UNIVERSITÀ DEGLI STUDI DI MILANO
FACOLTÀ DI SCIENZE E TECNOLOGIE

Dipartimento di Matematica “Federigo Enriques”

Dottorato di Ricerca in Scienze Matematiche

Ciclo XXXV

Tesi di Dottorato di Ricerca

Advances in stochastic analysis on spaces of measures: Kolmogorov equations related to stochastic filtering and mean field optimal stopping

Settore scientifico-disciplinare: MAT/06

Relatore:
Prof. Marco Fuhrman

Coordinatore del dottorato:
Prof. Dario Paolo Bambusi

Candidato:
Mattia Martini

A. A. 2021-2022

Thesis Summary

The purpose of this Thesis is to study some problems arising in the context of stochastic analysis and stochastic optimal control, where some relevant variables take values in the space of positive or probability measures.

Most of the work is dedicated to the introduction and study of some backward Kolmogorov equations on spaces of measures associated to filtering problems. Measure-valued processes arise naturally in the context of stochastic filtering and one can formulate two stochastic differential equations, called Zakai (Z) and Kushner-Stratonovich (KS) equations, that are solved by a positive and a probability measure-valued process respectively. A classical way to study these problems is to assume that the measure-valued processes admit a density and then exploit stochastic calculus in Hilbert spaces.

The approach used in this Thesis differs from this since we do not assume the existence of a density and we work directly in the context of measures. We formulate two backward Kolmogorov equations, which are parabolic partial differential equations with a given final condition, one over a space of positive (in the Z case) and one over a space of probability (in the KS case) measures. In the recent literature, PDEs on spaces of probability measures have been a topic of great interest, thanks to the connection with mean field games and McKean-Vlasov equations. Here we state a new equation on a space of probability measures, obtained in a completely different framework, and we introduce a PDE on a space of positive measures, which is one of the first problems of this type. We study the existence and uniqueness of a solution from two different points of view.

First, we focus on classical solutions. We prove that both the equations introduced admit a unique classical solution, as long as the final condition is chosen regular enough. To achieve this, we need some intermediate results of independent interest. In particular, we prove Itô formulas for the composition of filtering processes and real-valued functions defined on the space of measures. Moreover, we study the regularity of the solution to the filtering equations with respect to the initial datum. Since we are dealing with functions over spaces of measures, a key point is to discuss proper notions of derivatives, especially in the case of positive measures.

After this, we investigate the case when the final condition is less regular. A first remark is that the approach we used in the classical case strongly depends on the regularity of the final condition. Thus, we need to change the notion of solution and look for viscosity solutions. To this aim, we focus on the PDE associated to the KS equation. In the literature, only a few results are available on viscosity solutions for this kind of problem, and in particular, uniqueness is a very challenging issue. In this

case, we prove the existence and uniqueness of a viscosity solution and, in particular, we provide a comparison theorem.

In the final part of this Thesis, we deal with a different problem, which arises in the context of optimal stopping theory. We study a class of finite horizon time-inconsistent optimal stopping problems (OSPs) of mean field type, which includes those related to mean field diffusion processes and recursive utility functions. The mean field interaction is due to the fact that we consider a terminal cost depending not only on the stopped process but also on its law (which is different from the law of the process evaluated at the stopping time).

Despite the time-inconsistency of the OSP, we show that it is optimal to stop when the value-process hits the reward process for the first time, as is the case for the standard time-consistent OSP. We solve the problem by approximating the corresponding value-process with a sequence of Snell envelopes of processes, for which a sequence of optimal stopping times is constituted of hitting times of each of the reward processes by the associated value-process. Then, under mild assumptions, we show that this sequence of hitting times converges in probability to the hitting time for the mean field OSP and that the limit is optimal.

The research for this Thesis resulted in the following works:

1. M. Martini, *Kolmogorov equations on spaces of measures associated to nonlinear filtering processes*, Preprint: arXiv:2107.11865, 2021.

2. M. Martini, *Kolmogorov equations on the space of probability measures associated to the nonlinear filtering equation: the viscosity approach*, Preprint: arXiv:2202.11072, 2022.

3. B. Djehiche, M. Martini, *Mean Field Optimal Stopping: a limit approach*, Preprint: arXiv:2209.04174, 2022.

All of these preprints have been submitted to a journal.

Riassunto della Tesi

Lo scopo di questa Tesi è di studiare alcuni problemi di analisi stocastica e controllo ottimo stocastico, dove alcune variabili prendono valore in spazi di misure positive e di probabilità.

La maggior parte del lavoro è dedicata all'introduzione e allo studio di alcune equazioni di Kolmogorov retrograde su spazi di misure associate a problemi di filtraggio. I processi a valori in spazi di misure sorgono naturalmente nel contesto del filtraggio stocastico, dove si possono formulare due equazioni differenziali stocastiche, dette equazioni di Zakai (Z) e di Kushner-Stratonovich (KS), che vengono risolte rispettivamente da un processo a valori nelle misure positive e da un provesso a valori nelle probabilità. Un modo classico per studiare questi problemi è assumere che i processi ammettano una densità per poi sfruttare il calcolo stocastico su spazi Hilbert.

L'approccio utilizzato in questa tesi differisce da questo poiché non assumiamo l'esistenza di una densità e lavoriamo direttamente nel contesto delle misure. Formuliamo due equazioni di Kolmogorov retrograde, che sono equazioni alle derivate parziali paraboliche con una data condizione finale, una su uno spazio di misure positive (nel caso Z) e una su uno spazio di misure di probabilità (nel caso KS). Nella letteratura recente, le EDP su spazi di misure di probabilità sono state un argomento di grande interesse, grazie alla connessione con i giochi a campo medio e le equazioni di McKean-Vlasov. Qui presentiamo una nuova equazione su uno spazio di misure di probabilità, ottenuta in un modo completamente diverso, e introduciamo una EDP su uno spazio di misure positive, che è uno dei primi problemi di questo tipo. Studiamo l'esistenza e l'unicità di una soluzione da due diversi punti di vista.

Innanzitutto, ci concentriamo sulle soluzioni classiche. Dimostriamo che entrambe le equazioni introdotte ammettono un'unica soluzione classica, purché la condizione finale sia scelta sufficientemente regolare. Per raggiungere questo risultato, abbiamo bisogno di alcuni risultati intermedi di interesse indipendente. In particolare, mostriamo delle formule di Itô per la composizione di processi di filtraggio e funzioni a valori reali definite su spazi di misure. Inoltre, studiamo la regolarità della soluzione delle equazioni di filtraggio rispetto al dato iniziale. Trattandosi di funzioni su spazi di misure, un punto chiave è discutere opportunamente le nozioni di derivate, specialmente nel caso di misure positive.

Successivamente, esaminiamo il caso in cui la condizione finale è meno regolare. Una prima osservazione è che l'approccio usato nel caso classico dipende fortemente dalla regolarità della condizione finale. Pertanto, dobbiamo cambiare la nozione di soluzione e cercare soluzioni di viscosità. A questo scopo, ci concentriamo sulla EDP associata all'equazione di KS. In letteratura sono disponibili solo pochi risultati sulle soluzioni di viscosità per questi tipi di problema e, in particolare, l'unicità è un problema molto difficile. In questo caso dimostriamo l'esistenza e l'unicità di una soluzione di viscosità e, in particolare, forniamo un teorema del confronto.

Nella parte finale di questa Tesi, affrontiamo un problema diverso, proveniente dalla teoria dell'arresto ottimale. Studiamo una classe di problemi di arresto ottimo tempo-inconsistenti ad orizzonte finito di tipo campo medio, che include quelli relativi a processi di diffusione di tipo McKean-Vlasov e funzioni di utilità ricorsive. L'interazione di tipo campo medio è dovuta al fatto che consideriamo un costo terminale dipendente non solo dal processo arrestato ma anche dalla sua legge (che è diversa dalla legge del processo valutata al tempo di arresto).

Nonostante l'inconsistenza temporale del problema, dimostriamo che è ottimale fermarsi quando il processo valore colpisce per la prima volta il reward, come nel caso standard tempo consistente. Risolviamo il problema approssimando il processo valore con una sequenza di inviluppi Snell. Una sequenza di tempi di arresto ottimali è quindi data dai tempi di contatto di ciascuno dei processi valore con il rispettivo reward. Quindi, sotto opportune ipotesi, mostriamo che questa sequenza di tempi di contatto converge in probabilità al tempo di contatto per il problema di arresto ottimo di tipo campo medio, e che il limite è ottimale.

L'attività di ricerca per questa Tesi ha portato alla stesura dei seguenti lavori:

1. M. Martini, *Kolmogorov equations on spaces of measures associated to nonlinear filtering processes*, Preprint: arXiv:2107.11865, 2021.

2. M. Martini, *Kolmogorov equations on the space of probability measures associated to the nonlinear filtering equation: the viscosity approach*, Preprint: arXiv:2202.11072, 2022.

3. B. Djehiche, M. Martini, *Mean Field Optimal Stopping: a limit approach*, Preprint: arXiv:2209.04174, 2022.

Questi lavori sono stati tutti inviati a delle riviste e sono in attesa di revisione.

Acknowledgements

I would like to express my gratitude to my supervisor Prof. Marco Fuhrman for all the essential help and encouragement during my Ph.D. studies, as well as for the precious discussions which led to the research in this Thesis.

Many thanks to Prof. Boualem Djehiche, for sharing inspiring ideas and time with me, which also led to a fruitful collaboration. Many thanks are owed to Prof. Andrea Cosso, for his interest in my work and for all the valuable and stimulating suggestions he gave to me. I am also grateful to Prof. Fulvia Confortola, for her support and for all the exchanges we had over these years.

I thank two anonymous Referees for their feedback on this Thesis, their positive comments and relevant suggestions.

I thank the students and researchers I met in conferences, schools, and teaching activities for the opportunity of engaging in interesting debates.

Finally, I am enormously grateful to all the people whose fundamental support, be it academical or not, made all of this possible.

Contents

Thesis Summary **3**

Riassunto della Tesi **5**

Introduction **11**

I.1 Kolmogorov equations on spaces of measures associated to nonlinear filtering processes: classical solutions . . . 13

I.2 Kolmogorov equations on the space of probability measures associated to the nonlinear filtering equation: viscosity solutions . . . 19

I.3 Mean field optimal stopping: a limit approach . . . 22

Notation and preliminaries . . . 25

1 Nonlinear filtering **29**

1.1 Formulation of the problem . . . 30

1.2 The filtering equations . . . 31

1.3 Weak solutions to the filtering equations . . . 33

1.3.1 Proof of Lemma 1.3.6 . . . 37

2 Properties of functions on spaces of measures **39**

2.1 Derivatives on spaces of measures . . . 39

2.1.1 Some properties . . . 43

2.1.2 The $\mathcal{P}(K)$ case . . . 45

2.2 Approximation of real-valued functions over the space of positive measures . . . 46

3 Itô formulas for filtering processes **55**

3.1 Itô formula for the Zakai equation . . . 56

3.2 Itô formula for Kushner-Stratonovich equation . . . 60

4 Classical solutions **63**

4.1 Regularity with respect to the initial condition for the Zakai equation . 64

4.2 The backward Kolmogorov equation associated to the Zakai equation . 67

4.2.1 Existence and uniqueness of a classical solution . . . 68

4.3 The backward Kolmogorov equation associated to the Kushner-Stratonovich equation . . . 70

4.3.1 The backward Kolmogorov equation . . . 70

5 Viscosity solutions in the Kushner-Stratonovich case **73**
5.1 Statement of the problem . . . 73
5.1.1 Some analytical results . . . 75
5.2 Viscosity solutions to the backward Kolmogorov equation . . . 77
5.2.1 Existence of viscosity solutions . . . 78
5.2.2 Comparison theorem and uniqueness . . . 78

6 Mean Field Optimal Stopping: a limit approach **83**
6.1 Optimal stopping of a recursive utility function . . . 84
6.1.1 Existence and uniqueness of the value-processes . . . 85
6.2 Convergence results . . . 86
6.2.1 Convergence of the particle system . . . 86
6.2.2 Convergence of the optimal stopping times . . . 90
6.2.3 Extension to the general mean field case . . . 94
6.3 Optimal stopping of mean field SDEs . . . 96
6.3.1 Optimal stopping of the variance of a mean field diffusion . . . 97

Introduction

The objective of this Thesis is to study some problems arising in the context of stochastic analysis and stochastic optimal control, where the main difficulty is due to the fact that some relevant variables take values in a space of positive or probability measures.

Most of the effort of this work is devoted to the introduction of a class of partial differential equations of parabolic type, namely a class of backward Kolmogorov equations. These equations are formulated backward in time and the datum is the final condition. The deep connection between these types of equations and Markovian stochastic processes is well known in the finite-dimensional case, and it has been intensively investigated also in some infinite-dimensional cases, such as the Hilbertian case.

The first novelty of this Thesis is the fact that the backward Kolmogorov equations we present are related to two measure-valued stochastic processes arising in the context of nonlinear filtering. Thus, what we obtain are two partial differential equations of parabolic type where the spatial domain is a set of positive or probability measures. More precisely, in the nonlinear filtering framework it is natural to formulate two stochastic differential equations, namely the Zakai and the Kushner-Stratonovich equation, that are solved (in a weak sense) by a positive measure-valued and a probability measure-valued Markov process respectively. Thus, it is natural to think about the backward Kolmogorov equations associated to these stochastic differential equations, which will be partial differential equations over a space of positive and probability measures.

Once the two equations have been stated, we can look at their solutions in many different ways. As a first attempt, we look for classical solutions. In particular, as long as we take a smooth final condition, we can prove that there exists a unique classical solution. The proofs rely on an explicit construction of the solutions, exploiting the measure-valued solutions to the Zakai and Kushner-Stratonovich equations, as in the classical finite-dimensional case. The particular structure of the spaces of measures introduces some technical problems, especially from the differential calculus point of view. Thus, we need to discuss and extend some notions of derivatives, as well as we need to develop a tailor-made approximation technique for smooth real-valued functions over these spaces.

A natural extension is to ask for a less regular final condition. A first remark is that the approach we used to prove the existence and uniqueness of classical solutions strongly depends on the regularity of the final condition and there is no hope to obtain the same result with a lighter hypothesis. Thus, a natural way to proceed is to change the notion of solution, and in our case to look for viscosity solutions. To this aim, we focus on one of the two partial differential equations, namely the one associated to the Kushner-Stratonovich equation. In this case, we prove that there exists a unique viscosity solution. As often happens in the case of viscosity solutions, the main obstacle one has to face is uniqueness, and the so-called comparison principle provides a sufficient condition to achieve it. We prove a version of this result for our equation on the space of probability measures, by using an approximation technique that exploits a family of equations that can be solved classically.

In the final part of this Thesis, we focus on a completely different problem, which arises in the context of stochastic optimal control and more precisely which is related to the optimal stopping theory. This problem is connected to the previous ones by the fact that in both cases the main issue is to introduce approximations techniques to deal with the presence of a measure. We study a class of finite horizon time-inconsistent optimal stopping problems of mean field type. In particular, the mean-filed interaction is due to the fact that we consider a terminal cost depending not only on the stopped process, but also on its law, and due to this the problem becomes time-inconsistent. We provide a characterization of an optimal stopping time as a hitting time, and more precisely we show that it is optimal to stop when the value process hits the reward process for the first time, as in the case of the standard time-consistent optimal stopping problem.

The outline of the thesis is the following. In Chapter 1 we give a quick introduction to the stochastic filtering framework, highlighting some fundamental results and common notations we will use repeatedly in the following chapters. In particular, the Zakai and the Kushner-Stratonovich equation are introduced here, as well as some notions of solutions. In Chapter 2 we present and discuss different types of derivatives on spaces of measures, that are necessary for studying partial differential equations on spaces of measures. Moreover, we introduce some approximation techniques that are used mainly in Chapter 3, and that allows us to approximate real-valued functions over the space of positive and probability measures. Chapter 3 and 4 contains the main results of the first part of the thesis, and they are devoted to the introduction and solution of the backward Kolmogorov equations associated to the nonlinear filtering equations. In particular, in Chapter 3 we prove two Itô formulas: one for the composition of a real-valued function with a solution to the Zakai equation, and one for the composition of a real-valued function with a solution to the Kushner-Stratonovich equation. Regarding Chapter 4, in the first part some regularity properties of the Zakai equation's solution are studied. In the second part, we write the backward Kolmogorov equation associated to the Zakai equation, which is a partial differential equation of parabolic type over a space of positive measures, and we prove that there exists a unique classical solution. Finally, in the third part, we repeat the previous steps for the Kushner-Stratonovich equation, thus we will end up studying a partial differential equation over a space of probability measures. In Chapter 5 we focus on the Kolmogorov equation associated to the Kushner-Stratonovich equation and, under some structural assumption and with a less regular final condition, we prove that there exists a unique viscosity solution. Finally, Chapter 6 is dedicated to the optimal stopping problem of mean field type. To lighten the exposition, in the first part we introduce a simplified version of the problem we are interested in and we solve it, providing a characterization of an optimal stopping time. Then we discuss how to achieve the same result in a more general case. In the final part of this chapter, we focus on the optimal stopping of mean field diffusions and we discuss the case of optimal stopping of the variance.

The original results obtained in this Thesis are contained in Chapters 2, 3, 4, 5 and 6. In particular:

- The results in Chapters 2, 3 and 4 can be found in:
  M. Martini, *Kolmogorov equations on spaces of measures associated to nonlinear filtering processes*, Preprint: arXiv:2107.11865, 2021.

- The results in Chapters 5 can be found in:
  M. Martini, *Kolmogorov equations on the space of probability measures associated to the nonlinear filtering equation: the viscosity approach*, Preprint: arXiv:2202.11072, 2022.

- The results in Chapters 6 are obtained in collaboration with B. Djehiche and can be found in:
  B. Djehiche, M. Martini, *Mean Field Optimal Stopping: a limit approach*, Preprint: arXiv: 2209.04174, 2022.

In the following sections, we give a more detailed introduction to each of the topics presented in this work and we present briefly the main results obtained. Finally, we collect some preliminary remarks and notations, as well as a list of the recurrent symbols.

I.1 Kolmogorov equations on spaces of measures associated to nonlinear filtering processes: classical solutions

The study of measure-valued stochastic processes is a classical topic that has attracted enormous interest. For instance, there is a large literature related to the superprocesses framework (see for instance [42]), but more recently it has been related to the topic of mean field games and McKean-Vlasov equations (see [28, 79] or [31, 32]), where probability measure-valued processes are used in problems with common noise to describe the evolution of the conditional laws of some finite-dimensional stochastic processes. Thanks to this recent interest, many new results are now available, such as Itô formulas ([31, 32]) and tools for differential calculus on spaces of measures ([28, 31, 79]). Moreover, a topic of great interest is the partial differential equations on spaces of probability measures associated to these problems, such as, for instance, the so-called master equation in the context of mean field games (see for instance [30, 32]), the backward Kolmogorov equation associated to McKean-Vlasov type equation (see for instance [24]), or certain Hamilton-Jacobi equations ([63]).

We give a contribution in this direction. Differently, from the previous contexts, we aim to study partial differential equations on space of measures associated to measure-valued processes arising in stochastic filtering problems. In particular, given a measure-valued process, we first introduce the so-called backward Kolmogorov equation associated to it, which is a partial differential equation on a space of measures. Then, we study the existence and uniqueness of its classical solutions.

Stochastic filtering has been intensively studied, see for instance [8, 104] and the references therein for a systematic exposition of the topic. Two basic notions of the theory are the so-called normalized and unnormalized filtering processes, which are a probability measure-valued process and a positive measure-valued process respectively, and are proved to be the solutions, in a sense that will be clarified later, to stochastic differential equations, called the Kushner-Stratonovich and the Zakai equation respectively.

A classical way to deal with these equations (see, for instance, [87, 94]) is to show that the solutions admit a density with respect to the Lebesgue measure, which possibly belongs to a suitable Hilbert space of functions. Thus, one can study the density processes instead of the measure-valued processes and rely on tools of stochastic calculus on Hilbert spaces to further explore their properties. The price to pay is the introduction of unnecessary assumptions entailing that the filtering processes have a density belonging to a suitably chosen Hilbert space of functions. In this work, we avoid these conditions and rather follow the approach of [8, 16, 75, 80, 99], where the filtering processes are studied as genuine measure-valued processes.

In the literature, some early attempts were made to relate the stochastic filtering problem (and, actually, the optimal control problem with partial observation) to a partial differential equation on the space of probability measures. In particular, the papers [20, 59] carefully studied the corresponding Nisio non-linear semigroup together with its infinitesimal generator. Notice that since at the time a suitable differential calculus on the space of probability measures was still missing, such an infinitesimal generator was only described on a particular set of functions (namely the set of cylinder functions).

Now we present in an informal way the nonlinear filtering equations and then we state our main results. Here we avoid some technical details, and the precise discussion can be found in the next chapters of this Thesis.

In this framework, the Zakai equation reads as

$$\mathrm{d}\langle \rho_t, \psi \rangle = \langle \rho_t, A\psi \rangle \, \mathrm{d}t + \langle \rho_t, h\psi + B\psi \rangle \cdot \mathrm{d}Y_t,$$

where the solution $\rho = \{\rho_t, t \in [0,T]\}$ is a positive measure-valued process, A, B are differential operators defined by the formulas

$$A\psi(x) \coloneqq \sum_{i=1}^d f(x)\partial_i \psi(x) + \frac{1}{2}\sum_{i,j=1}^d (\sigma\sigma^\top)_{ij}(x)\partial_{ij}\psi(x) + (\bar\sigma\bar\sigma^\top)_{ij}(x)\partial_{ij}\psi(x), \quad x \in \mathbb{R}^d,$$

$$B_k\psi(x) \coloneqq \sum_{i=1}^d \bar\sigma_{ik}(x)\partial_i\psi(x), \quad x \in \mathbb{R}^d, k = 1, \dots d,$$

and $f, \sigma, \bar{\sigma}, h$ are functions that have to satisfy some hypotheses we will formulate later. The Kushner-Statonovitch equation reads as

$$\mathrm{d}\langle \Pi_t, \psi \rangle = \langle \Pi_t, A\psi \rangle \,\mathrm{d}t + (\langle \Pi_t, h\psi + B\psi \rangle - \langle \Pi_t, \psi \rangle \langle \Pi_t, h \rangle) \cdot \mathrm{d}I_t,$$

where $\Pi = \{\Pi_t, [0,T]\}$ is probability measure-valued. In the previous equations the processes Y and I are Brownian motions (with respect to appropriate probability measures), and we will explain later in this thesis that they are the so-called observation and innovation processes in the context of nonlinear filtering. The equalities are understood to hold for every ψ in a certain class of test functions, we used the notation $\langle \mu, \psi \rangle = \int \psi(x)\mu(\mathrm{d}x)$. In the following we also denote with $\mathcal{M}^+(\mathbb{R}^d)$ and $\mathcal{P}(\mathbb{R}^d)$ the spaces of positive and probability measures on $\mathbb{R}^d$ respectively, and with $\mathcal{M}_2^+(\mathbb{R}^d)$ and $\mathcal{P}_2(\mathbb{R}^d)$ the spaces of positive and probability measures with finite second moment.

