f(t, x, a) + \sum_{i=1}^d b(t, x, a)_i \delta_i + \frac{1}{2} \sum_{i,j=1}^d \sigma \sigma^T(t, x, a)_{ij} \gamma_{ij} \right], \quad (1.8)$$

avec  $A \subseteq \mathbb{R}^k$ , étant donné  $k \in \mathbb{N}^*$ . L'équation (1.7) est en fait l'équation de la programmation dynamique associée au problème de contrôle stochastique avec fonction valeur

$$v : (t, x) \rightarrow \inf_{\alpha} \mathbb{E} \int_t^T f(s, X_s^{t,x,\alpha}, \alpha_s) ds + g(X_T^{t,x,\alpha}), \quad (1.9)$$

où  $X^{t,x,\alpha}$  est une solution, lorsqu'elle existe, de l'EDS *contrôlée*

$$dX_t = b(t, X_t, \alpha_t) dt + \sigma(t, X_t, \alpha_t) dW_t, \quad t \in [0, T], \quad (1.10)$$

débutant au temps  $t \in [0, T]$  avec valeur  $x \in \mathbb{R}^d$ , étant donné un processus progressivement mesurable  $\alpha$  à valeurs dans  $A$ .

Sous des hypothèses générales,  $v$  est l'unique solution de viscosité de (0.6), voir par exemple [108, 42, 94].

L'équation (1.7) est une EDP semi-linéaire lorsque le contrôle n'intervient pas dans la fonction  $\sigma$ . Par ce qui précède, les solutions de viscosité peuvent alors être représentées à l'aide d'EDSR. Dans ce contexte, le choix naturel de dérive mentionné auparavant est guidé par le problème de contrôle stochastique.

Notre idée est de sélectionner une dérive qui localise le processus progressif dans les régions visitées par les trajectoires optimalement contrôlées. Ainsi, le processus optimalement contrôlé est un bon candidat pour notre grille d'estimation, en terme de coût de calcul. Toutefois, ce processus ne peut être simulé de manière progressive puisque les contrôles sont découverts de manière rétrograde en temps.

Cette observation suggère la stratégie suivante pour utiliser effectivement le processus optimalement contrôlé en tant que grille d'estimation. Si des informations sont disponibles sur sa loi terminale, l'idée est de le simuler de manière rétrograde en temps. Ceci permettra en particulier de contourner les problèmes de mémoire et de dimension.

Avec cette approche, nous nous attendons à ce que le nombre de trajectoires nécessaires pour estimer la fonction valeur décroisse significativement.

Dans le cas général d'EDP semi-linéaires du type (1.6), une dérive ad hoc ayant été choisie pour guider le processus progressif d'une EDSR, une manière de tenir compte de la condition terminale  $g$  est de le simuler de manière rétrograde en temps. Sur le plan théorique, ceci requiert la mise en place de représentations *complètement rétrogrades* d'EDP semi-linéaires, à savoir des représentations à base d'EDSR (markoviennes) dans lesquelles le processus progressif choisi  $\xi$  évolue de manière rétrograde en temps. Bien que la dérive soit fixée, la loi du processus  $X$  n'est pas déterminée, puisqu'aucune condition initiale  $X_0$  n'est fixée. En accord avec le précédent objectif,  $\xi$  sera le retourné en temps d'un processus markovien  $X$ .

Le retournement des processus markoviens a été exploré par de nombreux auteurs, voir par exemple [55] pour le cas des diffusions en dimension finie, [45] pour le cas des diffusions en dimension infinie et [64] pour le cas des processus à sauts. Nous mentionnons également les deux preprints récents [31, 32] en relation avec l'entropie.

Etant donné un processus de diffusion  $X$  avec coefficients  $b, \sigma$ , [55] fournit des conditions suffisantes sur les coefficients et sur les densités marginales (si elles existent)  $(q_t)_{t \in [0, T]}$  de  $X$  garantissant que  $X := X_T \dots$  soit également un processus de diffusion. De plus, il est prouvé que  $X$  est une solution (en loi) de l'EDS

$$X_t = X_T - \int_0^t b(T-r, X_r) dr + \int_0^t \text{div}_y \left( \frac{\sigma \sigma_i^T(T-r, X_r) q_{T-r}(X_r)}{q_{T-r}(X_r)} \right)_{i \in [1, d]} dr + \int_0^t \sigma(T-r, X_r) d\beta_r, \tag{1.11}$$

où  $\beta$  est un mouvement brownien.

Pour simuler  $X$ , (1.11) ne peut être utilisée en pratique. En effet, même en choisissant une loi terminale  $X_T$  compatible avec  $g$ , les lois marginales inconnues de  $X$  apparaissent dans (1.11).

Une idée pour dépasser cette limitation consiste à remplacer l'équation (1.11) par une EDS de type McKean, dans laquelle les lois marginales apparaissant dans la dynamique du retourné en temps ne

sont plus des paramètres exogènes mais une partie de la solution, i.e.

$$\begin{aligned} \square & \xi_t = \xi_0 - \int_0^t b(T-r, \xi_r) dr + \int_0^t \frac{\text{div}_y \sigma \sigma_i^{-T}(T-r, \xi_r) p_r(\xi_r)}{p_r(\xi_r)} dr + \int_0^t \sigma(T-r, \xi_r) d\beta_r, \\ \square & \xi_0 \sim \nu \\ \square & p_t \text{ densité de } \xi_t, t \in ]0, T[ \end{aligned} \tag{1.12}$$

où  $\nu$  est une probabilité donnée, dont la solution est le couple  $(\xi, \mathbf{p})$  où  $\mathbf{p}$  est la fonction à valeurs probabilités telle que pour tout  $t \in ]0, T[$ ,  $\mathbf{p}_t$  admet  $p_t$  pour densité. Par [55], le couple  $(X, \mathbf{q}_{T-\cdot})$  est une solution de (1.12) lorsque  $\nu$  est la loi de  $X_T$ .

Si l'EDS de McKean (1.12) est bien posée, l'unique solution  $\xi$  coïncidera avec le processus recherché  $X$ . Dans ce cas, (1.12) fournit une dynamique intrinsèque pour  $X$ .

Au delà du retournement des processus de diffusion et des problèmes de contrôle stochastique, l'étude des questions d'existence et d'unicité pour l'équation (1.12) a son intérêt propre. En effet, cette équation fournit une représentation probabiliste de l'équation de *Fokker-Planck* avec condition terminale

$$\begin{aligned} \square & \partial_t \mathbf{u}(t) = \frac{1}{2} \sum_{i,j=1}^d \partial_{ij} \sigma \sigma^{-T}(t, \cdot)_{ij} \mathbf{u}(t) - \sum_{i=1}^d \partial_i (b(t, \cdot)_i \mathbf{u}(t)), t \in [0, T[ \\ \square & \mathbf{u}(T) = \nu. \end{aligned} \tag{1.13}$$

Les problèmes d'évolution avec condition terminale tels que (1.13) émergent dans de nombreuses applications impliquant des problèmes *inverses*, en particulier en sciences physiques, avec l'exemple de la conduction de chaleur [14], des exemples en science des matériaux [95] ou encore en hydrologie [5]. Dans ces applications, la configuration terminale d'un système physique (modélisé par un processus de diffusion) est observée; ainsi les questions d'unicité (et de stabilité) sont plus importantes que la question de l'existence: le but est de déterminer des configurations initiales admissibles. Les problèmes d'unicité et de stabilité ont été en particulier traités dans la littérature par des techniques de régularisation, voir e.g [107, 74]. Toutefois, un faible nombre de résultats concernant l'unicité et l'approximation des solutions de l'équation générale (1.13) sont disponibles et les méthodes de preuves sont purement analytiques. Trouver une représentation probabiliste pourrait apporter un regard nouveau sur (1.13) et l'approximation de ses solutions.

Dans le Chapitre 1, le problème de l'unicité pour l'EDP (1.13) est abordé avec des méthodes analytiques et probabilistes. De plus, nous étudions les connections entre l'EDP (1.13) et l'EDS de type McKean (1.12). Ces connections sont utilisées pour obtenir des résultats d'existence et d'unicité pour (1.12). Il apparaît en particulier que l'unicité pour (0.12) est un outil fondamental pour étudier l'EDS de type McKean; ceci mis en évidence par la Proposition 1.3.2 qui établit que si  $(\xi, \mathbf{p})$  est une solution de (1.12) alors  $\mathbf{p}_{T-\cdot}$  est une solution de l'EDP (1.13). Plus précisément, les contributions du Chapitre 1 sont de deux natures:

1. L'unicité pour l'EDP (1.13) est abordée dans différentes classes de coefficients  $b, \sigma$  et différentes classes de solutions à valeurs mesures. Nous obtenons essentiellement trois types de résultats:

S'agissant du premier, l'unicité pour (1.13) est prouvée parmi les solutions ayant valeur initiale dans une certaine classe  $C$  de mesures finies, supposant que l'équation de Fokker-Planck progressive (1.2) admet au plus une solution avec valeur initiale dans  $C$ . En particulier nous traitons l'unicité parmi les solutions à valeurs mesures positives finies, ayant un multiple d'une masse de Dirac pour valeur initiale. Nous obtenons des résultats dans le cas de coefficients Lipschitz potentiellement dégénérés et dans le cas de coefficients bornés, non-dégénérés et potentiellement irréguliers. Nous utilisons ici des méthodes purement probabilistes. Ces résultats sont l'objet des Théorèmes 1.3.9 et 1.3.10.