Our main results in this context are two theorems on the existence and uniqueness of classical solutions to the backward Kolmogorov equations, associated to the Zakai and the Kushner-Stratonovich equations, introduced here for the first time. The solutions are functions $u : [0,T] \times \mathcal{M}^+(\mathbb{R}^d) \to \mathbb{R}$ or $u : [0,T] \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ respectively, for which the final condition is a given datum. Despite the theory of backward Kolmogorov equation in finite dimension being a classical and well-known topic (see for instance the chapter by Krylov in [74] and references therein), we note that finding solutions to Kolmogorov equations on infinite-dimensional spaces is a challenging problem and it has been studied intensively, see for instance [74, 41], and the search for classical solutions is often addressed, as in [58]. Most results are only concerned with the Hilbert space case, namely when $u : [0,T] \times H \to \mathbb{R}$, where H is a Hilbert space. The extension to spaces of measure requires entirely different methods and in particular new tools from differential calculus, as we will explain later. We also notice that, in addition to their intrinsic interest, understanding the backward Kolmogorov equations is a preliminary step for tackling the more complicated Hamilton-Jacobi-Bellman equations, which are non-linear equations strongly connected with the stochastic optimal control problem.

Along all the discussion, we will need the following

Assumption 1. *All the mappings $f, \sigma, \bar{\sigma}, h$ are taken Borel-measurable. Moreover we assume:*

a. the mappings $f \colon \mathbb{R}^d \to \mathbb{R}^d$, $\sigma \colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ and $\bar{\sigma} \colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are Lipschitz continuous;

b. the mapping $a := \sigma\sigma^\top \colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is uniformly elliptic, that is there exists $\lambda > 0$ such that $\sum_{i,j=1}^d a_{ij}\xi_i\xi_j \geq \lambda|\xi|^2$ for every $x, \xi \in \mathbb{R}^d$;

c. the mappings $f, \sigma, \bar{\sigma}, h$ are bounded.

The uniform ellipticity in Assumption 1, *b.* is needed for a technical purpose in the proof of Proposition 4.1.4. Indeed, we need the existence and uniqueness of a classical solution for a Kolmogorov equation over $\mathbb{R}^d$, with differential operator A. If one can obtain this result with different conditions, then a can be only non-degenerating, to have the Markov property for the solutions of the filtering equations. See Remark 1.3.12 for a detailed discussion.

The first result, given in Theorem 4.2.5, concerns the backward Kolmogorov equation associated to the Zakai equation, that reads as

$$\begin{cases} \partial_s u(\mu, s) + \mathcal{L}u(\mu, s) = 0, & (\mu, s) \in \mathcal{M}_2^+(\mathbb{R}^d) \times [0, T], \\ u(\mu, T) = \Phi(\mu), & \mu \in \mathcal{M}_2^+(\mathbb{R}^d), \end{cases} \tag{I.1}$$

where

$$\begin{aligned} \mathcal{L}u(\mu) = \langle \mu, \mathrm{D}_\mu u(\mu) \cdot f \rangle + \frac{1}{2} \langle \mu, \operatorname{tr} \left\{ \mathrm{D}_x \mathrm{D}_\mu u(\mu) \sigma \sigma^\top \right\} \rangle \\ + \frac{1}{2} \langle \mu, \operatorname{tr} \left\{ \mathrm{D}_x \mathrm{D}_\mu u(\mu) \bar{\sigma} \bar{\sigma}^\top \right\} \rangle + \frac{1}{2} \langle \mu \otimes \mu, \delta_\mu^2 u(\mu) h \cdot h \rangle \\ + \langle \mu \otimes \mu, h \cdot \bar{\sigma}^\top \delta_\mu \mathrm{D}_\mu u(\mu) \rangle + \frac{1}{2} \langle \mu \otimes \mu, \operatorname{tr} \left\{ \mathrm{D}_\mu^2 u(\mu) \bar{\sigma} \bar{\sigma}^\top \right\} \rangle. \end{aligned} \tag{I.2}$$

In (I.2), $\delta_\mu u, \delta_\mu^2 u, \mathrm{D}_\mu u, \mathrm{D}_\mu^2 u$ are notions of first and second-order derivatives on $\mathcal{M}^+(\mathbb{R}^d)$ we will discuss later, whilst D_x denotes the gradient on $\mathbb{R}^d$. In Theorem 4.2.5 we show that if the terminal condition $\Phi \in \mathrm{C}^2_{\mathrm{L}}(\mathcal{M}_2^+(\mathbb{R}^d))$, which means that it admits all the derivatives listed above, then there exists a unique classical solution to (I.1). The theorem is the following:

Theorem 4.2.5. *Let $\rho^{s,\mu}$ be the solution to the Zakai equation starting at time s from $\mu \in \mathcal{M}_2^+(\mathbb{R}^d)$, let $\Phi \in \mathrm{C}^2_{\mathrm{L}}(\mathcal{M}_2^+(\mathbb{R}^d))$ and let Assumption 1 holds. Then the function*

$$u(\mu, s) = \mathbb{E}\left[\Phi(\rho_T^{s,\mu})\right]$$

is the unique classical solution to the backward Kolmogorov equation (I.1).

Analogously, in Theorem 4.3.3 we prove existence and uniqueness for classical solutions to the backward Kolmogorov equation on $\mathcal{P}_2(\mathbb{R}^d)$ associated to the Kushner-Stratonovich equation, that is

$$\begin{cases} \partial_s u(\pi, s) + \mathcal{L}^{KS}u(\pi, s) = 0, & (\pi, s) \in \mathcal{P}(\mathbb{R}^d) \times [0, T], \\ u(\pi, T) = \Phi(\pi), & \pi \in \mathcal{P}(\mathbb{R}^d), \end{cases} \tag{I.3}$$

where

$$\begin{aligned} &\mathcal{L}^{KS}u(\pi) \\ &= \langle \pi, f \cdot \mathrm{D}_\pi u(\pi) \rangle + \frac{1}{2} \langle \pi, \operatorname{tr}\{\mathrm{D}_x \mathrm{D}_\pi u(\pi) \sigma \sigma^\top\} \rangle + \frac{1}{2} \langle \pi, \operatorname{tr}\{\mathrm{D}_x \mathrm{D}_\pi u(\pi) \bar{\sigma} \bar{\sigma}^\top\} \rangle \\ &+ \frac{1}{2} \langle \pi \otimes \pi, \delta_\pi^2 u(\pi) h \cdot h \rangle + \frac{1}{2} \langle \pi \otimes \pi, \operatorname{tr}\{\mathrm{D}_\pi^2 u(\pi) \bar{\sigma} \bar{\sigma}^\top\} \rangle + \frac{1}{2} \langle \pi \otimes \pi, \delta_\pi^2 u(\pi) \rangle |\langle \pi, h \rangle|^2 \\ &+ \langle \pi \otimes \pi, h \cdot \bar{\sigma}^\top \delta_\pi \mathrm{D}_\pi u(\pi) \rangle - \langle \pi \otimes \pi, \delta_\pi^2 u(\pi) h \rangle \cdot \langle \pi, h \rangle - \langle \pi \otimes \pi, \bar{\sigma}^\top \delta_\pi \mathrm{D}_\pi u(\pi) \rangle \cdot \langle \pi, h \rangle. \end{aligned} \tag{I.4}$$

Theorem 4.3.3. *Let $\Pi^{s,\pi}$ be the solution to the Kushner-Stratonovich equation starting at time s from $\pi \in \mathcal{P}_2(\mathbb{R}^d)$, let $\Phi \in \mathrm{C}^2_{\mathrm{L}}(\mathcal{P}_2(\mathbb{R}^d))$ and let Assumption 1 holds. Then the function*

$$u(\pi, s) = \mathbb{E}\left[\Phi(\Pi^{s;\pi}_T)\right]$$

is the unique classical solution to the backward Kolmogorov equation (I.3).

We notice that in both cases, the solution is given by the usual probabilistic representation formula.

In the study of (I.1) we have to face a specific difficulty: since we deal with functions over $\mathcal{M}^+(\mathbb{R}^d)$ we can not rely on the various notions in differential calculus that have been developed in the last years with reference to the space of probability measures (see [5, 31, 63, 79] for different notions of derivative and a comparison among them). For instance, the technique introduced by P.-L. Lions in [79] where the problem is lifted on a space of random variables is no longer available. However, it turns out that the notion of linear functional derivative, given in [31], can be extended to our framework. More precisely, we say $\delta_\mu u \colon \mathcal{M}^+(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ is the derivative of $u \colon \mathcal{M}^+(\mathbb{R}^d) \to \mathbb{R}$ in linear functional sense if u and $\delta_\mu u$ have some regularity properties and if for every $\mu, \mu' \in \mathcal{M}^+(\mathbb{R}^d)$ it holds

$$u(\mu') - u(\mu) = \int_0^1 \int_{\mathbb{R}^d} \delta_\mu u\,(t\mu' + (1-t)\mu, x)\,[\mu' - \mu](\,\mathrm{d}x)\,\mathrm{d}t.$$

By iterating the definition, one can introduce a notion of second-order linear functional derivative $\delta^2_\mu u \colon \mathcal{M}^+(\mathbb{R}^d) \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. Once $\delta_\mu u$ has been introduced, one can set

$$\mathrm{D}_\mu u(\mu, x) \coloneqq \mathrm{D}_x \delta_\mu u(\mu, x), \quad (\mu, x) \in \mathcal{M}^+(\mathbb{R}^d) \times \mathbb{R}^d.$$

If we restrict ourselves to the space of probability measures with finite second moment, under certain hypotheses this last definition coincides with the notion of derivative introduced by Lions, which we will call L-derivative. Moreover, we can define a second-order L-derivative by $\mathrm{D}^2_\mu u(\mu, x, y) := \mathrm{D}_x \mathrm{D}^\top_y \delta^2_\mu u(\mu, x, y) \in \mathbb{R}^{d\times d}$. Thus, we can introduce the space $\mathrm{C}^2_{\mathrm{L}}(\mathcal{M}^+(\mathbb{R}^d))$ as the set of real-valued functions over $\mathcal{M}^+(\mathbb{R}^d)$ which are twice differentiable in linear functional and L sense, and with regular derivatives. Most of the Itô formulas available in the literature involve only L-derivatives, whilst in our case, both linear functional and L-derivatives are needed.

We point out that in the literature other notions of derivative for functionals over sets of positive measures have been introduced. For instance, a derivative over $\mathcal{M}^+(\mathbb{R}^d)$ has been introduced in the framework of measure-valued processes related to particle systems (see for instance [42] and [71]) and it has been intensively used in the context of Fleming-Viot processes. It turns out that, under certain conditions, this notion coincides with the one adopted in this paper (see [93]). Another example is [2], where the authors give a definition of derivative for functionals over Poisson spaces.

As expected, to show the uniqueness property in (I.1) and (I.3) one needs to prove a suitable Itô formula, in our case for the composition of a real-valued function and a measure-valued process. In the recent literature, formulas of this kind have been proved when the process takes values in a space of probability measures, see for instance [30, 32] and it is constructed as the time evolution of the one-dimensional law (in certain situations the one-dimensional conditional law) of a given finite-dimensional process. For our purposes, we need very different results. In Proposition 3.1.1 we provide an Itô formula for the composition of a real-valued function over $\mathcal{M}^+(\mathbb{R}^d)$ and the $\mathcal{M}^+(\mathbb{R}^d)$-valued process solution to the Zakai equation.

Proposition 3.1.1. *Let $\rho = \{\rho_t, t \in [0,T]\}$ be a solution to the Zakai equation starting at $\mu \in \mathcal{M}_2^+(\mathbb{R}^d)$ and let u be in $\mathrm{C}^2_{\mathrm{L}}(\mathcal{M}_2^+(\mathbb{R}^d))$. Let also Assumption 1 holds. Then the following Itô formula holds:*

$$\begin{aligned} u(\rho_t) = u(\mu) &+ \int_0^t \langle \rho_s, \mathrm{D}_\mu u(\rho_s)\cdot f\rangle \,\mathrm{d}s \\ &+ \int_0^t \frac{1}{2}\langle \rho_s, \mathrm{tr}\left\{\mathrm{D}_x\mathrm{D}_\mu u(\rho_s)\sigma\sigma^\top\right\}\rangle \,\mathrm{d}s + \int_0^t \frac{1}{2}\langle \rho_t, \mathrm{tr}\left\{\mathrm{D}_x\mathrm{D}_\mu u(\rho_s)\bar\sigma\bar\sigma^\top\right\}\rangle \,\mathrm{d}s \\ &+ \int_0^t \langle \rho_s, h\delta_\mu u(\rho_s)\rangle \cdot \mathrm{d}Y_s + \int_0^t \langle \rho_s, \bar\sigma^\top \mathrm{D}_\mu u(\rho_s)\rangle \cdot \mathrm{d}Y_s \\ &+ \int_0^t \frac{1}{2}\langle \rho_s\otimes\rho_s, \delta^2_\mu u(\rho_s)h\cdot h\rangle \,\mathrm{d}s + \int_0^t \langle \rho_s\otimes\rho_s, h\cdot\bar\sigma^\top\delta_\mu\mathrm{D}_\mu u(\rho_s)\rangle \,\mathrm{d}s \\ &+ \int_0^t \frac{1}{2}\langle \rho_s\otimes\rho_s, \mathrm{tr}\left\{\mathrm{D}^2_\mu u(\rho_s)\bar\sigma\bar\sigma^\top\right\}\rangle \,\mathrm{d}s, \quad t\in[0,T]. \end{aligned}$$

Similarly, in Proposition 3.2.1 we prove an Itô formula related to the $\mathcal{P}(\mathbb{R}^d)$-valued process solution to the Kushner-Stratonovich equation.

Proposition 3.2.1. *Let $\Pi = \{\Pi, t \in [0,T]\}$ be a solution to the Kushner-Stratonovich equation starting from $\pi \in \mathcal{P}_2(\mathbb{R}^d)$ and let $u \in \mathrm{C}^2_{\mathrm{L}}(\mathcal{P}_2(\mathbb{R}^d))$. Moreover, let Assumption 1 be in force. Then for every $t\in[0,T]$ it holds:*

$$\begin{aligned} u(\Pi_t) = u(\pi) &+ \int_0^t \langle \Pi_s, \mathrm{D}_\mu u(\Pi_s)\cdot f\rangle \,\mathrm{d}s \\ &+ \int_0^t \frac{1}{2}\langle \Pi_s, \mathrm{tr}\left\{\mathrm{D}_x\mathrm{D}_\mu u(\Pi_s)\sigma\sigma^\top\right\}\rangle \,\mathrm{d}s + \int_0^t \frac{1}{2}\langle \Pi_t, \mathrm{tr}\left\{\mathrm{D}_x\mathrm{D}_\mu u(\Pi_s)\bar\sigma\bar\sigma^\top\right\}\rangle \,\mathrm{d}s \\ &+ \int_0^t \frac{1}{2}\langle \Pi_s\otimes\Pi_s, \delta^2_\mu u(\Pi_s)h\cdot h\rangle \,\mathrm{d}s + \int_0^t \frac{1}{2}\langle \Pi_s\otimes\Pi_s, \mathrm{tr}\left\{\mathrm{D}^2_\mu u(\Pi_s)\bar\sigma\bar\sigma^\top\right\}\rangle \,\mathrm{d}s \\ &+ \int_0^t \frac{1}{2}|\langle \Pi_s, h\rangle|^2\langle \Pi_s\otimes\Pi_s, \delta^2_\mu u(\Pi_s)\rangle \,\mathrm{d}s + \int_0^t \langle \Pi_s\otimes\Pi_s, h\cdot\bar\sigma^\top\delta_\mu\mathrm{D}_\mu u(\Pi_s)\rangle \,\mathrm{d}s \\ &- \int_0^t \langle \Pi_s\otimes\Pi_s, \delta^2_\mu u(\Pi_s)h\rangle\cdot\langle\Pi_s, h\rangle \,\mathrm{d}s - \int_0^t \langle \Pi_s\otimes\Pi_s, \bar\sigma^\top\delta_\mu\mathrm{D}_\mu u(\Pi_s)\rangle\cdot\langle\Pi_s, h\rangle \,\mathrm{d}s \\ &+ \int_0^t \langle \Pi_s, h\delta_\mu u(\Pi_s)\rangle\cdot \mathrm{d}I_s + \int_0^t \langle \Pi_s, \bar\sigma^\top\mathrm{D}_\mu u(\Pi_s)\rangle\cdot \mathrm{d}I_s - \int_0^t \langle \Pi_s, \delta_\mu u(\Pi_s)\rangle\langle\Pi_s, h\rangle\cdot \mathrm{d}I_s. \end{aligned}$$

Both results are new, but the latter can be viewed as a generalization of the one obtained in the context of mean field games with common noise ([32, Section 4.3]), as explained with more detail in Remark 3.2.4. One major technical difference from existing cases is the fact that the Zakai and Kushner-Stratonovich equations are understood to hold in a weak form, namely for arbitrary choice of the occurring test function. In our proof we first show the formula for a smaller class of functions with good properties by exploiting the classical Itô formula, then we obtain the general result by an approximation argument. In order to pass to the limit in the Itô formula one needs the convergence of the approximating functions as well as their first and second derivatives (linear functional and L-derivatives). The required constructions have some interest in themselves and can be used again in similar contexts (see Remark 3.2.5).

Concerning the existence of classical solutions to (I.1) and (I.3), the most difficult part of the proof is the investigation of the regularity of the solutions to the Zakai and the Kushner-Stratonovich equations with respect to the initial datum. The dependence of the filtering processes on the initial condition has been the object of intense study since it is related to the problem of assessing the effect of a misspecification of the initial distribution of the signal process in the filtering problem. However, the study of the differentiability properties of the solution with respect to the initial conditions seems to be addressed here for the first time. As this relates to the differentiability of measure-valued processes with respect to a measure (the starting point of the process itself) we need to introduce novel notions of differentiability for mappings from $\mathcal{M}^+(\mathbb{R}^d)$ (or $\mathcal{P}(\mathbb{R}^d)$) to $\mathcal{M}^+(\mathbb{R}^d)$(or $\mathcal{P}(\mathbb{R}^d)$). We point out that a similar definition has been recently introduced and studied in [83].

These results lay the foundations for the study of partial differential equations on spaces of measures associated to stochastic filtering problems. A natural prosecution will be to consider non-linear partial differential equations, in particular, the Hamilton-Jacobi-Bellman equation arising from the optimal control problem with partial observation problem will be investigated, see for instance the book [13] for a systematic introduction to the problem or the recent paper [14] for a modern approach in the density case, based on mean field techniques. This problem has been already tackled for the Zakai equation in [9, 10] exploiting the randomization method and BSDEs techniques, but in the more restrictive case where the function h is identically equal to zero. A look at (I.1) shows that this assumption allows the authors to rely only on L-derivatives and exploit previous results on the well-posedness of related partial differential equations, an approach that is not possible in our situation.

I.2 Kolmogorov equations on the space of probability measures associated to the nonlinear filtering equation: viscosity solutions

The theory of viscosity solutions is well-established in the finite-dimensional case (see for instance [37]) and in the infinite-dimensional Hilbert case (see for instance [98, 57]). Recently some developments have been obtained for more complicated spaces, where it is not easy to provide a notion of derivative. Two examples are the space of continuous paths (see for instance [36] and references therein) and the space of probability measures, to which the present work refers. In particular, in [12, 91, 92, 34, 10] the existence of viscosity solutions to second-order equations on the space of probability measures is addressed. On the other hand, uniqueness results are very hard to achieve, and only few papers are available for second-order equations ([27, 103, 35, 95]). As said in the previous section, the study of partial differential equations on spaces of probability measures is a topic of increasing research interest. For instance, in [4, 24, 63, 12, 92, 29] and references therein one can find many results both for the first and second-order case.

Our work gives a contribution in this direction. We study a class of backward Kolmogorov equations on spaces of probability measures, arising in the context of stochastic filtering. More precisely, we consider the backward Kolmogorov equation associated to the Kushner-Stratonovich stochastic differential equation (see Section 4.3), which is a second-order linear partial differential equation of parabolic type on the space of probability measures. Our goal is to continue the analysis started in [81] (here presented mainly in Chapter 4 and outlined in the previous paragraph), by discussing the case of a less regular final condition, namely just continuous and bounded. In this case, the notion of classical solution turns out to be too strong, thus a weaker notion is necessary. In particular, the notion of viscosity solution seems to be the natural one for these kinds of problems. Thus, our efforts will be concentrated on proving that there exists a unique viscosity solution.

In (I.4) the Kolmogorov operator associated to the Kushner-Stratonovich has been introduced:

$$\begin{aligned}
&\mathcal{L}^{KS}u(\pi) \\
&\quad= \langle \pi, f \cdot \mathrm{D}_\pi u(\pi)\rangle + \frac{1}{2}\langle \pi, \mathrm{tr}\{\mathrm{D}_x\mathrm{D}_\pi u(\pi)\sigma\sigma^\top\}\rangle + \frac{1}{2}\langle \pi, \mathrm{tr}\{\mathrm{D}_x\mathrm{D}_\pi u(\pi)\bar\sigma\bar\sigma^\top\}\rangle \\
&\quad+ \frac{1}{2}\langle \pi\otimes\pi, \delta^2_\pi u(\pi)h\cdot h\rangle + \frac{1}{2}\langle \pi\otimes\pi, \mathrm{tr}\{\mathrm{D}^2_\pi u(\pi)\bar\sigma\bar\sigma^\top\}\rangle + \frac{1}{2}\langle \pi\otimes\pi, \delta^2_\pi u(\pi)\rangle|\langle\pi,h\rangle|^2 \\
&\quad+ \langle \pi\otimes\pi, h\cdot\bar\sigma^\top\delta_\pi\mathrm{D}_\pi u(\pi)\rangle - \langle \pi\otimes\pi, \delta^2_\pi u(\Pi)h\rangle\cdot\langle\pi,h\rangle - \langle \pi\otimes\pi, \bar\sigma^\top\delta_\pi\mathrm{D}_\pi u(\pi)\rangle\cdot\langle\pi,h\rangle,
\end{aligned}$$

where $\delta_\pi u, \delta^2_\pi u, \mathrm{D}_\pi u, \mathrm{D}^2_\pi u$ are suitably defined derivatives of real-valued functions over $\mathcal{P}(\mathbb{R}^d)$, as shortly explained above (see Section I.1 for an informal definition or Section 2.1 for the precise one). Given a final condition $\Phi\colon \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$, the backward Kolmogorov equation reads as

$$\begin{cases}\partial_t u(\pi,t) + \mathcal{L}^{KS}u(\pi,t) = 0, & (\pi,t)\in\mathcal{P}_2(\mathbb{R}^d)\times[0,T),\\ u(\pi,T) = \Phi(\pi), & \pi\in\mathcal{P}_2(\mathbb{R}^d).\end{cases} \tag{I.5}$$

As reminded in Section I.1, in [81] it has been proved that if Φ is twice continuously differentiable (namely in $\mathrm{C}^2_\mathrm{L}(\mathcal{P}_2(\mathbb{R}^\mathrm{d}))$), then there exists a unique classical solution to (I.5) given by

$$u(\pi,s) = \mathbb{E}\left[\Phi(\Pi^{s,\pi}_T)\right], \tag{I.6}$$

where $\Pi^{s,\pi}$ is the solution to the Kushner-Stratonovich equation starting at time s from $\pi\in\mathcal{P}_2(\mathbb{R}^d)$. We recall that this result holds under the following assumption.

Assumption 1. *All the mappings $f, \sigma, \bar{\sigma}, h$ are taken Borel-measurable. Moreover we assume:*

a. the mappings $f\colon \mathbb{R}^d \to \mathbb{R}^d$, $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ and $\bar{\sigma}\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are Lipschitz continuous;

b. the mapping $a := \sigma\sigma^\top\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is uniformly elliptic, that is there exists $\lambda > 0$ such that $\sum_{i,j=1}^d a_{ij}\xi_i\xi_j \geq \lambda|\xi|^2$ for every $x, \xi \in \mathbb{R}^d$;

c. the mappings $f, \sigma, \bar{\sigma}, h$ are bounded.

As said before, since we want to reduce the regularity of the final condition Φ, we need to weaken the notion of solution and consider the viscosity one. This change introduces some extra difficulties and forces us to use a completely new approach. In particular, in this case, the main obstacle is the so-called comparison principle, which is a key step to prove the uniqueness property of a solution. To achieve this result, we need to restrict ourselves to the compact case, by looking at (I.5) over $\mathcal{P}(K) \subset \mathcal{P}_2(\mathbb{R}^d)$, with $K \subset \mathbb{R}^d$ compact. Thus, we need to ask for the following invariance property.

Assumption 3. *The probability measure-valued process Π has trajectories confined in the subset $\mathcal{P}(K)$, $K \subset \mathbb{R}^d$ compact.*

This is for instance the case when the filtering equation is associated to a signal with trajectories confined on a compact subset or on a compact manifold like a torus.

In this context, we are able to give our existence and uniqueness result. In Proposition 5.2.2, we show that under Assumption 1, the function u defined by (I.6) is a viscosity solution to (I.5). Then, by adding Assumption 3, we can also prove a uniqueness result via comparison principle.

Theorem 5.2.4. *Let $v_1, v_2\colon \mathcal{P}(K) \times [0,T] \to \mathbb{R}$ be respectively a viscosity subsolution and supersolution of* (5.1). *Moreover, let Assumption 1 and Assumption 3 hold. Then, $v_1(\pi,t) \leq v_2(\pi,t)$ for every $(\pi,t) \in \mathcal{P}(K) \times [0,T]$.*

Corollary 5.2.5. *Let $\Phi \in \mathrm{C}(\mathcal{P}(K))$ and let Assumption 1 and Assumption 3 hold. Then*

$$u(\pi,t) = \mathbb{E}\left[\Phi(\Pi_T^{\pi,t})\right], \quad (\pi,t) \in \mathcal{P}(K) \times [0,T],$$

is the unique viscosity solution to the backward Kolmogorov equation (5.1).

The technique used to prove the comparison principle is based on the fact that we have a candidate solution (given by a probabilistic representation formula of the usual type) that can be used to obtain partial comparison principles. A remarkable feature of this approach is to avoid the lifting of the problem on a Hilbert space, which is a procedure often used to solve these types of second-order problems. In that case, for viscosity solutions, it is not yet clear what is the relation between the lifted problem and the original one. We also point out that, differently from the classical solution case, the techniques we used can not be immediately extended to the case of positive measures with compact support, and so different ideas should be necessary to study the Kolmogorov equation associated to the Zakai equation.

I.3 Mean field optimal stopping: a limit approach

Optimal stopping problems (OSP) with cost depending on the mean of the stopped process arise for instance when the goal is to minimize the variance. In the work by Pedersen and Peskir [88, 89]), the problems of optimal variance stopping and optimal mean-variance stopping have been investigated in the case where the underlying process is for instance a Geometric Brownian motion, highlighting connections with portfolio choice. Even though the OSP of the variance is *time-inconsistent*, which means that the value process is not a supermartingale, they succeeded to solve the problem. In particular, they derive a variational inequality for the value function with an explicit stopping region. More recently, the interest in optimal stopping problems with a more general mean field type interaction such as the dependence on the law of the stopped process increased significantly, due mainly to the connection with the theory of mean field games and mean field optimal control (see e.g. [31] for a systematic presentation of the topic). The first contribution in this direction is the work by Bertucci [15] where an optimal stopping problem for a mean field game is studied using mainly PDE techniques. Then, other results and extensions have been obtained with different techniques in [33, 85, 23, 47] and in the recent papers [100, 48, 1].

A powerful tool to study optimal stopping problems is based on the Snell envelope of processes (see [17, 53] and Appendix D in [68]). Given the general OSP

$$Y_0 = \sup_{\tau \in \mathcal{T}_0} \mathbb{E}\left[L_\tau\right]$$

where the reward (a.k.a. the barrier or obstacle) process $L = \{L_t, t \in [0,T]\}$ is right continuous with left limits (càdlàg) and adapted to a filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\in[0,T]}$ which satisfies the usual conditions, the Snell envelope of the process L is defined by

$$Y_t = \operatorname*{ess\,sup}_{\tau \in \mathcal{T}_t} \mathbb{E}\left[L_\tau \mid \mathcal{F}_t\right],$$

where $\mathcal{T}_t$ is set of $\mathbb{F}$-stopping times with values in $[t,T]$. Under mild uniform integrability conditions on L, it turns out that Y is the smallest supermartingale that dominates L, an important property which implies that, when L has only nonnegative jumps, then Y is continuous and it is optimal to stop when Y hits L. More precisely, this means that the hitting time

$$\tau_t^* = \inf\{s \geq t,\ Y_s = L_s\} \wedge T \tag{I.7}$$

is optimal after t. In particular, $\tau^* := \tau_0^*$ is optimal for Y_0.

In the recent papers [44, 43], this result could be successfully applied to a large class of mean field OSPs whose value-process solves a mean field reflected BSDEs i.e. it is a supermartingale. That class of OSPs includes the following recursive OSP (we ignore the integral running cost term)

$$Y_0 = \sup_{\tau \in \mathcal{T}_0} \mathbb{E}\left[h(Y_\tau, \mathbb{P}_{Y_s}|_{s=\tau})\mathbb{1}_{\{\tau<T\}} + \xi\mathbb{1}_{\{\tau=T\}}\right],$$

where the mean field coupling is in terms of the marginal law of Y *evaluated* at a stopping time. We remark that it is not the law of the random variable Y_τ. Another example of a mean field OSP for which the supermartingale property is preserved is considered in the recent work by Talbi, Touzi and Zhang [100] where the mean field OSP for a mean field diffusion is studied in a weak (or relaxed) formulation, that means in terms of the joint marginal law of the stopped underlying process X and the survival process I associated with the stopping time. Moreover, the performance function is a deterministic function of the marginal laws of (X, I). They characterized the value function by a dynamic programming equation on the Wasserstein space.

Going back to the OSP of the variance problem, it can be seen as an OSP where the stopped obstacle is of the form $L_\tau = (X_\tau - \mathbb{E}[X_\tau])^2$. In this case, the results in [44, 43] do not apply, whilst the ones in [100] can solve the associated relaxed problem. Nevertheless, Pedersen and Peskir [88, 89]) could solve the OSP of the variance of an underlying Markov diffusion process by embedding it into an auxiliary standard OSP whose value function solves a standard variational inequality.

We consider the following class of time-inconsistent mean field OSPs beyond the mean-variance case. Let $T > 0$ be a finite time horizon, W a Brownian motion defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathbb{F} = \{\mathcal{F}_t\}_{t\in[0,T]}$ the complete Brownian filtration. On the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ let X be a diffusion process of mean field type:

$$X_t = X_0 + \int_0^t b(s, X_s, \mathbb{P}_{X_s})ds + \int_0^t \sigma(s, X_s, \mathbb{P}_{X_s})dW_s,$$

where $\mathbb{P}_{X_s}$ denotes the law of the random variable X_s. For a certain $\mathcal{F}_T$-measurable final condition ξ and a performance function h, we consider the following OSPs:

(OSPa) Optimal stopping of a mean field diffusion:

$$Y_0 = \sup_{\tau\in\mathcal{T}_0} \mathbb{E}\left[h(X_\tau, \mathbb{P}_{X_\tau})\mathbb{1}_{\{\tau<T\}} + \xi\mathbb{1}_{\{\tau=T\}}\right]; \tag{I.8}$$

(OSPb) Optimal stopping of a recursive utility function defined on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$:

$$Y_0 = \sup_{\tau\in\mathcal{T}_0} \mathbb{E}\left[h(Y_\tau, \mathbb{P}_{Y_\tau})\mathbb{1}_{\{\tau<T\}} + \xi\mathbb{1}_{\{\tau=T\}}\right]. \tag{I.9}$$

Our goal is to show under certain conditions that an optimal stopping time for each of the OSPs (I.8) and (I.9) can be charaterized as a hitting time similar to τ^* in (I.7). A straightforward extension is to consider the combination of (I.8) and (I.9) given by

$$Y_0 = \sup_{\tau\in\mathcal{T}_0} \mathbb{E}\left[h(X_\tau, \mathbb{P}_{X_\tau}, Y_\tau, \mathbb{P}_{Y_\tau})\mathbb{1}_{\{\tau<T\}} + \xi\mathbb{1}_{\{\tau=T\}}\right]. \tag{I.10}$$

To solve the above problems we use a limit approach which consists of introducing a family of interacting Snell envelopes $\{Y^{i,n}\}_{i=1}^n$ (see Section 6.1 for a precise definition for (OSPb) and Section 6.3 for (OSPa)) as approximation of the value-process of the mean field OSP. For example, we approximate the OSP (I.9) with the following family of interacting OSPs:

$$Y_0^{i,n} = \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_0^i} \mathbb{E}\left[h(Y_\tau^{i,n}, \frac{1}{n}\sum_{j=1}^n \delta_{Y_\tau^{j,n}})\mathbb{1}_{\{\tau<T\}} + \xi^i\mathbb{1}_{\{\tau=T\}}\right], \quad i = 1, 2, \ldots, n.$$

These problems are time-consistent and it can be shown (see Corollary 6.1.2) that it is optimal to stop at the hitting time $\hat{\tau}^{i,n}$ at which the value-process (which is now a Snell envelope) $(Y_t^{i,n})_{t\geq 0}$ hits the barrier $(\mathbb{E}[h(Y_t^{i,n}, \frac{1}{n}\sum_{j=1}^n \delta_{Y_t^{j,n}}) \,|\, \mathcal{F}_t^i])_{t\in[0,T]}$.

In Theorem 6.2.6 we prove that the stopping time

$$\tau^* = \inf\{t \geq 0,\ Y_t = h(Y_t, \mathbb{P}_{Y_t})\} \wedge T \tag{I.11}$$

is optimal for Y_0 given by (I.9), by showing that it is the limit in probability of $\hat{\tau}^{1,n}$ as $n \to \infty$. Thus, in this time-inconsistent framework an optimal stopping is also given by the usual hitting time. More precisely, under

Assumption 5. *The coefficients h and $\{\xi^i\}_{i\geq 1}$ satisfy*

(i) for each $i \geq 1$, $\xi^i \in L^2(\mathcal{F}_T^i)$. Moreover, the ξ^i's are independent copies of ξ, with $\xi^1 = \xi$;

(ii) $h\colon \Omega \times \mathbb{R} \times \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}$

(ii-a) is Lipschitz with respect to (y, μ), that is there exist two positive constants γ_1 and γ_2 such that

$$|h(\omega, y_1, \mu_1) - h(\omega, y_2, \mu_2)| \leq \gamma_1|y_1 - y_2| + \gamma_2\mathcal{W}_2(\mu_1, \mu_2),$$

for any $\omega \in \Omega$, any $y_1, y_2 \in \mathbb{R}$ and any $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R})$.

(ii-b) $h(\cdot, 0, \delta_0) \in L^2(\mathbb{P})$.

we obtain the following:

Theorem 6.2.6. *Let Assumption 5 holds and assume that γ_1 and γ_2 satisfy*

$$16(\gamma_1^2 + \gamma_2^2) < 1. \tag{I.12}$$

Then, the stopping time τ^* defined by (I.11) *is optimal for the OSP* (I.9).

To derive this result, we need to investigate the convergence of $\{Y_0^{i,n}\}_{n\geq 1}$ to Y_0 (Proposition 6.2.3) and the convergence of the associated optimal stopping times (Lemma 6.2.8).

To have a more clear exposition, we prove our results in the simplified case of dependence on the mean of the stopped process, and then in Subsection 6.2.3 we discuss the general case. Moreover, in Section 6.3 we discuss how to apply the suggested techniques to the problem (OSPa) associated to a mean field diffusion process and in Section 6.3.1 we discuss the OSP of the variance problem for Markov diffusion processes, we mentioned above.

Finally, we also point out that this class of OSPs should be studied using randomization techniques. In particular, one can consider the OSP where the control is chosen in a wider set, namely the set of randomized stopping times, which is compact in the Baxter-Chacon topology (see [11]). Then, following many papers including Edgar, Millet and Sucheston [52], Arenas [6] and El Karoui, Lepeltier and Millet [54], one can show that there exists an optimal randomized stopping time for Y_0, without any further characterization compared to the explicit optimal stopping time (I.11).

Notation and preliminaries

We collect here some recurrent notations and facts we will use during all our discussion.

Let $x, y \in \mathbb{R}^d$ and let $u\colon \mathbb{R}^d \to \mathbb{R}$:

- $|x|$: euclidian norm of x and so its absolute value if $d = 1$;
- $\mathrm{D}_x u$: gradient of u, seen as a column vector;
- $x \cdot y = x^\top y$: scalar product in $\mathbb{R}^d$.

Let X be a topological space, let $\mathcal{B}(X)$ be its Borel sigma algebra and the and let $E \subseteq \mathbb{R}^d$. We introduce the following spaces of functions:

- C(X): continuous real-valued functions over X;
- $\mathrm{C_b}(X)$: continuous and bounded real-valued functions over X;
- $\mathrm{C_c}(X)$: continuous real-valued functions over X with compact support;
- $\mathrm{C}(X;\mathbb{R}^d), \mathrm{C_b}(X;\mathbb{R}^d), \mathrm{C_c}(X;\mathbb{R}^d)$: as above but with $\mathbb{R}^d$-valued functions;
- $\mathrm{B_b}(X)$: Borel measurable and bounded functions over X;
- $\|u\|_\infty := \sup_{x\in X}|u(x)|$, with $u\colon X \to \mathbb{R}^d$, $d \geq 1$; if u depends on several arguments, in $\|u\|_\infty$ the supremum is taken over all of them;
- $\mathrm{C}^k(E)$, $k \in \mathbb{N}$: real-valued functions over $\mathbb{R}^d$ which are k-times continuously differentiable ($k = 0$ denotes the space of continuous functions); if E is closed, we say that a function is continuously differentiable over E if it has bounded and uniformly continuous derivative over $\mathrm{Int}\,(E)$;
- $\mathrm{C}^k_\mathrm{b}(E)$, $k \in \mathbb{N}$: real-valued functions in $\mathrm{C}^k(E)$ bounded and with bounded derivatives up to order k;
- $\mathrm{C}^k_\mathrm{c}(E)$, $k \in \mathbb{N}$: real-valued functions over $\mathbb{R}^d$ which are k-times continuously differentiable and with compact support;
- $\mathrm{C}^\infty(E), \mathrm{C}^\infty_\mathrm{b}(E), \mathrm{C}^\infty_\mathrm{c}(E)$: as above but with functions that have continuous derivatives of any order;
- $\mathrm{C}^k(E;\mathbb{R}^d), \mathrm{C_b}^k(E;\mathbb{R}^d), \mathrm{C_c}^k(E;\mathbb{R}^d)$: as above but with $\mathbb{R}^d$-valued functions;
- $\|u\|_{\mathrm{C}^k} := \|u\|_\infty + \sum_{i=1}^k \|\mathrm{D}^i_x u\|_\infty$ for any $u \in \mathrm{C}^k(\mathbb{R}^d)$;
- $\mathbf{1}$: is the function identically equal to 1, that is $\mathbf{1}(x) = 1$ for every x;
- $\mathbb{1}_A$: in the indicator function of the set A.

Let E be a subset of $\mathbb{R}^d$ with the Borel sigma algebra $\mathcal{B}(E)$.

- $\mathcal{M}(E)$: signed measures over E with finite total variation, that is $|\mu|(E) < \infty$ with $|\mu| := \mu^+ + \mu^-$;
- $\mathcal{M}^+(E)$: positive measures over E;
- $\mathcal{P}(E)$: probability measures over E;
- $\mathcal{M}_p^+(E)$, $p \in [1, +\infty)$: positive measures with finite p-th moment, that is

$$\mathcal{M}_p^+(E) := \left\{ \mu \in \mathcal{M}^+(E) \colon \int_E |x|^p \mu(\,\mathrm{d}x) < \infty \right\};$$

- $\mathcal{P}_p(E)$, $p \in [1, +\infty)$: probability measures with finite p-th moment;
- $\mu(\psi) = \langle \mu, \psi \rangle = \int_E \psi(x)\mu(\,\mathrm{d}x)$, for a measure μ and an integrable function ψ;
- $\mu(B)$, $B \in \mathcal{B}(E)$: measure of B;
- $\langle \mu, \mathbf{1} \rangle = \mu(E)$.

All the previous definitions extends analogously when E is a complete separable metric space endowed with its Borel sigma algebra.

We say that a sequence of positive or probability measures $\{\mu_n\}_{n \geq 1}$ converges weakly to a measure μ if $\mu_n(\psi) \to \mu(\psi)$ for any $\psi \in \mathrm{C_b}(E)$. If $E \subseteq \mathbb{R}^d$, we can notice that the weak convergence is equivalent to check that $\mu_n(\psi) \to \mu(\psi)$ for any $\psi \in \mathrm{C_b^2}(E)$. Indeed, it can be shown that it is enough to verify the convergence for every function which is bounded and uniformly continuous (see Remark 8.3.1 in [18]). Then, the result follows since uniformly continuous functions can be uniformly approximated by $\mathrm{C_b^2}(E)$ functions.

It will be useful to introduce a distance over $\mathcal{P}_p(E)$. We define the Wasserstein distance of order $p \geq 1$ between $\mu, \nu \in \mathcal{P}_p(E)$ as

$$\begin{aligned} &\mathcal{W}_p(\mu, \nu) \\ &\quad = \inf \left\{ \left(\int_{E \times E} |x - y|^p \gamma(\,\mathrm{d}x, \,\mathrm{d}y) \right)^{\frac{1}{p}} : \gamma \in \mathcal{P}_p(E \times E) \text{ with marginals } \mu, \nu \right\} \\ &\quad = \inf \left\{ \mathbb{E}\left[|X - Y|^p\right]^{\frac{1}{p}} : \mathfrak{L}(X) = \mu, \mathfrak{L}(Y) = \nu \right\}. \end{aligned}$$

The pair $(\mathcal{P}_p(E), \mathcal{W}_p)$ is a complete and separable metric space (see for instance Theorem 6.18 in [101]). Moreover, the convergence in $\mathcal{W}_p$ implies the weak convergence, whilst the converse is true if there is also the convergence of the moments up to order p-th.

Let $K \subset \mathbb{R}^r$ be compact. In this case $\mathcal{P}(K) = \mathcal{P}_p(K)$ for any $p \in [1, +\infty)$ and the convergence in $\mathcal{W}_p$ is equivalent to the weak convergence. Moreover, $\mathcal{P}(K)$ is compact in the weak topology.

A detailed discussion on these topics can be found for instance in [5, Chapter 7], [101, Chapter 6]. or from a more probabilistic point of view in [31, Chapter 5].

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and $T > 0$ a finite time horizon. Let $\mathbb{F} = \{\mathcal{F}_t\}_{t\in[0,T]}$ be a complete and right-continuous filtration (namely it satisfies the usual conditions). Let ξ be a random variable.

- $\mathfrak{L}(\xi), \mathbb{P}_\xi$: the law of ξ;
- $\mathfrak{P}$: σ-algebra on $\Omega \times [0,T]$ of $\mathcal{F}_t$-progressively measurable sets;
- $\mathcal{T}_t$ is the set of $\mathbb{F}$-stopping times τ such that $\tau \in [t,T]$ almost surely;
- $\mathcal{S}^2$: real-valued $\mathfrak{P}$-measurable càdlàg processes Y such that

$$\|Y\|^2_{\mathcal{S}^2} := \mathbb{E}[\sup_{u\in[0,T]} |Y_u|^2] < \infty;$$

- $\mathcal{S}^2_c$ is the space of $\mathcal{S}^2$-valued continuous processes. This space is complete and separable.

Chapter 1

Nonlinear filtering

This chapter is devoted to a quick introduction to the stochastic filtering topic and to collect some fundamental definitions and results that will be useful in the other chapters of this Thesis. All the statements presented here can be found in the literature (detailed references are given in the latter) and so we will omit the proofs.