S'agissant du deuxième, l'unicité est prouvée dans la classe des solutions à valeurs mesures finies, sans supposer l'unicité pour l'EDP progressive. Nous traitons le cas général de coefficients indépendants du temps, bornés, Hölder et non-dégénérés avec extension au cas de coefficients indépendants du temps par morceaux. Nous utilisons ici une méthode analytique reposant sur une propriété cruciale du semigroupe généré par l'opérateur de diffusion associé aux coefficients  $b, \sigma$ . Ces résultats sont l'objet du Théorème 1.3.13 du Corollaire 1.3.16.

Finalement, nous abordons la situation d'une EDP associée à un semigroupe inhomogène en temps de type *Ornstein-Uhlenbeck*. Ici, la méthode consiste à opérer un changement de variable dans l'EDP initiale qui permet de se ramener au cas de l'équation de la *chaleur rétrograde* pour laquelle l'unicité est aisément obtenue en utilisant la transformée de Fourier comme dans le cas progressif, voir Théorème 1.3.19.

2. Nous prouvons des résultats d'existence et d'unicité pour l'EDS de type McKean (1.12). Nous commençons par lier existence et unicité pour l'EDP (1.13) et l'EDS (1.12). Ensuite, nous discutons existence et unicité en loi pour (1.12) dans le cas de coefficients bornés réguliers, voir Théorème 1.4.15. Finalement, nous établissons des résultats d'existence forte et d'unicité trajectorielle dans le cas Ornstein-Uhlenbeck, voir Théorème 1.4.19.

Revenant au contrôle stochastique et aux considérations suivant (1.10), les résultats d'existence et d'unicité obtenus dans le Chapitre 1 donnent un point de départ théorique pour obtenir une représentation totalement rétrograde de l'équation d'Hamilton-Jacobi-Bellman impliquant le retourné en temps d'une approximation du processus optimalement contrôlé. Toutefois, ceci amène des défis tant sur le plan théorique que numérique.

1. Sur le plan théorique, Il est nécessaire de trouver une classe ad hoc de dérivées approchant raisonnablement celle du processus optimalement contrôlé et nous permettant d'utiliser les résultats obtenus dans le Chapitre 1.
2. Sur le plan numérique, nous souhaitons approcher le processus optimalement contrôlé. Suivant les méthodes développées dans le Chapitre 1, l'apparition de densités marginales dans le

problème de McKean est un important obstacle; le coût de leur estimation (non paramétrique) compensant le gain obtenu en termes de mémoire et de dimension, see e.g. [103]. Par ailleurs, il s'avère nécessaire d'avoir des informations sur la loi terminale; une telle information étant hautement dépendante de la nature du problème d'optimisation en jeu, en particulier du coût terminal.

La classe des dérivées affines nous permet de remplir les précédents objectifs. Dans le Chapitre 2, nous nous intéressons au choix d'une dérivée affine dépendant du temps  $b : (s, x) \rightarrow a(s)x + c(s)$  où  $a$  (resp.  $c$ ) est une fonction continue par morceaux définie sur  $[0, T]$  à valeurs dans  $M_d(\mathbb{R})$  (resp.  $\mathbb{R}^d$ ).

Dans un premier temps, nous obtenons une représentations totalement rétrograde pour des EDP semi linéaires générales du type

$$\begin{aligned} \square & \partial_t v(t, \cdot) + \frac{1}{2} \sum_{i,j=1}^d \sigma \sigma^T(t)_{ij} \partial_{ij} v(t, \cdot) + H(t, \cdot, v(t, \cdot), \nabla_x v(t, \cdot)) = 0, \quad t \in [0, T[ \\ \square & v(T, \cdot) = g. \end{aligned} \tag{1.14}$$

où  $\sigma$  est une fonction déterministe continue à valeurs matricielles.

Le concept de représentation totalement rétrograde évoqué auparavant, repose sur un processus  $\xi$  évoluant de manière rétrograde vis à vis du temps de l'EDP (1.14).

Pour définir  $\xi$ , nous fixons une mesure gaussienne  $\nu$  qui sera la loi de  $\xi_0$ . Soit  $\bar{m}^\nu$  (resp.  $\bar{Q}^\nu$ ) la moyenne (resp. la matrice de covariance) de la mesure  $\nu$ . Le processus  $\xi$  sera alors l'unique solution forte de

$$\xi_t = \xi_0 - \int_0^t b(T-r, \xi_r) + \sigma \sigma^T(T-r) Q(T-r)^{-1} (\xi_r - m(T-r)) dr + \int_0^t \sigma(T-r) d\beta_r, \quad t \in [0, T], \tag{1.15}$$

où  $\beta$  est un mouvement brownien,  $t \mapsto m(T-t) = \mathbf{E}(\xi_t)$  and  $t \mapsto Q(T-t)$ , la matrice de covariance de  $\xi_t$ . Il est aisé de prouver existence et unicité pour l'EDS (1.15) et le fait que  $m : [0, T] \mapsto \mathbb{R}^d$ ,  $Q : [0, T] \mapsto S_d(\mathbb{R})$  soient solutions des EDO rétrogrades suivantes

$$\begin{aligned} \square & \frac{d}{dt} m(t) = a(t) m(t) + c(t), \quad t \in [0, T] \\ \square & m(T) = \bar{m}^\nu, \end{aligned} \tag{1.16}$$

$$\begin{aligned} \square & \frac{d}{dt} Q(t) = Q(t) a(t)^T + a(t) Q(t) + \sigma \sigma^T(t), \quad t \in [0, T] \\ \square & Q = \bar{Q} \end{aligned} \tag{1.17}$$

à condition que  $Q(0)$  soit définie positive.

Sous des hypothèses générales concernant les fonctions  $H, g$ , le théorème ?? établit qu'une fonction  $v : [0, T] \times \mathbb{R}^d \mapsto \mathbb{R}$  continue en temps et continûment différentiable en espace avec gradient à croissance polynomiale, est une solution de viscosité de (1.14) si et seulement si pour tout  $t \in [0, T]$

$$v(t, \xi_t) = \mathbf{E} \int_t^T H(s, \xi_s, v(s, \xi_s), \nabla_x v(s, \xi_s) - b(s, \xi_s), \nabla_x v(s, \xi_s)) ds + g(\xi_T | \xi_t), \tag{1.18}$$

where  $\xi := \xi_{T-\cdot}$ .

La représentation totalement rétrograde de (1.14) consiste en le couplage (1.15)-(1.18). Remarquons ici que le lemme 2.3.6 établit que  $\xi$  est en fait le retourné en temps de la solution d'une EDS du type

$$dX_t = b(t, X_t) dt + \sigma(t) dW_t, \quad (1.19)$$

avec loi initiale non fixée.

Dans le corollaire 2.4.6, ce résultat est appliqué au cas où l'EDP (1.14) est l'équation de HJB associé à un problème de contrôle avec coefficient de diffusion  $\sigma$ . Dans ce cas, le couplage (1.15)-(1.18) devient une formule de représentation pour la fonction valeur du problème de contrôle mentionné ci-dessus. Ceci nous mène à suggérer un algorithme heuristique pour résoudre ce problème de contrôle, i.e. pour approcher sa fonction valeur et déterminer une stratégie optimale.

Notre approche consiste à déterminer un processus  $\xi$  solution de (1.15), dont la dérive affine  $\tilde{b}$  (et donc  $a$  et  $c$ ) est découverte simultanément avec la fonction valeur,  $\tilde{b}$  étant déterminée par régression linéaire, à chaque pas de temps, de manière à être aussi proche que possible de la dérive du processus optimalement contrôlé. Le processus  $\xi$  obtenu sera notre grille d'estimation, générée de manière rétrograde and supposée explorer efficacement les régions d'intérêt du problème de contrôle. Cette grille est *adaptive* dans le sens où elle utilise, à chaque pas de temps, l'estimation obtenue de la fonction valeur, en particulier de son gradient. Ainsi, cette méthode remplit l'objectif d'exploration parcimonieuse de l'espace d'état. Elle a aussi l'avantage de contourner le problème de mémoire puisqu'il n'est pas nécessaire de stocker la grille d'estimation. Finalement, la précision de la procédure est illustrée sur un problème de contrôle stochastique intervenant dans le contexte de la gestion de la demande dans les réseaux électriques.