The aim of stochastic filtering is to estimate an evolving dynamical system, the *signal*, customarily modelled by a stochastic process. In general, it will be denoted by $X = \{X_t, t \geq 0\}$ where t is the temporal parameter. The process can not be measured directly, however a partial measurement of the process can be obtained. This measurement is modelled by another stochastic process $Y = \{Y_t, t \geq 0\}$, the *observation*. The main goal is to find the so-called *filter*, which is the conditional law of the signal process at time t given the information contained in the observation process up to t.

The stochastic filtering problem was first studied in discrete time by Kolmogorov in [69, 70] and Krein in [72, 73]. The problem in continuous time was discussed firstly by Wiener in a context regarding optimal estimation of dynamical systems in the presence of noise [102]. He solved the problem by using spectral theory of stationary processes.

The next main result has been the formulation of the linear Gaussian filter in the 1960s, where the signal is described by a linear stochastic differential equation. The pioneers in this field have been Kalman and Bucy. One of the first article on this topic is [67], by Kalman, where the author presents a solution to the problem in the discrete case with gaussian initial condition. Bucy obtained similar results independently (see for instance [26]). This was a crucial result, both from a theoretical and applicative point of view, making the linear filter used in many applied and engineering problems. After that, a plethora of extensions have been produced, removing in certain case the gaussian and the linear hypotheses. Initially many of the algorithms developed were quite empirical and not based on a solid mathematical theory.

In the 1960s, Kushner in [76, 77] and Stratonovich in [96, 97] formulated independently a rigorous mathematical theory for nonlinear filtering that led to the so-called nonlinear filtering equation, also known as Kushner-Stratonovich equation or Fujisaki-Kallianpur-Kunita equation. Indeed Fujisaki, Kallianpur and Kunita obtained the same equation in [62] by exploiting the innovation approach introduced by Kailath in

[66]. In the same period, the equation for the unnormalized filter was introduced by Duncan [49, 50], Mortensen [82] and Zakai [105] and consequently referred as Zakai equation or Duncan-Mortensen-Zakai equation.

Since then, these equations have been intensively studied, see for instance [87, 94] and references therein. In the recent literature, a systematic exposition of the topic can be found for instance in the book by Bain and Crisan [8] on in the book by Xiong [104]. We also highlight that the so-called stochastic optimal control with partial observation problem is strongly related with the filtering framework, see for instance the book by Bensoussan [13].

1.1 Formulation of the problem

Our intention here is to present the stochastic filtering problem for a diffusion type signal process. The most general version of the stochastic filtering problem is exhaustively addressed for instance in Chapter 3 of [8], and some comments can be found later in Remark 1.1.1.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $0 < T < \infty$ be a finite time horizon and let $\mathbb{F} := \{\mathcal{F}_t\}_{t\in[0,T]}$ be a filtration that satisfies the usual conditions, that is $\mathbb{F}$ is complete and right continuous. Let $W = \{W_t, t \in [0,T]\}$ and $B = \{B_t, t \in [0,T]\}$ be two independent d-dimensional $\{\mathcal{F}_t\}$-Brownian motions (we take them with the same dimension for simplicity, but the generalization to the case with different dimensions is straightforward). We introduce the signal process $X = \{X_t, t \in [0,T]\}$ as the d-dimensional solution to the stochastic differential equation

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}s + \sigma(X_t)\,\mathrm{d}W_t + \bar{\sigma}(X_t)\,\mathrm{d}B_t, \quad X_0 \in L^2(\Omega, \mathcal{F}_0), \tag{1.1}$$

where the Borel measurable mappings $f\colon \mathbb{R}^d \to \mathbb{R}^d$, $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ and $\bar{\sigma}\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ will be chosen later in order to have existence of a strong solutions, uniqueness (up to indistinguishability), continuity and Markovianity of X. The idea behind stochastic filtering is that we can not observe directly the signal, but we can only observe a process $Y = \{Y_t, t \in [0,T]\}$, called observation process. The dynamic of Y is given by

$$Y_t = \int_0^t h(X_s)\,\mathrm{d}s + B_t, \tag{1.2}$$

where $h\colon \mathbb{R}^d \to \mathbb{R}^d$ is Borel measurable and such that $\mathbb{P}\left(\int_0^T |h(X_s)|\,\mathrm{d}s < \infty\right) = 1$ in order to have the integral in (1.2) well defined. If we introduce the observation filtration $\mathbb{F}^Y := \{\mathcal{F}_t^Y\}_{t\in[0,T]}$, where $\mathcal{F}_t^Y$ is the completion with respect to the $\mathbb{P}$-null sets of the σ-algebra generated by Y up to the time t, we can say that the filtering problem consists in finding the "best" approximation of the signal X given the information contained in $\mathbb{F}^Y$.

In this case, the notion of best approximation relates to the conditional expectation. More precisely, the aim becomes to find a $\mathcal{P}(\mathbb{R}^d)$-valued process $\Pi = \{\Pi_t, t \in [0,T]\}$, called filter, such that

$$\langle \Pi_t, \varphi \rangle = \mathbb{E}\left[\varphi(X_t) | \mathcal{F}_t^Y\right], \tag{1.3}$$

almost surely, for every $t \in [0,T]$ and $\varphi \in \mathrm{B_b}(\mathbb{R}^d)$. Due to the presence of the observation noise B in the signal equation (1.1), we will refer to this problem as stochastic filtering with *correlated noise*, in contrast with the case where $\bar{\sigma}$ is null, that is called without correlated noise.

Now, let us state the assumptions on the coefficients necessary for our discussion. They are not always necessary together (see Remark 1.2.1 and Remark 1.3.12), but we group them to simplify the exposition.

Assumption 1. *All the mappings $f, \sigma, \bar{\sigma}, h$ are taken Borel-measurable. Moreover we assume:*

a. the mappings $f \colon \mathbb{R}^d \to \mathbb{R}^d$, $\sigma \colon \mathbb{R}^d \to \mathbb{R}^{d \times d}$ and $\bar{\sigma} \colon \mathbb{R}^d \to \mathbb{R}^{d \times d}$ are Lipschitz continuous;

b. the mapping $a := \sigma\sigma^\top \colon \mathbb{R}^d \to \mathbb{R}^{d \times d}$ is uniformly elliptic, that is there exists $\lambda > 0$ such that $\sum_{i,j=1}^d a_{ij}\xi_i\xi_j \geq \lambda|\xi|^2$ for every $x, \xi \in \mathbb{R}^d$;

c. the mappings $f, \sigma, \bar{\sigma}, h$ are bounded.

Under these conditions, the signal is a uniquely characterized Markov process and moreover there exists a $\mathcal{P}(\mathbb{R}^d)$-valued $\{\mathcal{F}_t^Y\}$-optional process which satisfies (1.3) (see for instance [8, Theorem 2.1]).

Remark 1.1.1. At this level one can deal with more general signals. Let S be a complete metric space and let us define the operator $A \colon \mathcal{D}(A) \subseteq \mathrm{B_b}(S) \to \mathrm{B_b}(S)$, with $\mathcal{D}(A)$ that denotes its domain. As reported for instance in [8, Section 3.1], the filter can be defined for an S-valued signal which solves the martingale problem with generator A.

1.2 The filtering equations

In the previous section, we introduced the filter $\Pi = \{\Pi_t, t \in [0,T]\}$ as the conditional law of the signal X given the observation filtration $\mathbb{F}^Y$. Despite its natural definition, this object is not easy to use and study directly. A helpful result in this direction is the fact that, under certain conditions, the filter satisfies a stochastic differential equation. Since the filter is a $\mathcal{P}(\mathbb{R}^d)$-valued process, the equation has to be formulated in a suitable weak sense. This result has been obtained independently by Kushner [76, 77], Stratonovich [96, 97] and then by Fujisaki, Kallianpur and Kunita [62], and it is presented for instance in [8, Chapter 3] or [104, Chapter 5].

Before writing the equation solved by the filter, we need to set some notations. Let us introduce two differential operators A and B with domain $\mathrm{C_b^2}(\mathbb{R}^d) \subset \mathrm{B_b}(\mathbb{R}^d)$.

In particular $A, B\colon \mathrm{C}^2_{\mathrm{b}}(\mathbb{R}^d) \to \mathrm{B}_{\mathrm{b}}(\mathbb{R}^d)$ are defined for every $\psi \in \mathrm{C}^2_{\mathrm{b}}(\mathbb{R}^d)$, $x \in \mathbb{R}^d$, by

$$\begin{aligned} A\psi(x) &:= \sum_{i=1}^d f_i(x)\partial_i\psi(x) + \frac{1}{2}\sum_{i,j=1}^d \left\{ \left(\sigma\sigma^\top\right)_{ij}(x)\partial_{ij}\psi(x) + \left(\bar\sigma\bar\sigma^\top\right)_{ij}(x)\partial_{ij}\psi(x) \right\}, \\ B_k\psi(x) &:= \sum_{i=1}^d \bar\sigma_{ik}(x)\partial_i\psi(x), \quad k = 1, \dots d. \end{aligned} \tag{1.4}$$

We can notice that A is the infinitesimal generator associated to the signal process X whilst B is a first-order differential operator which vanishes in the case of filtering problem without correlated noise (that is $\bar\sigma = 0$). Moreover, under Assumption 1, we have that $A, B\colon \mathrm{C}^2_{\mathrm{b}}(\mathbb{R}^d) \to \mathrm{C}_{\mathrm{b}}(\mathbb{R}^d)$.

Let Assumption 1 hold and let us introduce the process

$$Z_t = \exp\left\{ -\int_0^t h(X_s)\cdot \mathrm{d}B_s - \frac{1}{2}\int_0^t |h(X_s)|^2\,\mathrm{d}s \right\}, \quad t \in [0, T],$$

which can be proved to be a martingale. Thus, we can define a new probability by

$$\mathrm{d}\mathbb{Q}|_{\mathcal{F}_t} = Z_t\,\mathrm{d}\mathbb{P}|_{\mathcal{F}_t}.$$

A first key result is the so-called Kallianpur-Striebel formula (see for instance [8, Proposition 3.16] or [104, Theorem 5.3]), that gives us a way to represent the filter in terms on a $\mathcal{M}^+(\mathbb{R}^d)$-valued process $\rho = \{\rho_t, t \in [0, T]\}$ called *unnormalized filter*. In particular, it holds that for every $\psi \in \mathrm{B}_{\mathrm{b}}(\mathbb{R}^d)$ and fixed $t \in [0, T]$, the filter Π can be represented as

$$\langle \Pi_t, \psi\rangle = \frac{\langle \rho_t, \psi\rangle}{\rho_t(\mathbb{R}^d)}, \tag{1.5}$$

where ρ is a $\mathcal{M}^+(\mathbb{R}^d)$-valued process such that $\rho_t(\mathbb{R}^d) > 0$ for every $t \in [0, T]$ and

$$\langle \rho_t, \psi\rangle = \mathbb{E}^{\mathbb{Q}}[Z_t^{-1}\psi(X_t)|\mathcal{F}^Y_t].$$

Once the unnormalized filter is introduced, one can show that it satisfies the following stochastic differential equation, called the Zakai equation (see, for instance, [104, Theorem 5.5]):

$$\langle \rho_t, \psi\rangle = \langle \pi, \psi\rangle + \int_0^t \langle \rho_s, A\psi\rangle\,\mathrm{d}s + \int_0^t \langle \rho_s, h\psi + B\psi\rangle \cdot \mathrm{d}Y_s, \tag{1.6}$$

for every $\psi \in \mathrm{C}^2_{\mathrm{b}}(\mathbb{R}^d)$ and with $\pi = \mathfrak{L}(X_0)$. We can notice that the equation is linear in ρ and written in weak sense, that is it must hold for every test function in $\mathrm{C}^2_{\mathrm{b}}(\mathbb{R}^d)$. We highlight that, for every ψ, the process $\langle \rho_t, \psi\rangle$ is real-valued and so there is no need to introduce a notion of stochastic integral for measure-valued processes. Moreover, from Girsanov's theorem it follows that the observation process Y, which drives the stochastic integral in (1.6), is a Brownian motion under $\mathbb{Q}$.

Starting from (1.6), by using the Kallianpur-Striebel formula it can be shown that the filter Π satisfies the following stochastic differential equation, called Kushner-Stratonovich equation (or Fujisaki-Kallianpur-Kunita equation):

$$\Pi_t(\psi) = \langle \pi, \psi \rangle + \int_0^t \langle \Pi_s, A\psi \rangle \, \mathrm{d}s + \int_0^t \left(\langle \Pi_s, h\psi + B\psi \rangle - \langle \Pi_s, \psi \rangle \langle \Pi_s, h \rangle \right) \cdot \mathrm{d}I_s, \quad (1.7)$$

for every $\psi \in \mathrm{C}_\mathrm{b}^2(\mathbb{R}^d)$ and with $\pi = \mathfrak{L}(X_0)$. The process $I = \{I_t, t \in [0,T]\}$, called *innovation process*, is defined for every $t \in [0,T]$ by

$$I_t := Y_t - \int_0^t \langle \Pi_s, h \rangle \, \mathrm{d}s, \quad (1.8)$$

and one can prove that it is a d-dimensional $\{\mathcal{F}_t^Y\}$-Brownian motion under $\mathbb{P}$ (see, for instance, [104, Lemma 5.6]). We also notice that, differently from the Zakai equation, the Kushner-Stratonovich equation is nonlinear in Π, due to the presence of the term $\langle \Pi_s, \psi \rangle \langle \Pi_s, h \rangle$.

Remark 1.2.1. We presented the filtering equations for a diffusion type signal and under Assumption 1, since our main results in the next chapters hold under these hypotheses. However, the results stated above hold for more general signals and for less restrictive conditions. For instance, we can consider a signal X which is the solution of a martingale problem (see Remark 1.1.1), and we can formulate both the Kushner-Stratonovich and the Zakai equations assuming that h is such that

$$\mathbb{E}^{\mathbb{P}}\left[\int_0^t |h(X_s)|^2 \, \mathrm{d}s\right] < \infty, \quad \mathbb{E}^{\mathbb{P}}\left[\int_0^t Z_s |h(X_s)|^2 \, \mathrm{d}s\right] < \infty,$$

and

$$\mathbb{Q}\left(\int_0^t \rho_s |h|^2 \, \mathrm{d}s < \infty\right) = 1.$$

A detailed exposition of the most general framework can be found, for instance, in [8, Chapters 2,3], and in the references therein.

1.3 Weak solutions to the filtering equations

In the previous sections we introduced the stochastic filtering problem and the related stochastic differential equations. As it can be seen directly from (1.6) and (1.7), the equations are strongly related to the structure of the filtering problem, since they are driven by the observation process Y and the innovation process I respectively. The purpose of this section is to present a framework to study directly the Zakai and the Kushner-Stratonovich equations as equation for measure-valued processes, without relying on the original filtering problem. To do so, we follow the ideas in the work by Szpirglas [99] (and the extension to the correlated noise case obtained by Heunis and Lucic in [80]), introducing proper notions of weak solution, pathwise uniqueness and uniqueness in law. We present all the necessary definitions, mainly following the exposition in [80].

First, we introduce a notion of weak solution of the Kushner-Stratonovich equation, which mimics the classical notion of weak solution of a stochastic differential equation.

Definition 1.3.1. *The pair* $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}_{t\in[0,T]}, \tilde{\mathbb{P}}), (\tilde{\Pi}, \tilde{I})\}$ *is a weak solution to the Kushner-Stratonovich equation starting at* $\pi \in \mathcal{P}(\mathbb{R}^d)$ *if:*

i. $(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}_{t\in[0,T]}, \tilde{\mathbb{P}})$ *is a complete filtered probability space.*

ii. $\tilde{I} = \{\tilde{I}_t, t \in [0,T]\}$ *is an* $\mathbb{R}^d$*-valued* $\{\tilde{\mathcal{F}}_t\}$*-Brownian motion on* $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$.

iii. $\tilde{\Pi} = \{\tilde{\Pi}_t, t \in [0,T]\}$ *is a* $\mathcal{P}(\mathbb{R}^d)$*-valued continuous* $\{\tilde{\mathcal{F}}_t\}$*-adapted process such that*

$$\tilde{\mathbb{P}}\left(\int_0^T \sum_{i=1}^d \langle \tilde{\Pi}_s, |h_i|\rangle^2 \,\mathrm{d}s < \infty\right) = 1,$$

and for every $\psi \in \mathrm{C}^2_{\mathrm{b}}(\mathbb{R}^d)$ *it holds*

$$\tilde{\Pi}_t(\psi) = \langle \pi, \psi\rangle + \int_0^t \langle \tilde{\Pi}_s, A\psi\rangle \,\mathrm{d}s + \int_0^t \Big(\langle \tilde{\Pi}_s, h\psi + B\psi\rangle - \langle \tilde{\Pi}_s, \psi\rangle\langle \tilde{\Pi}_s, h\rangle\Big) \cdot \mathrm{d}\tilde{I}_s, \tag{1.9}$$

for every $t \in [0,T]$*, almost surely.*

Analogously, we can define the Zakai equation's weak solutions:

Definition 1.3.2. *The pair* $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}_{t\in[0,T]}, \tilde{\mathbb{Q}}), (\tilde{\rho}, \tilde{Y})\}$ *is a weak solution to the Zakai equation starting at* $\mu \in \mathcal{M}^+(\mathbb{R}^d)$ *if:*

i. $(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}_{t\in[0,T]}, \tilde{\mathbb{Q}})$ *is a complete filtered probability space.*

ii. $\tilde{Y} = \{\tilde{Y}_t, t \in [0,T]\}$ *is an* $\mathbb{R}^d$*-valued* $\{\tilde{\mathcal{F}}_t\}$*-Brownian motion on* $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{Q}})$.

iii. $\tilde{\rho} = \{\tilde{\rho}_t, t \in [0,T]\}$ *is a* $\mathcal{M}^+(\mathbb{R}^d)$*-valued continuous* $\{\tilde{\mathcal{F}}_t\}$*-adapted process such that*

$$\tilde{\mathbb{Q}}\left(\int_0^T \sum_{i=1}^d \langle \tilde{\rho}_s, |h_i\psi + B_i\psi|\rangle^2 \,\mathrm{d}s < \infty\right) = 1,$$

and for every $\psi \in \mathrm{C}^2_{\mathrm{b}}(\mathbb{R}^d)$ *it holds*

$$\langle \tilde{\rho}_t, \psi\rangle = \mu(\psi) + \int_0^t \langle \tilde{\rho}_s, A\psi\rangle \,\mathrm{d}s + \int_0^t \langle \tilde{\rho}_s, h\psi + B\psi\rangle \cdot \mathrm{d}\tilde{Y}_s, \tag{1.10}$$

for every $t \in [0,T]$*, almost surely.*

Remark 1.3.3. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$ be the probabilistic setup introduced in Section 1.1. If we consider the filter Π and the innovation process I defined by (1.8), we have that $\{(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P}), (\Pi_t, I_t)\}$ is a weak solution to the Kushner-Stratonovich equation. Analogously, if we take the unnormalized filter ρ and the probability $\mathbb{Q}$ introduced in Section 1.2, we have that Y a Brownian motion under $\mathbb{Q}$ and then $\{(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{Q}), (\rho_t, Y_t)\}$ is a weak solution to the Zakai equation.

Remark 1.3.4. A useful result, proved in [80] (see Fact 3.2), is the fact that the trajectories of weak solutions to the Zakai equation have total mass that does not touch zero and with uniformly bounded moments. More precisely, if $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}_{t\in[0,T]}, \tilde{\mathbb{Q}}), (\tilde{\rho}, \tilde{Y})\}$ is a weak solution to the Zakai equation starting at $\mu \in \mathcal{M}^+(\mathbb{R}^d)$, then $\tilde{\rho}_t(\mathbb{R}^d) > 0$ for every $t \in [0, T]$, $\tilde{\mathbb{Q}}$ almost surely. Moreover for every $\alpha \in (1, +\infty)$ there exists a positive constant $\gamma(\alpha, \mu)$ such that

$$\mathbb{E}^{\tilde{\mathbb{Q}}}\left[\sup_{t\in[0,T]} |\tilde{\rho}_t(\mathbb{R}^d)|^\alpha\right] \leq \gamma(\alpha, \mu).$$

Remark 1.3.5. In [99] and [80], the Zakai equation is always taken with initial condition in the space of probability measures. However, it is easy to consider the case in which the initial condition is a positive measure, different from the null measure. Indeed one can always reconduct the problem to the one starting from a probability by performing a standardization, thanks to the linearity of the Zakai equation.

In the following lemma we show that if a weak solution to a filtering equation starts from a measure with finite second moment, then its trajectories will take value in a space of measures with finite second moment. This result will be useful in the next chapters and we postpone its proof (see Section 1.3.1) to avoid technicalities in this expository chapter on nonlinear filtering.

Lemma 1.3.6. *Let $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}_{t\in[0,T]}, \tilde{\mathbb{Q}}), (\tilde{\rho}, \tilde{Y})\}$ be a weak solution to the Zakai equation starting at $\mu \in \mathcal{M}_2^+(\mathbb{R}^d)$. Then $\tilde{\rho}_t \in \mathcal{M}_2^+(\mathbb{R}^d)$ for every $t \in [0, T]$, $\tilde{\mathbb{Q}}$-almost surely. Similarly, if $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}_{t\in[0,T]}, \tilde{\mathbb{P}}), (\tilde{\Pi}, \tilde{I})\}$ is a weak solution to the Kushner-Stratonovich equation starting at $\pi \in \mathcal{P}_2(\mathbb{R}^d)$, then $\tilde{\Pi}_t \in \mathcal{P}_2(\mathbb{R}^d)$ for every $t \in [0, T]$, $\tilde{\mathbb{P}}$-almost surely.*

Regarding the notion of uniqueness, we have the following two definitions, which follow the classical Yamada-Watanabe formalism.

Definition 1.3.7. *The Kushner-Stratonovich equation has the pathwise uniqueness property if: given two weak solutions*

$$\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}_{t\in[0,T]}, \tilde{\mathbb{P}}), (\tilde{\Pi}^1, \tilde{I})\}, \quad \{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}_{t\in[0,T]}, \tilde{\mathbb{P}}), (\tilde{\Pi}^2, \tilde{I})\},$$

of the equation starting at $\pi \in \mathcal{P}(\mathbb{R}^d)$, it holds that

$$\tilde{\mathbb{P}}\left(\tilde{\Pi}^1_t = \tilde{\Pi}^2_t, \forall t \in [0, T]\right) = 1.$$

In the same way we state the pathwise uniqueness property for the Zakai equation.

Definition 1.3.8. *The Kushner-Stratonovich equation has the uniqueness in joint law property if: given two weak solutions*

$$\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}_{t\in[0,T]}, \tilde{\mathbb{P}}), (\tilde{\Pi}, \tilde{I})\}, \quad \{(\hat{\Omega}, \hat{\mathcal{F}}, \{\hat{\mathcal{F}}_t\}, \hat{\mathbb{P}}), (\hat{\Pi}, \hat{I})\},$$

of the equation starting at $\pi \in \mathcal{P}(\mathbb{R}^d)$, it holds that $(\tilde{\Pi}, \tilde{I}) = \{(\tilde{\Pi}_t, \tilde{I}_t), t \in [0, T]\}$ and $(\hat{\Pi}, \hat{I}) = \{(\hat{\Pi}_t, \hat{I}_t), t \in [0, T]\}$ have the same finite-dimensional distributions, where we endowed $\mathcal{P}(\mathbb{R}^d)$ with the Borel σ-algebra induced by the weak convergence topology.
In the same way we define the uniqueness in joint law property for the Zakai equation.

The main result in [80] is the following theorem regarding the uniqueness for the two equations of nonlinear filtering:

Theorem 1.3.9 ([80])**.** *Let Assumption 1 holds. Then:*

i. the Zakai equation has the pathwise uniqueness and the uniqueness in joint law properties;

ii. the Kushner-Stratonovich equation has the uniqueness in joint law property.

Remark 1.3.10. In the case without correlated noise, studied in [99], it is possible to prove pathwise uniqueness also for the weak solutions of the Kushner-Stratonovich equation. Moreover, in [99] (Théorème V.6) it is shown how the classical Yamada-Watanabe result also apply to this situation, namely pathwise uniqueness and existence of a weak solution implies existence of a strong solution. The technique is not affected by the addition of a correlated noise, so at least for the Zakai equation we also have existence of a strong solution. This allows us to fix a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}_{t\in[0,T]}, \tilde{\mathbb{Q}})$, equipped with a $\{\tilde{\mathcal{F}}_t\}$-Brownian motion $\tilde{Y}$, and to solve the Zakai equation with respect to different initial conditions on the same probabilistic setup.

Remark 1.3.11. Since the uniqueness in law property holds, the Markov property follows for both the weak solution of the Kushner-Stratonovich and the Zakai equation.

Remark 1.3.12. The fact that $a := \sigma\sigma^\top$ must be uniformly elliptic (Assumption 1 - *b.*) is necessary only for the proof of Proposition 4.1.4 (in particular, it is required to ensure the existence of a smooth solution to (4.7)). For all the previous results, one can just assume a non-degeneracy condition on a, that is $a(x)$ positive definite for every $x \in \mathbb{R}^d$. Moreover, the non-degeneracy of σ is required only for Theorem 1.3.9. Thus, if one assume the conclusions of Theorem 1.3.9 and the Markov property for the solutions, as well as the existence of a smooth solution to (4.7), then all the following results still hold without Assumption 1 - *b*.

We conclude this section by stating how to obtain a weak solution to the Kushner-Stratonovich equation from a weak solution to the Zakai equation and viceversa. First, let us assume that Assumption 1 holds and let $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}_{t\in[0,T]}, \tilde{\mathbb{P}}), (\tilde{\Pi}, \tilde{I})\}$ be a weak solution to the Kushner-Stratonovich equation. Then, we can define two processes $\tilde{Y} = \{\tilde{Y}_t, t \in [0,T]\}$ and $\chi := \{\chi_t, t \in [0,T]\}$ by

$$\tilde{Y}_t = \tilde{I}_t + \int_0^t \langle \tilde{\Pi}_s, h\rangle \,\mathrm{d}s, \quad \chi_t = \exp\left\{-\int_0^t \langle \tilde{\Pi}_s, h\rangle \cdot \mathrm{d}\tilde{I}_s + \frac{1}{2}\int_0^t |\langle \tilde{\Pi}_s, h\rangle|^2 \,\mathrm{d}s\right\}, \quad (1.11)$$

where it is easy to see that χ in a $\tilde{\mathbb{P}}$-martingale. Thus, if we introduce the probability measure

$$\mathrm{d}\tilde{\mathbb{Q}} = \chi_T \,\mathrm{d}\tilde{\mathbb{P}}$$

and

$$\tilde{\rho} = \{\tilde{\rho}_t = \mu(\mathbb{R}^d)\chi_t^{-1}\tilde{\Pi}_t, t \in [0,T]\},$$

we have that $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}_{t\in[0,T]}, \tilde{\mathbb{Q}}), (\tilde{\rho}, \tilde{Y})\}$ is a weak solution to the Zakai equation starting at $\mu \in \mathcal{M}_2^+(\mathbb{R}^d)$. We remark that the presence of $\mu(\mathbb{R}^d)$ in the definition of

$\tilde{\rho}$ is necessary to keep track that the initial condition μ is not a probability measure (see also Remark 1.3.5). On the other hand, if $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}_{t\in[0,T]}, \tilde{\mathbb{Q}}), (\tilde{\rho}, \tilde{Y})\}$ is a weak solution to the Zakai equation, we can set

$$\tilde{I}_t = \tilde{Y}_t - \int_0^t \frac{\langle \tilde{\rho}_s, h\rangle}{\tilde{\rho}_s(\mathbb{R}^d)} \,\mathrm{d}s, \quad \xi_t = \exp\left\{ \int_0^t \frac{\langle \tilde{\rho}_s, h\rangle}{\tilde{\rho}_s(\mathbb{R}^d)} \cdot \mathrm{d}\tilde{Y}_s - \frac{1}{2}\int_0^t \left| \frac{\langle \tilde{\rho}_s, h\rangle}{\tilde{\rho}_s(\mathbb{R}^d)} \right|^2 \mathrm{d}s \right\}. \quad (1.12)$$

Since ξ is a martingale, we can introduce

$$\mathrm{d}\tilde{\mathbb{P}} = \xi_T \,\mathrm{d}\tilde{\mathbb{Q}}$$

and set

$$\tilde{\Pi} = \{\tilde{\Pi}_t = \tilde{\rho}_s / \tilde{\rho}_s(\mathbb{R}^d), t \in [0,T]\}.$$

Thus, the couple $\{(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{\mathcal{F}}_t\}_{t\in[0,T]}, \tilde{\mathbb{P}}), (\tilde{\Pi}, \tilde{I})\}$ is a weak solution to the Kushner-Stratonovich equation.

The proofs of these results in the context of weak solutions to the nonlinear filtering equations can be found in [99] for the case without correlated noise. However, the case with correlated noise has no extra difficulties. The technique used to prove these relations is based on Girsanov's theorem and on the classical Itô formula, and it follows the way to link the filter Π and the unnormalized filter ρ. A detailed discussion about this change of probability method in the context of the filtering problem can be found for instance in [8, Chapter 3] or in [104, Chapter 5].

1.3.1 Proof of Lemma 1.3.6

For simplicity, we provide a sketch of the proof for the Zakai equation in the one dimensional case ($d = 1$). The general case with $d > 1$ and the case of Kushner-Stratonovich equation are immediate extensions. In order to keep the notation lighter, we remove the tildes in the notation for the weak solutions and we will denote the expectation $\mathrm{E}^{\tilde{\mathbb{Q}}}$ with respect to $\tilde{\mathbb{Q}}$ just with $\mathbb{E}$.

First, we show that if $\mu \in \mathcal{M}_1^+(\mathbb{R})$, then $\rho_t \in \mathcal{M}_1^+(\mathbb{R})$ for every $t \in [0,T]$, almost surely. Let us consider a smooth function $\psi\colon \mathbb{R} \to \mathbb{R}$ which is greater than the mapping $x \mapsto |x|$ in a neighbourhood of 0 and equal to $x \mapsto |x|$ outside that neighbourhood. In particular, ψ has bounded first and second-order derivatives. If we show that $\langle \rho_t, \psi\rangle < \infty$ for every $t \in [0,T]$, almost surely, then this first claim is proved.

Let us consider an increasing family of smooth cut-off functions $\{\phi^N\}_{N\geq 1}$ which are equal to one in $[-N, N]$ and equal to zero outside $[-N-1, N+1]$. These functions can be chosen to be bounded together with their first and second-order derivatives by a constant $C > 0$ independent of N. We set $\psi^N(x) = \psi(x)\phi^N(x)$ for every $x \in \mathbb{R}$, thus it holds

$$\begin{aligned} \mathrm{D}_x\psi^N(x) &= \psi(x)\mathrm{D}_x\phi^N(x) + \mathrm{D}_x\psi(x)\phi^N(x), \\ \mathrm{D}_x^2\psi^N(x) &= \psi(x)\mathrm{D}_x^2\phi^N(x) + 2\mathrm{D}_x\psi(x)\mathrm{D}_x\phi^N(x) + \mathrm{D}_x^2\psi(x)\phi^N(x), \end{aligned}$$

and moreover $\{\psi^N\}_{N\ge1}, \{\mathrm{D}_x\psi^N\}_{N\ge1}$ and $\{\mathrm{D}_x^2\psi^N\}_{N\ge1}$ converge pointwise to $\psi, \mathrm{D}_x\psi$ and $\mathrm{D}_x^2\psi$ respectively, where the first convergence takes place monotonically. We also notice that for every $N \ge 1$, $\psi^N \in \mathrm{C}_\mathrm{b}^2(\mathbb{R})$.

Let us fix $N \ge 1$. Since ρ is a weak solution to the Zakai equation, it holds for every $t \in [0, T]$

$$\langle \rho_t, \psi^N\rangle = \langle \mu, \psi^N\rangle + \int_0^t \langle \rho_s, A\psi^N\rangle \,\mathrm{d}s + \int_0^t \langle \rho_s, (h + B)\psi^N\rangle \cdot \mathrm{d}Y_s.$$

By taking the square, the expectation and then by Itô isometry, we get

$$\begin{aligned}&\mathbb{E}\left[\langle \rho_t, \psi^N\rangle^2\right]\\ &\qquad \le 3\langle \mu, \psi^N\rangle^2 + 3T\int_0^t \mathbb{E}\left[\langle \rho_s, A\psi^N\rangle^2\right] \mathrm{d}s + 3\int_0^t \mathbb{E}\left[\langle \rho_s, (h+B)\psi^N\rangle^2\right] \mathrm{d}s.\end{aligned}$$

Now, if we write explicitly the operators A, B and we use the boundedness of $b, \sigma, \bar{\sigma}, h$ jointly with the boundedness of $\mathrm{D}_x\psi, \mathrm{D}_x^2\psi, \phi^N, \mathrm{D}_x\phi^N, \mathrm{D}_x^2\phi^N$ (recalling that the bound for $\phi^N, \mathrm{D}_x\phi^N, \mathrm{D}_x^2\phi^N$ does not depend on N), we obtain the inequality

$$\mathbb{E}\left[\langle \rho_t, \psi^N\rangle^2\right] \le M_1\left(\langle \mu, \psi^N\rangle^2 + \int_0^t \mathbb{E}\left[\langle \rho_s, \psi^N\rangle^2\right] + \mathbb{E}\left[\rho_s(\mathbb{R})^2\right] \mathrm{d}s\right), \tag{1.13}$$

where $M_1 = M_1(b, \sigma, \bar{\sigma}, h, T, C, \psi)$ is a positive constant independent of t and N. Thanks to the monotone convergence theorem, we can pass to the limit as $N \to +\infty$ in (1.13) and get

$$\mathbb{E}\left[\langle \rho_t, \psi\rangle^2\right] \le M_1\left(\langle \mu, \psi\rangle^2 + \int_0^t \mathbb{E}\left[\langle \rho_s, \psi\rangle^2\right] + \mathbb{E}\left[\rho_s(\mathbb{R})^2\right] \mathrm{d}s\right).$$

Thus, in view of Remark 1.3.4 and by Gronwall's lemma, there exists a positive constant $M_2 = M_2(b, \sigma, \bar{\sigma}, h, T, C, \psi, \mu)$ such that for every $t \in [0, T]$ it holds that

$$\mathbb{E}\left[\langle \rho_t, \psi\rangle^2\right] \le M_2\langle \mu, \psi\rangle^2. \tag{1.14}$$

Since the bound (1.14) does not depend on t, we can proceed similarly to the previous steps and by Burkholder inequality and monotone convergence we can also deduce that there exists a positive constant $M_3 = M_3(b, \sigma, \bar{\sigma}, h, T, C, \psi, \mu)$ such that

$$\mathbb{E}\left[\sup_{0\le t\le T} \langle \rho_t, \psi\rangle^2\right] \le M_3\langle \mu, \psi\rangle^2, \tag{1.15}$$

thus if $\mu \in \mathcal{M}_1^+(\mathbb{R})$ then $\sup_{0\le t\le T}\langle \rho_t, \psi\rangle < +\infty$ almost surely and so $\rho_t \in \mathcal{M}_1^+(\mathbb{R})$ for every $t \in [0, T]$, almost surely.

To conclude, we need to prove that if $\mu \in \mathcal{M}_2^+(\mathbb{R})$ then $\rho_t \in \mathcal{M}_2^+(\mathbb{R})$ for every $t \in [0, T]$, almost surely. To this aim, we can proceed analogously to the above case in which $\mu \in \mathcal{M}_1^+(\mathbb{R})$, choosing $\psi(x) = x^2$ and noticing that its first and second-order derivatives are linear and constant respectively. Then, we can still use the approximation technique, combined with Remark 1.3.4 and (1.15) and so the lemma is proved.

Chapter 2

Properties of functions on spaces of measures

The main aim of this chapter is to introduce some analytical tools that are necessary to deal with functions over spaces of positive and probability measures. The first question that we face is which notion of derivative fits our problem best. Then, we provide a procedure that allows us to approximate "smooth" functions over the space of positive measures with a special class of function, namely the cylindrical functions.

2.1 Derivatives on spaces of measures

Since our final goal is to discuss some Kolmogorov equations on suitable spaces of measures, we need to introduce notions of derivatives for real-valued or measure-valued functions over spaces of measures. Regarding the real-valued functions, we take inspiration from the literature recently developed for real-valued functions over the space of probability measures. We give a little extension of the notion of linear functional derivative (or flat derivative) discussed for instance in [29, 30, 31]. Another definition we need mimics the so-called L-derivative introduced in the context of mean field games and discussed for instance in [28, 31]. Following [30], we define it as the spatial gradient of the linear functional derivative. For probability measures, under proper hypotheses, this definition coincides with the original one given by Lions in [79] through the lifting on a Hilbert space. A discussion on the relations among these definitions can be found in [31] in the case of probability measures or in [93] in a more general case. The last definition we introduce is a notion of derivative of functions from $\mathcal{M}^+(\mathbb{R}^d)$ to $\mathcal{M}^+(\mathbb{R}^d)$. This is a new definition, strongly inspired by the previous ones.

Definition 2.1.1 (Linear functional derivative)**.** *A function $u\colon \mathcal{M}^+(\mathbb{R}^d) \to \mathbb{R}$ is said to have linear functional derivative if it is continuous, bounded and if there exists a function*

$$\delta_\mu u\colon \mathcal{M}^+(\mathbb{R}^d) \times \mathbb{R}^d \ni (\mu, x) \mapsto \delta_\mu u(\mu, x) \in \mathbb{R},$$

that is bounded, continuous for the product topology, $\mathcal{M}^+(\mathbb{R}^d)$ equipped with the weak *topology, and such that for all* μ *and* μ' *in* $\mathcal{M}^+(\mathbb{R}^d)$*, it holds:*

$$u(\mu') - u(\mu) = \int_0^1 \int_{\mathbb{R}^d} \delta_\mu u\,(t\mu' + (1-t)\mu, x)\,[\mu' - \mu](\,\mathrm{d}x)\,\mathrm{d}t. \tag{2.1}$$

We call $\mathrm{C}^1(\mathcal{M}^+(\mathbb{R}^d))$ *the class of functions from* $\mathcal{M}^+(\mathbb{R}^d)$ *to* $\mathbb{R}$ *that are differentiable in linear functional sense.*

Remark 2.1.2. If $u \in \mathrm{C}^1(\mathcal{M}^+(\mathbb{R}^d))$, we can introduce a notion of second-order derivative by asking that the mapping $\mu \mapsto \delta_\mu u(\mu, x)$ is differentiable in linear functional sense for every x and that $(\mu, x, y) \mapsto \delta^2_\mu u(\mu, x, y)$ is bounded and continuous. In general one can define derivatives of order $k \in \mathbb{N}$ and introduce the space $\mathrm{C}^k(\mathcal{M}^+(\mathbb{R}^d))$ of functions that are k times differentiable in linear functional sense. Notice that every time we differentiate, the derivative depends on a new spatial variable.

We introduce now the second notion of derivative, namely the L-derivative, for real-valued functions over $\mathcal{M}^+(\mathbb{R}^d)$. We follow the definition given in [30], since for positive measure we cannot rely on the lifting procedure of [79].

Definition 2.1.3 (L-derivative)**.** *A function* u *is said to be* L*-differentiable if it is in* $\mathrm{C}^1(\mathcal{M}^+(\mathbb{R}^d))$ *and if, for every* $\mu \in \mathcal{M}^+(\mathbb{R}^d)$*, the mapping* $\mathbb{R}^d \ni x \mapsto \delta_\mu u(\mu, x) \in \mathbb{R}$ *is everywhere differentiable, with* $\mathcal{M}^+(\mathbb{R}^d) \times \mathbb{R}^d \ni (\mu, x) \mapsto \mathrm{D}_x \delta_\mu u(\mu, x) \in \mathbb{R}^d$ *continuous and bounded. We set:*

$$\mathrm{D}_\mu u(\mu, x) \coloneqq \mathrm{D}_x \delta_\mu u(\mu, x) \in \mathbb{R}^d, \tag{2.2}$$

and we denote this class of functions with $\mathrm{C}^1_{\mathrm{L}}(\mathcal{M}^+(\mathbb{R}^d))$*.*

Remark 2.1.4. If we consider functions over $\mathcal{P}_2(\mathbb{R}^d)$ (see Remark 2.1.7), Definition 2.1.3 turns out to coincide with the definition of L-derivative introduced by Lions in [79] and discussed for instance in [28, 31]. More relations with other notions of derivative in the case of signed measures have been also investigated in [93].

Regarding the second-order L-derivative, again in view of [30], we give the following definition:

Definition 2.1.5. *A function* $u\colon \mathcal{M}^+(\mathbb{R}^d) \to \mathbb{R}$ *is said to be in* $\mathrm{C}^2_{\mathrm{L}}(\mathcal{M}^+(\mathbb{R}^d))$ *if the following conditions hold:*

i. *u is in* $\mathrm{C}^2(\mathcal{M}^+(\mathbb{R}^d))$*;*

ii. *the mapping* $\mathbb{R}^d \ni x \mapsto \delta_\mu u(\mu, x) \in \mathbb{R}$ *is everywhere twice differentiable, with continuous and bounded derivatives on* $\mathcal{M}^+(\mathbb{R}^d) \times \mathbb{R}^d$*;*

iii. *the mapping* $\mathbb{R}^d \times \mathbb{R}^d \ni (x, y) \mapsto \delta^2_\mu u(\mu, x, y)$ *is twice differentiable, with continuous and bounded derivatives on* $\mathcal{M}^+(\mathbb{R}^d) \times \mathbb{R}^d \times \mathbb{R}^d$*.*

We define the second-order L*-derivative of* $u \in \mathrm{C}^2_{\mathrm{L}}(\mathcal{M}^+(\mathbb{R}^d))$ *as follows:*

$$\mathrm{D}^2_\mu u(\mu, x, y) \coloneqq \mathrm{D}_x \mathrm{D}_y^\top \delta^2_\mu u(\mu, x, y) \in \mathbb{R}^{d \times d},$$

where $\mathrm{D}_y^\top = [\partial_{y_1}, \dots, \partial_{y_d}]$ *is the gradient (with respect to* y*) operator, seen as a row.*

Remark 2.1.6. In order to define the second-order L-derivative $\mathrm{D}^2_\mu u$, it is enough to ask for less regularity of the mappings $x \mapsto \delta_\mu u(\mu, x)$ and $(x, y) \mapsto \delta^2_\mu u(\mu, x, y)$ (for instance $x \mapsto \delta_\mu u(\mu, x)$ can be one time continuously differentiable). However, for our scopes, it is necessary to require further regularity of these mappings and so we included it in the definition to keep the exposition clearer.

Remark 2.1.7. If we consider $u\colon \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$, then we ask that the condition (2.1) in Definition 2.1.1 holds for every $\mu, \mu' \in \mathcal{P}(\mathbb{R}^d)$. In this case we will denote by $\mathrm{C}^1(\mathcal{P}(\mathbb{R}^d))$ the space of all the functions differentiable in this sense. Of course, we can proceed in the same way for the derivatives of higher-order or for the L-derivatives. Moreover, this also works for subsets like $\mathcal{M}^+_p(\mathbb{R}^d)$ and $\mathcal{P}_p(\mathbb{R}^d)$, $p \in [1, +\infty)$.

Remark 2.1.8. If we consider functions defined over the space of probability measures $\mathcal{P}(\mathbb{R}^d)$, we have that the linear functional derivative is defined up to an additive constant (see for instance [31, Remark 5.46]). A way to guarantee uniqueness (see for instance [30, Section 2.2.1]) is to adopt the convention

$$\int_{\mathbb{R}^d} \delta_\mu u(\mu, x)\mu(\,\mathrm{d}x) = 0, \quad \mu \in \mathcal{P}(\mathbb{R}^d). \tag{2.3}$$

Remark 2.1.9. We can also give the definitions of linear functional and L-derivative in the case of measures with compact support $\mathcal{M}^+(K)$, $K \subset \mathbb{R}^d$ compact with sufficiently regular boundary. In this case the additional variable generated by the differentiation belongs to K and the spatial differentiability required for the L-derivative has to be meant only in the proper direction at the boundary of K.

Remark 2.1.10. Let $(\mathrm{B}, \|\cdot\|_{\mathrm{B}})$ be a Banach space and let us consider $u\colon \mathcal{M}^+(\mathbb{R}^d) \to \mathrm{B}$. Then, all the previous definitions can be trivially extended to this framework. We will use the notations $\mathrm{C}^{\mathrm{k}}(\mathcal{M}^+(\mathbb{R}^d); \mathrm{B})$, $\mathrm{C}^1_{\mathrm{L}}(\mathcal{M}^+(\mathbb{R}^d); \mathrm{B})$ and $\mathrm{C}^2_{\mathrm{L}}(\mathcal{M}^+(\mathbb{R}^d); \mathrm{B})$. In this case $\delta_\mu u\colon \mathcal{M}^+(\mathbb{R}^d) \times \mathbb{R}^d \to \mathrm{B}$ and $\mathrm{D}_\mu u\colon \mathcal{M}^+(\mathbb{R}^d) \times \mathbb{R}^d \to \mathrm{B}^d$, and the same holds for higher-order derivatives.