Dans le Chapitre 3, nous nous concentrons sur l'utilisation des solutions d'EDS de type McKean pour représenter les solutions d'extensions non-conservatives (et non linéaires) de l'équation (1.2). Plus précisément, nous considérons des versions *perturbées* de (1.2) de la forme

$$\begin{aligned} \square \quad \partial_t u(t, \cdot) &= \frac{1}{2} \sum_{i,j=1}^d \partial_{ij} \sigma \sigma^T(t, \cdot, u(t, \cdot))_{ij} u(t, \cdot) - \sum_{i=1}^d \partial_i (b(t, \cdot, u(t, \cdot)))_i u(t, \cdot) \\ \square \quad &+ \Lambda(t, \cdot, u(t, \cdot), \nabla_x u(t, \cdot)) u(t, \cdot), \quad t \in ]0, T], \\ \square \quad u(0, \cdot) &= v \end{aligned} \quad (1.20)$$

où  $\Lambda : [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}$ . Les objets probabilistes qui émergent pour représenter (1.20) sont solutions de l'EDS de type McKean suivante

$$\begin{aligned} \square \quad Y &= Y_0 + \int_0^t \sigma(s, Y_s, u(s, Y_s)) dW_s + \int_0^t b(s, Y_s, u(s, Y_s)) ds \\ \square \quad Y_0 &\sim v \\ \square \quad \int \phi(x) u(t, x) dx &= \mathbf{E} \left[ \phi(Y_t) \exp \left\{ \int_0^t \Lambda(s, Y_s, u(s, Y_s), \nabla u(s, Y_s)) ds \right\} \right], \quad t \in ]0, T], \end{aligned} \quad (1.21)$$

pour toute fonction test à valeurs réelles  $\phi$  continue bornée. La seconde ligne apparaissant dans (1.21), est appelée *linking equation* et peut être vue comme une extension *trajectoire-dépendante* de la dépendance vis à vis de la loi observée dans les EDS de type McKean usuelles (à savoir lorsque  $\Lambda = 0$ ). Ceci ouvre la voie à de nouvelles méthodes numériques fondées sur des *systèmes de particules en interaction*. Nous détaillons ces méthodes et donnons certaines classes d'EDP non-conservatives pour lesquelles la précédente représentation présente un intérêt.

**Titre:** Techniques numériques rétrogrades de type McKean pour les EDPs et une application à la gestion de l'énergie.

**Mots clés:** Equation Différentielle Stochastique (EDS) de type McKean; représentation probabiliste d'EDPs; équation d'Hamilton-Jacobi-Bellman (HJB); problème inverse; équation de Fokker-Planck; retourné en temps d'une diffusion; contrôle stochastique; schéma de régression de type Monte-Carlo; gestion de la demande.

**Résumé:** Cette thèse s'intéresse aux équations différentielles stochastiques de type McKean (EDS) et à leur utilisation pour représenter des équations aux dérivées partielles (EDP) non linéaires. Ces équations ne dépendent pas seulement du temps et de la position d'une certaine base, nous proposons un algorithme de Monte-Carlo particule mais également de sa loi. En particulier nous traitons le cas inhabituel de la représentation d'EDP de type Fokker-Planck avec condition terminale fixée. Nous discutons l'existence et l'unicité pour ces EDP et de leur représentation progressive-rétrograde. Nous appliquons cette méthode de représentation sous la forme d'une EDS de type McK- dans le contexte de la gestion de la demande dans les réseaux électriques. Pour finir, nous faisons le point sur le retourné dans le temps d'un processus de diffusion. Nous introduisons la notion de représentation complète-ment rétrograde d'une EDP semilinéaire. Celle-ci consiste dans le couplage d'une EDS rétrograde et d'un pro-

**Title:** Probabilistic backward McKean numerical methods for PDEs and one application to energy management.

**Keywords:** McKean stochastic differential equation; probabilistic representation of PDEs; Hamilton Jacobi Bellman (HJB) equation; inverse problem; Fokker-Planck equation; time-reversal of diffusion; stochastic control; regression Monte-Carlo scheme; demand-side management.

**Abstract:** This thesis concerns McKean Stochastic Differential Equations (SDEs) to represent possibly non-linear Partial Differential Equations (PDEs). Those depend not only on the time and position of a given particle, but also on its probability law. In particular, we treat the unusual case of Fokker-Planck type PDEs with prescribed final data. We discuss existence and uniqueness for those equations and provide a probabilistic representation in the form of McKean type equation, whose unique solution corresponds to the time-reversal dynamics of a diffusion process. We introduce the notion of fully backward representation of a semilinear PDE: that consists in fact in the coupling of a classical Backward