We conclude this first part of the section by presenting an example of computations of linear functional and L-derivatives.

Example 2.1.11. Let $g\colon \mathbb{R}^n \to \mathbb{R}$ be in $\mathrm{C}^2_{\mathrm{b}}(\mathbb{R}^n)$ and let $\{\psi\}_{i=1}^n \subset \mathrm{C}_{\mathrm{b}}(\mathbb{R}^d)$. We define

$$u\colon \mathcal{M}^+(\mathbb{R}^d) \ni \mu \mapsto g\left(\langle \mu, \psi_1\rangle, \dots, \langle \mu, \psi_n\rangle\right) \in \mathbb{R}.$$

Then $u \in \mathrm{C}^2(\mathcal{M}^+(\mathbb{R}^d))$ and it holds:

$$\delta_\mu u(\mu, x) = \sum_{k=1}^n \partial_k g\left(\langle \mu, \psi_1\rangle, \dots, \langle \mu, \psi_n\rangle\right) \psi_k(x),$$
$$\delta^2_\mu u(\mu, x, y) = \sum_{k,l=1}^n \partial_l \partial_k g\left(\langle \mu, \psi_1\rangle, \dots, \langle \mu, \psi_n\rangle\right) \psi_k(x)\psi_l(y).$$

Moreover, if $\{\psi_i\}_{i=1}^n \subset \mathrm{C}^2_\mathrm{b}(\mathbb{R}^d)$, then u is in $\mathrm{C}^2_\mathrm{L}(\mathcal{M}^+(\mathbb{R}^d))$ and it holds:

$$\mathrm{D}_\mu u(\mu, x) = \sum_{k=1}^n \partial_k g\left(\langle \mu, \psi_1\rangle, \dots, \langle \mu, \psi_n\rangle\right) \mathrm{D}_x \psi_k(x),$$
$$\mathrm{D}^2_\mu u(\mu, x, y) = \sum_{k,l=1}^n \partial_l \partial_k g\left(\langle \mu, \psi_1\rangle, \dots, \langle \mu, \psi_n\rangle\right) \mathrm{D}_x \psi_k(x) \mathrm{D}_y^\top \psi_l(y).$$

These functions, which we call cylindrical, play a key role in the proof of the Itô formula in Section **??**. We will discuss in Section 2.2 some approximation properties of this class.

The last definition we give concerns the differentiability for functions from $\mathcal{M}^+(\mathbb{R}^d)$ to $\mathcal{M}^+(\mathbb{R}^d)$. The idea is to ask for a relation similar to (2.1) for the measure-valued function tested against regular functions.

Definition 2.1.12 (Linear functional derivative for measure-valued functions)**.** *We say that a function $m\colon \mathcal{M}^+(\mathbb{R}^d) \to \mathcal{M}^+(\mathbb{R}^d)$ is differentiable in linear functional sense if there exists a mapping*

$$\tilde{\delta}_\mu m\colon \mathcal{M}^+(\mathbb{R}^d) \times \mathbb{R}^d \ni (\mu, x) \mapsto \tilde{\delta}_\mu m(\mu, x) \in \mathcal{M}^+(\mathbb{R}^d),$$

bounded in total variation, continuous for the product topology, $\mathcal{M}^+(\mathbb{R}^d)$ equipped with the weak topology, and such that for all μ and μ' in $\mathcal{M}^+(\mathbb{R}^d)$, it holds:

$$\langle m(\mu') - m(\mu), \psi\rangle = \int_0^1 \int_{\mathbb{R}^d} \langle \tilde{\delta}_\mu m\left(t\mu' + (1-t)\mu, x\right), \psi\rangle [\mu' - \mu](\mathrm{d}x)\,\mathrm{d}t, \tag{2.4}$$

for every $\psi \in \mathrm{C}_\mathrm{b}(\mathbb{R}^d)$. We call $\tilde{\mathrm{C}}^1(\mathcal{M}^+(\mathbb{R}^d))$ the class of functions from $\mathcal{M}^+(\mathbb{R}^d)$ to $\mathcal{M}^+(\mathbb{R}^d)$ that are differentiable in linear functional sense and, analogously, we denote by $\tilde{\mathrm{C}}^k(\mathcal{M}^+(\mathbb{R}^d))$ the space of functions that are k times differentiable.

Remark 2.1.13. The joint continuity required in Definition 2.1.12 implies that, for every fixed $\mu \in \mathcal{M}^+(\mathbb{R}^d)$, the mapping $x \mapsto \tilde{\delta}_\mu m(\mu, x)$ is measurable and so $\tilde{\delta}_\mu u$ is a transition kernel.

Remark 2.1.14. It easy to see that if $m\colon \mathcal{M}^+(\mathbb{R}^d) \to \mathcal{M}^+(\mathbb{R}^d)$ is in $\tilde{\mathrm{C}}^1(\mathcal{M}^+(\mathbb{R}^d))$ then the mapping $\mu \mapsto \langle m(\mu), \psi\rangle$ is in $\mathrm{C}^1(\mathcal{M}^+(\mathbb{R}^d))$ for every $\psi \in \mathrm{C}_\mathrm{b}(\mathbb{R}^d)$. In particular, $\langle \tilde{\delta}_\mu m(\mu, x), \psi\rangle = \delta_\mu(\langle m(\cdot), \psi\rangle)(\mu, x)$ for every $\psi \in \mathrm{C}_\mathrm{b}(\mathbb{R}^d)$, $\mu \in \mathcal{M}^+(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$. The converse is also true if we assume that the regularity of the mapping $\mu \mapsto \langle m(\mu), \psi\rangle$ is uniform with respect to ψ.

Example 2.1.15. Let us consider $\rho \in \mathrm{C}_\mathrm{b}(\mathbb{R}^d; [0, +\infty))$, and let us define $m_\rho\colon \mathcal{M}^+(\mathbb{R}^d) \to \mathcal{M}^+(\mathbb{R}^d)$ as

$$\langle m_\rho(\mu), \psi\rangle = \langle \rho\mu, \psi\rangle = \int_{\mathbb{R}^d} \psi(x)\rho(x)\mu(\mathrm{d}x), \quad \psi \in \mathrm{C}_\mathrm{b}(\mathbb{R}^d).$$

From Definition 2.1.12 it holds that $\tilde{\delta}_\mu m(\mu, x) = \rho(x)\delta_x$, where δ_x is the Dirac measure in $x \in \mathbb{R}^d$. Moreover, for any $\mu \in \mathcal{M}^+(\mathbb{R}^d)$, $x, y \in \mathbb{R}^d$, we have $\tilde{\delta}^2_\mu m(\mu, x, y) = 0$.

Example 2.1.16. Let us consider $f \in \mathrm{C_b}(\mathbb{R}^d;\mathbb{R}^d)$ and let us define the mapping $m\colon \mathcal{M}^+(\mathbb{R}^d) \to \mathcal{M}^+(\mathbb{R}^d)$ as the push-forward measure through f, namely

$$\mathcal{M}^+(\mathbb{R}^d) \ni \mu \mapsto m(\mu) = f_\# \mu \in \mathcal{M}^+(\mathbb{R}^d).$$

We recall that for every $\psi \in \mathrm{C_b}(\mathbb{R}^d)$ it holds

$$\int_{\mathbb{R}^d} \psi(y) f_\#\mu(\,\mathrm{d}y) = \int_{\mathbb{R}^d} \psi(f(x))\mu(\,\mathrm{d}x).$$

Thus, we have $\tilde{\delta}_\mu m(\mu, x) = \delta_{f(x)}$, where $\delta_{f(x)}$ is the Dirac measure centered in $f(x)$, $x \in \mathbb{R}^d$.

2.1.1 Some properties

We list here some properties which will be required in the next sections and that help us to understand how these derivatives can be combined. A first property we need concerns the symmetry of the second-order derivatives.

Proposition 2.1.17. *Let $u\colon \mathcal{M}^+(\mathbb{R}^d) \to \mathbb{R}$ be of class $\mathrm{C}^2_{\mathrm{L}}(\mathcal{M}^+(\mathbb{R}^d))$. The following facts hold:*

i. the second-order linear functional derivative is symmetric in the spatial arguments, that is $\delta^2_\mu u(\mu, x, y) = \delta^2_\mu u(\mu, y, x)$ for every $x, y \in \mathbb{R}^d$ and $\mu \in \mathcal{M}^+(\mathbb{R}^d)$;

ii. $\mathrm{D}_x(\delta^2_\mu u(\mu, x, y)) = \delta_\mu \left(\mathrm{D}_\mu u(\cdot, x)\right)(\mu, y)$;

iii. $\mathrm{D}_\mu(\mathrm{D}_\mu u(\cdot, x))(\mu, y) = \mathrm{D}^2_\mu u(\mu, x, y)$.

Proof. The proof follows the one of [30, Lemma 2.2.4], without relevant modifications in the argument. □

Remark 2.1.18. Thanks to Proposition 2.1.17, we can characterize the second-order L-derivative as the L-derivative of the mapping $\mu \mapsto \mathrm{D}_\mu u(\mu, x)$, for every fixed $x \in \mathbb{R}^d$, as we do for the linear functional case. We also notice that the first property in Proposition 2.1.17 holds more generally when $u \in \mathrm{C}^2(\mathcal{M}^+(\mathbb{R}^d))$.

Remark 2.1.19. If we consider functions over $\mathcal{P}(\mathbb{R}^d)$, the Schwarz-type identity *i.* in Proposition 2.1.17 is not true, but it holds with an additional correction term (see [30, Lemma 2.2.4]). In particular, it holds that

$$\delta^2_\mu u(\mu, x, y) = \delta_\mu u^2(\mu, y, x) + \delta_\mu u(\mu, x) - \delta_\mu u(\mu, y), \quad \mu \in \mathcal{P}(\mathbb{R}^d), x, y \in \mathbb{R}^d.$$

We can notice that the correction terms disappear if we integrate with respect to $\mu \otimes \mu$.

The next proposition shows how to compute the linear functional derivative of the composition of a function from $\mathbb{R}$ to $\mathbb{R}$ and a function from $\mathcal{M}^+(\mathbb{R}^d)$ to $\mathbb{R}$, by a chain rule similar to the classical one.

Proposition 2.1.20. *Let $h \in \mathrm{C}^1_{\mathrm{b}}(\mathbb{R})$ and $g \in \mathrm{C}^1(\mathcal{M}^+(\mathbb{R}^d))$. Then the composition*

$$\mathcal{M}^+(\mathbb{R}^d) \ni \mu \mapsto h(g(\mu)) \in \mathbb{R}$$

is in $\mathrm{C}^1(\mathcal{M}^+(\mathbb{R}^d))$ and the following chain rule holds:

$$\delta_\mu h(g(\cdot))(\mu, x) = h'(g(\mu))\delta_\mu g(\mu, x), \quad x \in \mathbb{R}^d, \mu \in \mathcal{M}^+(\mathbb{R}^d).$$

Proof. For every $\mu, \mu' \in \mathcal{M}^+(\mathbb{R}^d)$ we have

$$h(g(\mu')) - h(g(\mu)) = \int_0^1 \frac{\partial}{\partial t} h(g(\mu_t))\,\mathrm{d}t = \int_0^1 h'\left(g(\mu_t)\right)\frac{\partial}{\partial t} g(\mu_t)\,\mathrm{d}t,$$

where $\mu_t = t\mu' + (1-t)\mu$. If we show that

$$\frac{\partial}{\partial t} g(\mu_t) = \int_{\mathbb{R}^d} \delta_\mu g(\mu_t, x)[\mu' - \mu](\,\mathrm{d}x),$$

then we are done. For h fixed, since $g \in \mathrm{C}^1(\mathcal{M}^+(\mathbb{R}^d))$ we can compute the increment

$$\frac{1}{h}\left(g(\mu_{t+h}) - g(\mu_t)\right) = \frac{1}{h}\int_0^1 \delta_\mu g(\tau h\mu' - \tau h\mu + \mu_t, x) h[\mu' - \mu](\,\mathrm{d}x)\,\mathrm{d}\tau.$$

Then by taking $h \to 0$ we get $\frac{\partial}{\partial t} g(\mu_t) = \int_{\mathbb{R}^d} \delta_\mu g(\mu_t, x)[\mu' - \mu](\,\mathrm{d}x)$, where we used the dominated convergence theorem and the joint continuity and boundedness of $\delta_\mu g$. □

Remark 2.1.21. An easy generalization holds for $h(g_1(\mu), \ldots, g_n(\mu))$, where $n \in \mathbb{N}$, $h \in \mathrm{C}^1_{\mathrm{b}}(\mathbb{R}^n)$ and $\{g_i\}_{i=1}^n \subset \mathrm{C}^1(\mathcal{M}^+(\mathbb{R}^d))$. Then $h \in \mathrm{C}^1(\mathcal{M}^+(\mathbb{R}^d))$ and it holds:

$$\delta_\mu h(g_1(\cdot), \ldots, g_n(\cdot))(\mu, x) = \sum_{i=1}^n \partial_i h(g_1(\mu), \ldots, g_n(\mu))\delta_\mu g_i(\mu, x), \quad x \in \mathbb{R}^d, \mu \in \mathcal{M}^+(\mathbb{R}^d).$$

Another natural result we would like to have is the chain rule for the composition between functions from $\mathcal{M}^+(\mathbb{R}^d)$ to $\mathbb{R}$ and from $\mathcal{M}^+(\mathbb{R}^d)$ to $\mathcal{M}^+(\mathbb{R}^d)$.

Proposition 2.1.22. *Let $m \in \tilde{\mathrm{C}}^1(\mathcal{M}^+(\mathbb{R}^d))$ and let $g \in \mathrm{C}^1(\mathcal{M}^+(\mathbb{R}^d))$. Then the composition mapping $\mathcal{M}^+(\mathbb{R}^d) \ni \mu \mapsto g(m(\mu)) \in \mathbb{R}$ is in $\mathrm{C}^1(\mathcal{M}^+(\mathbb{R}^d))$ and it holds:*

$$\delta_\mu g(m(\cdot))(\mu, x) = \langle \tilde{\delta}_\mu m(\mu, x), \delta_\mu g(\cdot)(m(\mu))\rangle.$$

Moreover, if $m \in \tilde{\mathrm{C}}^2(\mathcal{M}^+(\mathbb{R}^d))$ and $g \in \mathrm{C}^2(\mathcal{M}^+(\mathbb{R}^d))$, then $g(m(\cdot)) \in \mathrm{C}^2(\mathcal{M}^+(\mathbb{R}^d))$ with

$$\delta^2_\mu g(m(\cdot))(\mu, x, y) = \langle \tilde{\delta}^2_\mu m(\mu, x, y), \delta_\mu g(\cdot)(m(\mu))\rangle + \langle \tilde{\delta}_\mu m(\mu, x) \otimes \tilde{\delta}_\mu m(\mu, y), \delta^2_\mu g(\cdot)(m(\mu))\rangle.$$

Proof. For every $\mu, \mu' \in \mathcal{M}^+(\mathbb{R}^d)$ we have

$$g(m(\mu')) - g(m(\mu)) = \int_0^1 \frac{\partial}{\partial t} g(m(\mu_t))\,\mathrm{d}t,$$

where $\mu_t = t\mu' + (1-t)\mu$. Then, by the regularity of g and m follows that

$$\frac{\partial}{\partial t} g(m(\mu_t)) = \int_{\mathbb{R}^d} \langle \tilde{\delta}_\mu m(\mu_t, x), \delta_\mu g(\cdot)(m(\mu_t))\rangle[\mu' - \mu](\,\mathrm{d}x), \tag{2.5}$$

and so the thesis. In the same way, one can deduce the result for the second-order derivative. □

Finally, we state a proposition regarding the differentiation of products. We omit the proof since it is analogue to the two above.

Proposition 2.1.23. *Let $f, g \in \mathrm{C}^1(\mathcal{M}^+(\mathbb{R}^d))$. Then the product map*

$$\mathcal{M}^+(\mathbb{R}^d) \ni \mu \mapsto f(\mu)g(\mu) \in \mathbb{R}$$

is in $\mathrm{C}^1(\mathcal{M}^+(\mathbb{R}^d))$ and the following product rule holds:

$$(\delta_\mu fg)(\mu, x) = f(\mu)\delta_\mu g(\mu, x) + g(\mu)\delta_\mu f(\mu, x).$$

Moreover, if $m \in \tilde{\mathrm{C}}^1(\mathcal{M}^+(\mathbb{R}^d))$ and if $\psi\colon \mathbb{R}^d \times \mathcal{M}^+(\mathbb{R}^d) \to \mathbb{R}$ is bounded, of class $\mathrm{C}^1(\mathcal{M}^+(\mathbb{R}^d))$ in the measure argument and continuous in the spatial argument, then the mapping

$$\mu \ni \mathcal{M}^+(\mathbb{R}^d) \mapsto h(\mu) := \langle m(\mu), \psi(\cdot, \mu)\rangle \in \mathbb{R}$$

is in $\mathrm{C}^1(\mathcal{M}^+(\mathbb{R}^d))$ and it holds

$$\delta_\mu h(\mu, x) = \langle \tilde{\delta}_\mu m(\mu, x), \psi(\cdot, \mu)\rangle + \langle m(\mu), \delta_\mu \psi(\cdot, \mu, x)\rangle.$$

2.1.2 The $\mathcal{P}(K)$ case

Let $K \subset \mathbb{R}^d$ be compact and let us consider the space $\mathcal{P}(K)$, which is compact with respect to the weak topology (see for instance [101]). In this section we consider mappings from $\mathcal{P}(K)$ to $\mathbb{R}$. For this particular case, we will rewrite the definitions presented in Section 2.1, together with some remarks. In particular, the purpose of this section is to fix the notation in view of the discussion in Chapter 5, where we will consider only measures in $\mathcal{P}(K)$.

Definition 2.1.24. *We say that a function $u\colon \mathcal{P}(K) \to \mathbb{R}$ has linear functional derivative if there exists a map*

$$\delta_\mu u\colon \mathcal{P}(K) \times K \to \mathbb{R},$$

which is jointly continuous and such that for every $\mu, \nu \in \mathcal{P}(K)$ it holds

$$u(\mu) - u(\nu) = \int_0^1 \int_K \delta_\mu u(\theta\mu + (1-\theta)\nu, x)[\mu - \nu](\,\mathrm{d}x)\,\mathrm{d}\theta. \tag{2.6}$$

We denote by $\mathrm{C}^1(\mathcal{P}(K))$ the space of functions that are differentiable in linear functional sense.

We recall that in this case the linear functional derivative is defined up to an additive constant. Following Remark 2.1.8, we choose the version such that, for every $\mu \in \mathcal{P}(K)$,

$$\int_K \delta_\mu u(\mu, x)\mu(\,\mathrm{d}x) = 0.$$

Remark 2.1.25. As in Remark 2.1.2, by iterating Definition 2.1.24 (keeping the extra spatial variable in $\delta_\mu u$ fixed), one can introduce the space $\mathrm{C}^2(\mathcal{P}(K))$ and analogously the space $\mathrm{C}^k(\mathcal{P}(K))$, $k \in \mathbb{N}$.

Definition 2.1.26. *We say that a function u is in $\mathrm{C}^2_{\mathrm{L}}(\mathcal{P}(K))$ if:*

i. $u \in \mathrm{C}^2(\mathcal{P}(K))$*;*

ii. for every $\mu \in \mathcal{P}(K)$*, the maps* $K \ni x \mapsto \delta_\mu u(\mu, x)$ *and* $K \times K \ni (x, y) \mapsto \delta^2_\mu u(\mu, x, y)$ *are twice continuously differentiable, with jointly continuous derivatives over* $\mathcal{P}(K) \times K$ *and* $\mathcal{P}(K) \times K \times K$*.*

We define the first and second-order L-derivatives as

$$\mathrm{D}_\mu u(\mu, x) := \mathrm{D}_x \delta_\mu u(\mu, x) \in \mathbb{R}^d, \quad \mathrm{D}^2_\mu u(\mu, x, y) := \mathrm{D}_x \mathrm{D}^\top_y \delta^2_\mu u(\mu, x, y) \in \mathbb{R}^{d\times d},$$

for every $\mu \in \mathcal{P}(K)$ *and* $x, y \in K$*.*

Remark 2.1.27. Differently from the definitions in Section 2.1, in this case the boundedness of the derivatives is guaranteed by the joint continuity, since we are considering functions over compact sets.

2.2 Approximation of real-valued functions over the space of positive measures

Here we discuss how to approximate real-valued functions over $\mathcal{M}^+(\mathbb{R}^d)$ with a class of simpler functions, which allows easier and explicit computations. Our technique is based on a construction which is well known for function over space of probability measures, in particular on $\mathcal{P}_2(\mathbb{R}^d)$ endowed with the Wasserstein metric. Given $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, we can introduce its empirical approximation

$$\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}, \quad n \geq 1, \tag{2.7}$$

where $\{X_i\}_{i=1}^n$ are independent identically distributed (i.i.d.) random variables with law μ (over an arbitrary probability space $(\Omega, \mathcal{F}, \mathbb{P})$). It can be shown, see for instance [31, Section 5.1.2], that $\mathcal{W}_2(\mu, \mu_n) \to 0$ almost surely and in $L^2(\Omega)$ as $n \to +\infty$. If $u\colon \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$, we can introduce its empirical projection $u^n(\mu) := u(\mu_n)$ and if u is bounded and continuous with respect to $\mathcal{W}_2$ one can conclude that, for every $\mu \in \mathcal{P}_2(\mathbb{R}^d)$,

$$\mathbb{E}\left[u^n(\mu)\right] = \mathbb{E}\left[u(\mu_n)\right] = \langle \mu^{\times n}, u\left(\frac{1}{n}\sum_{i=1}^n \delta_{\cdot_i}\right)\rangle \to u(\mu), \quad n \to +\infty, \tag{2.8}$$

where we used the notation

$$\langle \mu^{\times n}, u\left(\frac{1}{n}\sum_{i=1}^n \delta_{\cdot_i}\right)\rangle = \int_{\mathbb{R}^{dn}} u\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\right)\mu(\,\mathrm{d}x_1)\dots\mu(\,\mathrm{d}x_n).$$

We can set $\phi^n(\mu) = \langle \mu^{\times n}, u\left(\frac{1}{n}\sum_{i=1}^n \delta_{\cdot_i}\right)\rangle$ and thus the family $\{\phi^n\}_{n\geq 1}$ approximates pointwise u.

The goal of our approximation technique is to find a class of functions with good properties that allows us to approximate functions over positive measures together with their derivatives, when they exist.

The first step in our procedure is to adapt the previous argument to a space of finite positive measures. Let us introduce, for $k > 1$,

$$\mathcal{M}^+_{2,k}(\mathbb{R}^d) := \left\{ \mu \in \mathcal{M}^+_2(\mathbb{R}^d) \colon \mu(\mathbb{R}^d) \in \left[\frac{1}{k}, k\right]\right\}.$$

In the end, we will be able to approximate a function $u\colon \mathcal{M}^+_{2,k}(\mathbb{R}^d) \to \mathbb{R}$, where $k > 1$ is fixed, in a way that allow us to approximate also its derivatives, when $u \in \mathrm{C}^2_{\mathrm{L}}(\mathcal{M}^+_{2,k}(\mathbb{R}^d))$. We can define, for every $\mu \in \mathcal{M}^+_{2,k}(\mathbb{R}^d)$,

$$u^n(\mu) := u\left(\frac{\mu(\mathbb{R}^d)}{n}\sum_{i=1}^n \delta_{X_i}\right), \quad n \geq 1, \tag{2.9}$$

where $\{X_i\}_{i=1}^n$ are i.i.d. random variables with law $\mu/\mu(\mathbb{R}^d)$. If we fix $\mu \in \mathcal{M}^+_{2,k}(\mathbb{R}^d)$ and we ask the mapping $\mathcal{P}_2(\mathbb{R}^d) \ni \pi \mapsto u\left(\mu(\mathbb{R}^d)\pi\right) \in \mathbb{R}$ to be bounded and continuous in $\mathcal{W}_2$, then we can conclude, as for (2.8), that

$$\phi^n(\mu) := \mathbb{E}\left[u^n(\mu)\right] = \frac{1}{\mu(\mathbb{R}^d)^n}\langle \mu^{\times n}, u\left(\frac{\mu(\mathbb{R}^d)}{n}\sum_{i=1}^n \delta_{\cdot_i}\right)\rangle \to u\left(\mu(\mathbb{R}^d)\frac{\mu}{\mu(\mathbb{R}^d)}\right) = u(\mu), \tag{2.10}$$

as $n \to +\infty$.

Remark 2.2.1. Notice that if $u\colon \mathcal{M}^+_{2,k}(\mathbb{R}^d) \to \mathbb{R}$ is continuous with respect to the weak topology, then the mapping $\mathcal{P}_2(\mathbb{R}^d) \ni \pi \mapsto u\left(\mu(\mathbb{R}^d)\pi\right) \in \mathbb{R}$ is continuous in $\mathcal{W}_2$ for every fixed $\mu \in \mathcal{M}^+_{2,k}(\mathbb{R}^d)$. Indeed if $\mathcal{W}_2(\pi, \pi') \to 0$, then $\langle \mu(\mathbb{R}^d)\pi, \psi\rangle \to \langle \mu(\mathbb{R}^d)\pi', \psi\rangle$ for every $\psi \in \mathrm{C}_{\mathrm{b}}(\mathbb{R}^d)$ and then one conclude thanks to the continuity of u.

Remark 2.2.2. It is easy to show that the mapping

$$\mathbb{R}^{d\times n} \ni (x_1, \dots, x_n) \mapsto \tilde{u}^n(x_1, \dots, x_n) := u\left(\frac{\mu(\mathbb{R}^d)}{n}\sum_{i=1}^n \delta_{x_i}\right)$$

is in $\mathrm{C}^2_{\mathrm{b}}(\mathbb{R}^{d\times n})$ if $u \in \mathrm{C}^2_{\mathrm{L}}(\mathcal{M}^+(\mathbb{R}^d))$. Indeed, let $h > 0$ and $k \in \{1, \dots, n\}$, then

$$\begin{aligned}\frac{1}{h}&\left[\tilde{u}^n(x_1, \dots, x_k + h, \dots, x_n) - \tilde{u}^n(x_1, \dots, x_k, \dots, x_n)\right] \\ &= \frac{\mu(\mathbb{R}^d)}{n}\int_0^1 \frac{\delta_\mu u\left(m^{x_k}_{\theta,h}, x_k + h\right) - \delta_\mu u\left(m^{x_k}_{\theta,h}, x_k\right)}{h}\,\mathrm{d}\theta,\end{aligned}$$

with $m^{x_1}_{\theta,h} := \frac{\mu(\mathbb{R}^d)}{n}\left(\sum_{i\neq k}^n \delta_{x_i} + \theta\delta_{x_k+h} + (1-\theta)\delta_{x_k}\right)$. If we let $h \to 0$, we obtain

$$\partial_k \tilde{u}^n(x_1, \dots, x_n) = \frac{\mu(\mathbb{R}^d)}{n}\mathrm{D}_\mu u\left(\frac{\mu(\mathbb{R}^d)}{n}\sum_{i=1}^n \delta_{x_i}, x_k\right),$$

which is continuous and bounded thanks to the regularity of $\mathrm{D}_\mu u$. We can proceed analogously for the second order derivatives. Moreover, if we consider $k > 1$ and $u \in \mathrm{C}^2_{\mathrm{L}}(\mathcal{M}^+_{2,k}(\mathbb{R}^d))$, then the mapping

$$\mathbb{R}^{d\times n} \times \left[\frac{1}{k}, k\right] \ni (x_1, \dots, x_n, z) \mapsto \bar{u}^n(x_1, \dots, x_n, z) \coloneqq u\left(\frac{z}{n}\sum_{i=1}^n \delta_{x_i}\right)$$

is in $\mathrm{C}^2_{\mathrm{b}}(\mathbb{R}^{d\times n} \times \left[\frac{1}{k}, k\right])$. Indeed, let h be an admissible increment, then

$$\frac{1}{h}\left[\bar{u}^n(x_1, \dots, x_n, z+h) - \bar{u}^n(x_1, \dots, x_n, z)\right] \\ = \frac{1}{n}\sum_{k=1}^d \int_0^1 \delta_\mu u\left(\frac{z+\theta h}{n}\sum_{i=1}^d \delta_{x_i}, x_k\right) \mathrm{d}\theta \to \frac{1}{n}\sum_{k=1}^d \delta_\mu u\left(\frac{z}{n}\sum_{i=1}^d \delta_{x_i}, x_k\right),$$

as $h \to 0$, which is continuous and bounded thanks to the regularity of $\delta_\mu u$. As before, we can proceed analogously for the second order derivatives.

The set $\mathcal{M}^+_{2,k}(\mathbb{R}^d)$ is not compact with respect to the weak topology and in the further computations this will give rise to some problems. So, we want to restrict to the case in which the measures are in a compact subset of $\mathcal{M}^+_{2,k}(\mathbb{R}^d)$. Let us introduce the family of compact rectangles $\{K_N = [-N, N]^d\}_{N\geq 1} \subset \mathbb{R}^d$. Then, for every $N \geq 1$, we define

$$\mathcal{H}^k_N \coloneqq \left\{\mu \in \mathcal{M}^+_{2,k}(\mathbb{R}^d) \colon \operatorname{supp}\mu \subset K_N\right\},$$

which is compact in the weak topology. Given $\mu \in \mathcal{M}^+_{2,k}(\mathbb{R}^d)$, we denote with $\rho^N\mu$ the measure such that $\frac{\mathrm{d}\rho^N\mu}{\mathrm{d}\mu} = \rho^N$, with ρ^N positive and smooth cut-off function equal to 1 in K_N and identically zero outside K_{N+1}. We notice that $\rho^N\mu$ is always in $\mathcal{H}^k_{N+1}$. Thus, for every $N \geq 1$ we set

$$\hat{u}^N(\mu) \coloneqq u(\rho^N\mu), \quad \mu \in \mathcal{M}^+_{2,k}(\mathbb{R}^d). \tag{2.11}$$

In the following lemma, we show how we can approximate a function $u\colon \mathcal{M}^+_{2,k}(\mathbb{R}^d) \to \mathbb{R}$ with $\{\hat{u}^N\}_{N\geq 1}$.

Lemma 2.2.3. *Let $k > 1$, let u be in $\mathrm{C}^2_{\mathrm{L}}(\mathcal{M}^+_{2,k}(\mathbb{R}^d))$ and let the sequence $\{\hat{u}^N\}_{N\geq 1}$ be defined by (2.11). Then, for every $\mu \in \mathcal{M}^+_{2,k}(\mathbb{R}^d)$, $\hat{u}^N(\mu) \to u(\mu)$ as $N \to +\infty$ and $\{\delta_\mu \hat{u}^N\}_{N\geq 1}$, $\{\delta^2_\mu \hat{u}^N\}_{N\geq 1}$, $\{\mathrm{D}_\mu \hat{u}^N\}_{N\geq 1}$, $\{\mathrm{D}^2_\mu \hat{u}^N\}_{N\geq 1}$, $\{\mathrm{D}_x\mathrm{D}_\mu \hat{u}^N\}_{N\geq 1}$, $\{\mathrm{D}_x\delta^2_\mu \hat{u}^N\}_{N\geq 1}$ pointwise converge to the respective derivatives of u. Moreover, $\|\hat{u}^N\|_\infty \leq \|u\|_\infty$, and there exists $C > 0$ independent of u, N, k such that*

$$\|\delta_\mu \hat{u}^N\| \leq \|\delta_\mu u\|, \qquad \|\delta^2_\mu \hat{u}^N\| \leq \|\delta^2_\mu u\|,$$
$$\|\mathrm{D}_\mu \hat{u}^N\| \leq C(\|\mathrm{D}_\mu u\| + \|\delta_\mu u\|), \qquad \|\partial_x\mathrm{D}_\mu \hat{u}^N\| \leq C(\|\mathrm{D}_\mu u\| + \|\delta_\mu u\| + \|\mathrm{D}_x\mathrm{D}_\mu u\|),$$
$$\|\mathrm{D}_x\delta^2_\mu \hat{u}^N\| \leq C(\|\delta^2_\mu u\| + \|\mathrm{D}_x\delta^2_\mu u\|), \qquad \|\mathrm{D}^2_\mu \hat{u}^N\| \leq C(\|\delta^2_\mu u\| + \|\mathrm{D}_x\delta^2_\mu u\| + \|\mathrm{D}^2_\mu u\|),$$

where the norms are meant as $\|\cdot\|_\infty$, that is the supremum norm over all the arguments of the function.

Proof. First, we notice that $\rho^N\mu \to \mu$ weakly as $N \to +\infty$. Indeed, for every $\varphi \in \mathrm{C_b}(\mathbb{R}^d)$, we have

$$\langle \rho^N\mu, \varphi\rangle = \int_{\mathbb{R}^d} \rho^N(x)\varphi(x)\mu(\,\mathrm{d}x) \to \int_{\mathbb{R}^d} \varphi(x)\mu(\,\mathrm{d}x),$$

as $n \to +\infty$, thanks to dominated convergence. Thus, since u is continuous with respect to the weak topology, we have that for every $\mu \in \mathcal{M}^+_{2,k}(\mathbb{R}^d)$, $\hat{u}^N(\mu) \to u(\mu)$ as $N \to +\infty$. Moreover, it is immediate that $\|\hat{u}^N\|_\infty \le \|u\|_\infty$.

Regarding the linear functional derivatives, from Proposition 2.1.22 with $m(\mu) = \rho^N\mu$ and $g(\mu) = u(\mu)$, it follows that for every $\mu \in \mathcal{M}^+_{2,k}(\mathbb{R}^d)$ and $x, y \in \mathbb{R}^d$

$$\begin{aligned} \delta_\mu \hat{u}^N(\mu, x) &= \rho^N(x)\delta_\mu u(\rho^N\mu, x) \to \delta_\mu u(\mu, x), && \text{as } N \to +\infty,\\ \delta^2_\mu \hat{u}^N(\mu, x, y) &= \rho^N(x)\rho^N(y)\delta^2_\mu u(\rho^N\mu, x, y) \to \delta^2_\mu u(\mu, x, y), && \text{as } N \to +\infty, \end{aligned}$$

thanks to the continuity of $\delta_\mu u$ and $\delta^2_\mu u$ and the fact that $\rho^N(x) \to 1$ as $N \to +\infty$, for every $x \in \mathbb{R}^d$. As before, the estimates on the norms easily follows.

To conclude, it remains to show the convergence of the derivatives in space of $\delta_\mu u$ and $\delta^2_\mu u$. For instance, we have that, for every $\mu \in \mathcal{M}^+_{2,k}(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$

$$\begin{aligned} \mathrm{D}_\mu \hat{u}^N(\mu, x) &= \mathrm{D}_x\delta_\mu \hat{u}^N(\mu, x)\\ &= \mathrm{D}_x\rho^N(x)\delta_\mu u(\rho^N\mu, x) + \rho^N(x)\mathrm{D}_\mu u(\rho^N\mu, x) \to \mathrm{D}_\mu u(\mu, x), \quad N \to +\infty. \end{aligned}$$

Indeed, the first summand tends to 0 since $\operatorname{supp} \mathrm{D}_x\rho^N \subset [N, N+1]$ and $\delta_\mu u$ is bounded, whilst the second one tends to $\mathrm{D}_\mu u(\mu, x)$ thanks to the continuity of $\mathrm{D}_\mu u$ and the fact that $\rho^N(x) \to 1$ as $N \to +\infty$, for every $x \in \mathbb{R}^d$. In a similar way, we get

$$\begin{aligned} \mathrm{D}_x\mathrm{D}_\mu \hat{u}^N(\mu, x) =& \mathrm{D}^2_x\rho^N(x)\delta_\mu u(\rho^N\mu, x)\\ &+ 2\mathrm{D}_x\rho^N(x)\mathrm{D}_\mu u(\rho^N\mu, x) + \rho^N(x)\mathrm{D}_x\mathrm{D}_\mu u(\rho^N\mu, x), \end{aligned}$$

$$\mathrm{D}_x\delta^2_\mu \hat{u}^N(\mu, x, y) = \rho^N(y)\mathrm{D}_x\rho^N(x)\delta^2_\mu u(\rho^N\mu, x, y) + \rho^N(y)\rho^N(x)\mathrm{D}_x\delta^2_\mu u(\rho^N\mu, x, y),$$

$$\begin{aligned} \mathrm{D}^2_\mu \hat{u}^N(\mu, x, y) =& \mathrm{D}_x\rho^N(x)\mathrm{D}_y\rho^N(y)^\top\delta^2_\mu u(\rho^N\mu, x, y)\\ &+ \rho^N(y)\mathrm{D}_x\rho^N(x)\mathrm{D}_y\delta^2_\mu u(\rho^N\mu, x, y)^\top\\ &+ \rho^N(x)\mathrm{D}_x\delta^2_\mu u(\rho^N\mu, x, y)\mathrm{D}_y\rho^N(y)^\top\\ &+ \rho^N(y)\rho^N(x)\mathrm{D}^2_\mu u(\rho^N\mu, x, y), \end{aligned}$$

and then we can prove the convergence of the remaining derivatives as before.

Regarding the estimates on the norm, it follows that

$$\begin{aligned} \|\mathrm{D}_\mu \hat{u}^N\| &\le C(\|\mathrm{D}_\mu u\| + \|\delta_\mu u\|), & \|\partial_x\mathrm{D}_\mu \hat{u}^N\| &\le C(\|\mathrm{D}_\mu u\| + \|\delta_\mu u\| + \|\partial_x\mathrm{D}_\mu u\|),\\ \|\partial_x\delta^2_\mu \hat{u}^N\| &\le C(\|\delta^2_\mu u\| + \|\partial_x\delta^2_\mu u\|), & \|\mathrm{D}^2_\mu \hat{u}^N\| &\le C(\|\delta^2_\mu u\| + \|\partial_x\delta^2_\mu u\| + \|\mathrm{D}^2_\mu u\|), \end{aligned}$$

where $C = \max\left\{1, 2\|\partial_x\rho^N\|, \|\partial_x^2\rho^N\|, \|\partial_x\rho^N(\partial_y\rho^N)^\top\|\right\}$ and all the norms are meant as $\|\cdot\|_\infty$. Notice that C can be chosen independent of N, k and u, thanks to the particular structure of ρ^N. $\square$

Thanks to Lemma 2.2.3, we can approximate functions in $\mathrm{C}^2_{\mathrm{L}}(\mathcal{M}^+_{2,k}(\mathbb{R}^d))$ with functions in $\mathrm{C}^2_{\mathrm{L}}(\mathcal{H}^k_N)$. Since we need more regular approximants, our goal now is to show that if $u \in \mathrm{C}^2_{\mathrm{L}}(\mathcal{H}^{\mathrm{k}}_{\mathrm{N}})$, with $k > 1$ and $N \geq 1$ fixed, then its derivatives can be approximated by the corresponding derivatives of the sequence $\{\phi^n\}_{n\geq 1}$ introduced in (2.10), namely

$$\phi^n(\mu) := \mathbb{E}\left[u^n(\mu)\right] = \frac{1}{\mu(\mathbb{R}^d)^n}\langle \mu^{\times n}, u\left(\frac{\mu(\mathbb{R}^d)}{n}\sum_{i=1}^n \delta_{\cdot_i}\right)\rangle, \quad \mu \in \mathcal{H}^k_N.$$

Lemma 2.2.4. *Let u be in $\mathrm{C}^2_{\mathrm{L}}(\mathcal{H}^{\mathrm{k}}_{\mathrm{N}})$, for $k > 1$ and $N \geq 1$ fixed, and let $\{\phi^n\}_{n\geq 1}$ be defined by (2.10). Then, for every $\mu \in \mathcal{H}^k_N$, $\phi^n(\mu) \to u(\mu)$ as $n \to +\infty$ and $\{\delta_\mu\phi^n\}_{n\geq 1}$, $\{\delta^2_\mu\phi^n\}_{n\geq 1}$, $\{\mathrm{D}_\mu\phi^n\}_{n\geq 1}$, $\{\mathrm{D}^2_\mu\phi^n\}_{n\geq 1}$, $\{\mathrm{D}_x\mathrm{D}_\mu\phi^n\}_{n\geq 1}$, $\{\mathrm{D}_x\delta^2_\mu\phi^n\}_{n\geq 1}$ converge pointwise to the respective derivatives of u. Moreover, $\|\phi^n\|_\infty \leq \|u\|_\infty$ and the same holds for the derivatives, up to a multiplicative constant independent of u, N and k.*

Remark 2.2.5. Thanks to Lemma 2.2.4, we can approximate functions in $\mathrm{C}^2_{\mathrm{L}}(\mathcal{H}^k_N)$ with functions $\phi^n\colon \mathcal{H}^k_N \to \mathbb{R}$ of the form

$$\phi^n(\mu) = \langle \frac{\mu^{\times n}}{\mu(\mathbb{R}^d)^n}, \varphi_n(\cdot, \dots, \cdot, \mu(\mathbb{R}^d))\rangle,$$

where $\varphi_n \in \mathrm{C}^2_{\mathrm{b}}(K^n_N \times \left[\frac{1}{k}, k\right])$ is symmetric in the first n arguments and $K_N = [-N, N]^d$.

Proof. First, $u\colon \mathcal{H}^k_N \to \mathbb{R}$ is absolutely continuous with respect to the weak topology and so it is also for the mapping $\mathcal{P}_2(K_N) \ni \pi \mapsto u(\mu(\mathbb{R}^d)\pi)$, with $\mu \in \mathcal{H}^k_N$. Moreover, since we are considering measures over a compact subset of $\mathbb{R}^d$, the absolute continuity of this last mapping also holds in $\mathcal{W}_2$. The same consideration is true also for $\delta_\mu u$, $\mathrm{D}_\mu u$ and $\mathrm{D}_x\mathrm{D}_\mu u$ (resp. $\delta^2_\mu u$, $\mathrm{D}_x\delta^2_\mu u$ and $\mathrm{D}^2_\mu u$) when restricted to $\mathcal{H}^k_N \times K_N$ (resp. $\mathcal{H}^k_N \times K_N \times K_N$).

We already showed the convergence of $\phi^n(\mu)$ to $u(\mu)$ for every $\mu \in \mathcal{H}^k_N$. Let us discuss the convergence of the linear functional derivatives of ϕ^n. First, we notice that $\phi^n(\mu) = f(\mu)g(\mu)$ with

$$g(\mu) := \frac{1}{\mu(\mathbb{R}^d)^n}, \quad f(\mu) := \langle \mu^{\times n}, u\left(\frac{\mu(\mathbb{R}^d)}{n}\sum_{i=1}^n \delta_{\cdot_i}\right)\rangle, \quad \mu \in \mathcal{H}^k_N.$$

From Proposition 2.1.20 it follows that $\delta_\mu g(\mu, x) = -\frac{n}{\mu(\mathbb{R}^d)^{n+1}}$, whilst from Proposition 2.1.23 combined with 2.1.22 we get

$$\begin{aligned}&\delta_\mu f(\mu, x)\\ &= n\langle \mu^{\times(n-1)}, u\left(\frac{\mu(\mathbb{R}^d)}{n}\sum_{i=1}^{n-1}\delta_{\cdot_i} + \frac{\mu(\mathbb{R}^d)}{n}\delta_x\right)\rangle + \frac{1}{n}\sum_{j=1}^n\langle \mu^{\times n}, \delta_\mu u\left(\frac{\mu(\mathbb{R}^d)}{n}\sum_{i=1}^n \delta_{\cdot_i}, \cdot_j\right)\rangle.\end{aligned}$$

Thus, from Proposition 2.1.23, we have that for $(\mu, x) \in \mathcal{H}_N^k \times K_N$,

$$\delta_\mu \phi^n(\mu, x) = \frac{n}{\mu(\mathbb{R}^d)^n} \langle \mu^{\times(n-1)}, u\left(\frac{\mu(\mathbb{R}^d)}{n}\sum_{i=1}^{n-1}\delta_{\cdot_i} + \frac{\mu(\mathbb{R}^d)}{n}\delta_x\right)\rangle$$
$$- \frac{n}{\mu(\mathbb{R}^d)^{n+1}}\langle \mu^{\times n}, u\left(\frac{\mu(\mathbb{R}^d)}{n}\sum_{i=1}^{n}\delta_{\cdot_i}\right)\rangle$$
$$+ \frac{1}{n}\sum_{j=1}^{n}\langle\left(\frac{\mu}{\mu(\mathbb{R}^d)}\right)^{\times n}, \delta_\mu u\left(\frac{\mu(\mathbb{R}^d)}{n}\sum_{i=1}^{n}\delta_{\cdot_i}, \cdot_j\right)\rangle,$$

which can be written by using the random variables $\{X_i\}_{i=1}^n$ as

$$\begin{aligned}
&\delta_\mu \phi^n(\mu, x) \\
&\quad = \frac{n}{\mu(\mathbb{R}^d)}\left\{\mathbb{E}\left[u\left(\frac{\mu(\mathbb{R}^d)}{n}\sum_{i=1}^{n-1}\delta_{X_i} + \frac{\mu(\mathbb{R}^d)}{n}\delta_x\right)\right] - \mathbb{E}\left[u\left(\frac{\mu(\mathbb{R}^d)}{n}\sum_{i=1}^{n}\delta_{X_i}\right)\right]\right\} \\
&\qquad + \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\left[\delta_\mu u\left(\frac{\mu(\mathbb{R}^d)}{n}\sum_{i=1}^{n}\delta_{X_i}, X_j\right)\right]. \qquad (2.12)
\end{aligned}$$

Regarding the first two terms of (2.12), since u is differentiable in linear functional sense, we get that their difference is equal to

$$\begin{aligned}
&\mathbb{E}\left[\int_0^1 \delta_\mu u\left(\frac{\mu(\mathbb{R}^d)}{n}\sum_{i=1}^{n-1}\delta_{X_i} + \theta\frac{\mu(\mathbb{R}^d)}{n}\delta_x + (1-\theta)\frac{\mu(\mathbb{R}^d)}{n}\delta_{X_n}, x\right) \mathrm{d}\theta\right] \\
&\quad - \mathbb{E}\left[\int_0^1 \delta_\mu u\left(\frac{\mu(\mathbb{R}^d)}{n}\sum_{i=1}^{n-1}\delta_{X_i} + \theta\frac{\mu(\mathbb{R}^d)}{n}\delta_x + (1-\theta)\frac{\mu(\mathbb{R}^d)}{n}\delta_{X_n}, X_n\right) \mathrm{d}\theta\right],
\end{aligned}$$

which tends to $\delta_\mu u(\mu, x) - \langle \frac{\mu}{\mu(\mathbb{R}^d)}, \delta_\mu u(\mu, \cdot)\rangle$, thanks to dominated convergence and the regularity of the linear functional derivative. On the other hand, the last therm in (2.12) converges to $\langle \frac{\mu}{\mu(\mathbb{R}^d)}, \delta_\mu u(\mu, \cdot)\rangle$ (which coincides, for instance, with $\mathbb{E}\left[\delta_\mu u(\mu, X_1)\right]$). Indeed,

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\left[\delta_\mu u\left(\frac{\mu(\mathbb{R}^d)}{n}\sum_{i=1}^{n}\delta_{X_i}, X_j\right)\right] - \mathbb{E}\left[\delta_\mu u(\mu, X_1)\right]\right| \\
&\quad \le \left|\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}\left[\delta_\mu u\left(\frac{\mu(\mathbb{R}^d)}{n}\sum_{i=1}^{n}\delta_{X_i}, X_j\right)\right] - \frac{1}{n}\mathbb{E}\left[\sum_{j=1}^{n}\delta_\mu u(\mu, X_j)\right]\right| \\
&\qquad + \left|\frac{1}{n}\mathbb{E}\left[\sum_{j=1}^{n}\delta_\mu u(\mu, X_j)\right] - \mathbb{E}\left[\delta_\mu u(\mu, X_1)\right]\right|,
\end{aligned}$$

then the first term tends to zero thanks to the uniform continuity of $\delta_\mu u$ and the second one tends to zero due to the law of large numbers. From (2.12) it also follows that $\|\delta_\mu \phi^n\|_\infty \le 3\|\delta_\mu u\|_\infty$. Similarly we are able to show the pointwise convergence of $\delta_\mu^2\phi^n$ and the estimate for $\|\delta_\mu^2\phi^n\|_\infty$.

Now we study the convergence of the L-derivatives of ϕ^n, by differentiating (2.12) with respect to x. The second and the third term do not depend on x, whilst for the first one we can compute the increment:

$$\frac{1}{h}\left[\delta_\mu\phi^n(\mu, x+he_k) - \delta_\mu\phi^n(\mu,x)\right] = \mathbb{E}\left[\int_0^1 \frac{\delta_\mu u\left(m_{\theta,h}^x, x+he_k\right) - \delta_\mu u\left(m_{\theta,h}^x, x\right)}{h}\,\mathrm{d}\theta\right],$$

where $h \in \mathbb{R}$, $k \in \{1,\dots,d\}$, e_k is the k-th vector of the canonical basis of $\mathbb{R}^d$ and

$$m_{\theta,h}^x := \frac{\mu(\mathbb{R}^d)}{n}\left(\sum_{i=1}^{n-1}\delta_{X_i} + \theta\delta_{x+he_k} + (1-\theta)\delta_x\right).$$

Then we can pass to the limit as $h \to 0$ and thanks to the uniform continuity and the differentiability of $\delta_\mu u$ we get

$$\mathrm{D}_\mu\phi^n(\mu,x) = \mathrm{D}_x\delta_\mu\phi^n(\mu,x) = \mathbb{E}\left[\mathrm{D}_x\delta_\mu u\left(\frac{\mu(\mathbb{R}^d)}{n}\sum_{i=1}^{n-1}\delta_{X_i} + \frac{\mu(\mathbb{R}^d)}{n}\delta_x, x\right)\right],$$

which tends to $\mathrm{D}_\mu u(\mu,x)$ as $n \to +\infty$ and from which we can also deduce $\|D_m u\phi^n\|_\infty \le \|\mathrm{D}_\mu u\|_\infty$. Again, with similar computations one can show the pointwise convergence and the norms estimates for $\{\mathrm{D}_x\delta_\mu^2\phi^n\}_{n\ge1}$, $\{\mathrm{D}_x\mathrm{D}_\mu\phi^n\}_{n\ge1}$, $\{\mathrm{D}_\mu^2\phi^n\}_{n\ge1}$.

□

A further approximation can be achieved by substituting φ in Remark 2.2.5 with a certain family of symmetric polynomials. The obtained function has a more regular structure and is easier to study. We call functions of this type cylindrical and they represent the last step in our approximation.

Definition 2.2.6 (Cylindrical functions)**.** *We say that $u : \mathcal{M}^+(\mathbb{R}^d) \to \mathbb{R}$ is a cylindrical function of order $k \in \mathbb{N}$ if there exist $n \in \mathbb{N}$, $g \in \mathrm{C}_\mathrm{b}^\mathrm{k}(\mathbb{R}^n)$ and $\{\psi_i\}_{i=1}^n \subset \mathrm{C}_\mathrm{b}^k(\mathbb{R}^d)$ such that*

$$u(\mu) = g\left(\langle\mu,\psi_1\rangle,\dots,\langle\mu,\psi_n\rangle\right), \quad \forall\,\mu \in \mathcal{M}^+(\mathbb{R}^d).$$

We denote with $\mathcal{C}_k(\mathcal{M}^+(\mathbb{R}^d))$ the set of cylindrical functions of order k.

In order to study the convergence in the next Lemma, we consider the following norm: let $N \ge 1$, $k > 1$ and let $u \in \mathrm{C}_\mathrm{L}^2(\mathcal{H}_N^k)$, then

$$\|u\|_{\mathrm{C}_\mathrm{L}^2(\mathcal{H}_N^k)} := \sup_{\substack{\mu\in\mathcal{H}_N^k,\\ x,y\in K_N}} \Big(|u(\mu)| + |\delta_\mu u(\mu,x)| + |\delta_\mu^2 u(\mu,x,y)| + |\mathrm{D}_x\delta_\mu^2 u(\mu,x,y)|$$
$$+ |\mathrm{D}_\mu u(\mu,x)| + |\mathrm{D}_x\mathrm{D}_\mu u(\mu,x)| + |\mathrm{D}_\mu^2 u(\mu,x,y)|\Big), \quad (2.13)$$

where we recall that $K_N = [-N,N]^d$.

Lemma 2.2.7. *Let $N \ge 1$ and $k > 1$ be fixed. Let $f : \mathcal{H}_N^k \to \mathbb{R}$ be defined, for $r \in \mathbb{N}$, as $f(\mu) := \langle \frac{\mu^{\times r}}{\mu(\mathbb{R}^d)^r}, \varphi(\cdot,\dots,\cdot,\mu(\mathbb{R}^d))\rangle$, where $\varphi \in \mathrm{C}_\mathrm{b}^2\left(K_N^r \times \left[\frac{1}{k},k\right]\right)$ is symmetric in the first r arguments. Then, there exists a sequence $\{f^n\}_{n\ge1} \subset \mathcal{C}_2(\mathcal{M}^+(\mathbb{R}^d))$ such that $\|f - f^n\|_{\mathrm{C}_\mathrm{L}^2(\mathcal{H}_\mathrm{N}^\mathrm{k})} \to 0$ as $n \to +\infty$.*

Proof. Since $K_N^r \times \left[\frac{1}{k}, k\right] \subset \mathbb{R}^{dr+1}$ is compact, we can find a family of symmetrical polynomials $\{\varphi_n\}_{n\geq 1}$ which approximates φ in C^2 norm (we can choose a symmetric version of Bernstein approximants, see for instance [25, Section 3.2]). More precisely,

$$\varphi_n(x_1, \dots, x_r, z) = \sum_{i=1}^{\ell(n)} h_i(z) \prod_{j=1}^{r} g_{i,j}(x_j),$$

where $\ell(n) \in \mathbb{N}$, $g_{i,j} \colon K_N \to \mathbb{R}$ and $h_i \colon \left[\frac{1}{k}, k\right] \to \mathbb{R}$ are monomials, for every $i = 1, \dots, \ell(n)$ and $j = 1, \dots, r$. Let us introduce $f^n(\mu) \coloneqq \langle \frac{\mu^{\times r}}{\mu(\mathbb{R}^d)^r}, \varphi_n(\cdot, \dots, \cdot, \mu(\mathbb{R}^d)\rangle$, for every $n \geq 1$. This sequence is in $\mathcal{C}_2(\mathcal{M}^+(\mathbb{R}^d))$, indeed, for every $n \geq 1$,

$$f^n(\mu) = g^n(\langle \mu, g_{1,1}\rangle, \dots, \langle \mu, g_{\ell(n),r}\rangle, \mu(\mathbb{R}^d)),$$

where g^n is defined over $\mathbb{R}^{dr\ell(n)} \times \mathbb{R}$ as

$$g^n(\xi_{1,1}, \dots, \xi_{\ell(n),r}, \zeta) = \sum_{i=1}^{\ell(n)} \zeta^{e(i)-1} \prod_{j=1}^{r} \xi_{i,j},$$

where we denoted by $e(i)$ the exponents in the monomials h_i. This function g^n is symmetric, twice continuously differentiable but not bounded. However, we can consider the product with a smooth symmetric (in the first r variables) cut-off function, equal to 1 in the rectangle $[-R-1, R+1]^{dr\ell(n)} \times \left[\frac{1}{k}, k\right]$, where

$$R \coloneqq \max\{|\langle \mu, g_{1,1}\rangle|, \dots, |\langle \mu, g_{\ell(n),r}\rangle|\},$$

and vanishing smoothly outside. This function, that we denote again with g^n for simplicity, is symmetric and in $\mathrm{C}^2_{\mathrm{b}}(\mathbb{R}^{\mathrm{d}r\ell(\mathrm{n})} \times \mathbb{R})$. We can notice that at this point we needed the fact that $\mu(\mathbb{R}^d) \geq 1/k$, which was included in the definition of $\mathcal{H}_N^k$, and not only $\mu(\mathbb{R}^d) > 0$. Indeed, we need $\varphi \in \mathrm{C}^2_{\mathrm{b}}\left(K_N^r \times \left[\frac{1}{k}, k\right]\right)$ in order to introduce the sequence $\{\varphi_n\}_{n\geq 1}$ and consequently $\{f^n\}_{n\geq 1}$.

Let us study the convergence of $\{f^n\}_{n\geq 1}$ in norm $\|\cdot\|_{\mathrm{C}^2_{\mathrm{L}}(\mathcal{H}_N^k)}$. First, it holds that $\sup_{\mu\in\mathcal{H}_N^k} |f^n(\mu) - f(\mu)| \to 0$ as $n \to +\infty$, thanks to the uniform convergence of $\{\varphi^n\}_{n\geq 1}$ and the bound on $\mu(\mathbb{R}^d)$. Regarding the first-order linear functional derivative, we have that for every $\mu \in \mathcal{H}_N^k$ and $x \in K_N$,

$$\begin{aligned}\delta_\mu f^n(\mu, x) = r\langle \frac{\mu^{\times(r-1)}}{\mu(\mathbb{R}^d)^r}, \varphi_n(\cdot_1, \dots, \cdot_{r-1}, x, \mu(\mathbb{R}^d))\rangle + \langle \frac{\mu^{\times r}}{\mu(\mathbb{R}^d)^r}, \partial_z \varphi^n(\cdot, \dots, \cdot, \mu(\mathbb{R}^d))\rangle \\ - r\langle \frac{\mu^{\times r}}{\mu(\mathbb{R}^d)^{r+1}}, \varphi^n(\cdot, \dots, \cdot, \mu(\mathbb{R}^d))\rangle. \quad (2.14)\end{aligned}$$

Due to dominated convergence, expression (2.14) converges as $n \to +\infty$ to the same one with φ instead of φ^n, which is $\delta_\mu f$. Analogously, $\delta^2_\mu f^n(\mu, x, y) \to \delta_\mu f^n(\mu, x, y)$ as $n \to +\infty$ for every $\mu \in \mathcal{H}_N^k$ and $x, y \in K_N$. Moreover, since φ_n and its derivatives up to order two converge uniformly to φ and its derivatives, we have that

$$\sup_{\substack{\mu\in\mathcal{H}_N^k,\\ x\in K_N}} |\delta_\mu f^n(\mu, x) - \delta_\mu f(\mu, x)| \to 0, \qquad \sup_{\substack{\mu\in\mathcal{H}_N^k,\\ x,y\in K_N}} \left|\delta^2_\mu f^n(\mu, x, y) - \delta^2_\mu f(\mu, x, y)\right| \to 0,$$

as $n \to +\infty$. To conclude the discussion on the convergence, we are able to bring the spatial derivative inside the integral in the first row of (2.14) and so for every $\mu \in \mathcal{H}_N^k$ and $x \in \mathbb{R}^d$

$$\begin{aligned}\mathrm{D}_\mu f^n(\mu, x) = r\langle \frac{\mu^{\times(r-1)}}{\mu(\mathbb{R}^d)^r}, \mathrm{D}_x\varphi_n(\cdot_1, \ldots, \cdot_{r-1}, x, \mu(\mathbb{R}^d))\rangle \\ \to r\langle \frac{\mu^{\times(r-1)}}{\mu(\mathbb{R}^d)^r}, \mathrm{D}_x\varphi(\cdot_1, \ldots, \cdot_{r-1}, x, \mu(\mathbb{R}^d))\rangle = \mathrm{D}_\mu f(\mu, x),\end{aligned}$$

as $n \to +\infty$. As before, the convergence takes place also uniformly since the first-order derivative of φ_n converges uniformly to the one of φ. In the same way we can show the uniform convergence over $\mathcal{H}_N^k \times K_N$ (or $\mathcal{H}_N^k \times K_N \times K_N$) of $\{\mathrm{D}_\mu^2 f^n\}_{n\geq 1}$, $\{\mathrm{D}_x\mathrm{D}_\mu f^n\}_{n\geq 1}$, $\{\mathrm{D}_x\delta_\mu^2 f^n\}_{n\geq 1}$ to $\mathrm{D}_\mu^2 f$, $\mathrm{D}_x\mathrm{D}_\mu f$, $\mathrm{D}_x\delta_\mu^2 f$ respectively. □

Remark 2.2.8. We can notice that both the results of Lemma 2.2.4 and Lemma 2.2.7 work for functions over $\mathcal{P}_2(K_N)$, and the approximants remain well defined over $\mathcal{P}_2(K_N)$. Regarding Lemma 2.2.3, we can not simply cut the measure support, since the mass must be kept normalized. In this case, we can approximate $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ with the family of probabilities obtained by concentrating all the mass outside the compact K_N in the origin, namely $\{\lambda_1^N\mu + \lambda_2^N\delta_0\}_{N\geq 1}$, where λ_1^N and λ_2^N are smooth positive cut-off functions, $\lambda_1^N = 0$ in K_{N+1}, $\lambda_2^N = 0$ in K_N^{c} and $\lambda_1^N + \lambda_2^N = 1$. This sequence converges weakly to $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and the approximation properties in Lemma 2.2.3 still hold for $\{\hat{u}^N\}_{N\geq 1} \coloneqq \{u(\lambda_1^N\mu + \lambda_2^N\delta_0)\}_{N\geq 1}$.

Remark 2.2.9. From the proofs of Lemma 2.2.3, we can see that if we ask for $u \in \mathrm{C}^2(\mathcal{M}_{2,\mathrm{k}}^+(\mathbb{R}^\mathrm{d}))$, then the approximation for u and its first- and second-order linear functional derivatives still holds. The same applies also for Lemma 2.2.4.

Remark 2.2.10. At the beginning of this section we used the fact that a function u on $\mathcal{M}^+(\mathbb{R}^d)$ can be regarded as a function defined on $\mathbb{R}_+ \times \mathcal{P}(\mathbb{R}^d)$ through the formula

$$v(\lambda, \pi) = u(\lambda\pi), \quad \lambda > 0,\ \pi \in \mathcal{P}(\mathbb{R}^d),$$

and conversely

$$u(\mu) = v\left(\mu(\mathbb{R}^d), \frac{\mu}{\mu(\mathbb{R}^d)}\right), \quad \mu \in \mathcal{M}^+(\mathbb{R}^d).$$

Probably this fact could be used to deduce some of the results we proved for functions over $\mathcal{M}^+(\mathbb{R}^d)$ (see for instance Section 2.1.1) directly from earlier results for functions over $\mathcal{P}(\mathbb{R}^d)$. However, we decided to give the proofs directly in the context of $\mathcal{M}^+(\mathbb{R}^d)$ since in this case the computations are more straightforward and since this is the natural framework to study the Zakai equation.

Chapter 3

Itô formulas for filtering processes

In this chapter, we present two Itô formulas concerning the solutions of the filtering equations. In the recent literature, formulas for the composition of a probability measure-valued process and a real-valued function have been recently obtained, especially in a context related to mean field games and McKean-Vlasov equations (see for instance [24, 31, 32]). In all these cases, the considered measure-valued process is the law of a stochastic process, and this feature is usually exploited in the proofs. Here we follow a different approach, based on the fact that the measure-valued processes we are considering satisfy an explicit stochastic differential equation (namely the Zakai or the Kushner-Stratonovich equation). To take advantage of this, we first prove the result for the family of cylindrical functions, introduced in the previous chapter. Then, we obtain a general result by approximation.

We first focus on the Zakai equation, since the fact that it describes the evolution of a positive measure-valued process represents a major novelty. Then, we present analog results for the Kushner-Stratonovich equation, and we discuss briefly some similarities and links with the problem of mean filed games with common noise.

The results contained in this section hold under more general conditions than Assumption 1. In particular, it will be only necessary to assume:

Assumption 2. *All the mappings $f, \sigma, \bar{\sigma}, h$ are taken Borel-measurable and bounded.*

We also recall that the operators A, B, introduced in Section 1.2, are defined for every $\psi \in \mathrm{C}^2_\mathrm{b}(\mathbb{R}^d)$, $x \in \mathbb{R}^d$, by

$$\begin{aligned} A\psi(x) &\coloneqq \sum_{i=1}^d f_i(x)\partial_i\psi(x) + \frac{1}{2}\sum_{i,j=1}^d \left\{ \left(\sigma\sigma^\top\right)_{ij}(x)\partial_{ij}\psi(x) + \left(\bar{\sigma}\bar{\sigma}^\top\right)_{ij}(x)\partial_{ij}\psi(x) \right\}, \\ B_k\psi(x) &\coloneqq \sum_{i=1}^d \bar{\sigma}_{ik}(x)\partial_i\psi(x), \quad k = 1, \dots d. \end{aligned} \tag{3.1}$$

3.1 Itô formula for the Zakai equation

In this section, we consider a complete filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{Q})$ endowed with a $\{\mathcal{F}_t\}$-Brownian motion Y, and a solution to the Zakai equation ρ on that probabilistic setup. In particular

$$\langle \rho_t, \psi\rangle = \langle \mu, \psi\rangle + \int_0^t \langle \rho_s, A\psi\rangle \,\mathrm{d}s + \int_0^t \langle \rho_s, h\psi + B\psi\rangle \cdot \mathrm{d}Y_s, \quad \psi \in \mathrm{C}^2_\mathrm{b}(\mathbb{R}^d), \tag{3.2}$$

where the coefficients are related to the filtering problem (1.1)-(1.2) and the operators A and B are defined by (5.3). Finally, we will consider only initial conditions μ in $\mathcal{M}^+_2(\mathbb{R}^d)$, so ρ is a $\mathcal{M}^+_2(\mathbb{R}^d)$-valued process.

The aim of this section is to identify an Itô formula for the composition of a regular function and a process that solves (3.2). This is a key step for our final purpose, that is to write and study the backward Kolmogorov equation associated to the Zakai equation.

Before stating the main result of this section, we need to introduce a notation for the integral with respect to the product measure $\mu \otimes \mu$, with $\mu \in \mathcal{M}^+(\mathbb{R}^d)$. Let $f, g \in \mathrm{C}_\mathrm{b}(\mathbb{R}^d)$, $h \in \mathrm{C}_\mathrm{b}(\mathbb{R}^d \times \mathbb{R}^d)$. We adopt the following notations:

- Every time $\mu(\mathrm{d}x) \otimes \mu(\mathrm{d}y)$ integrates a product of functions over $\mathbb{R}^d$, it is meant that the first one is integrated with respect to $\mu(\mathrm{d}x)$ and the second one with respect to $\mu(\mathrm{d}y)$, that is

$$\langle \mu \otimes \mu, fg\rangle := \int\int f(x)g(y)\mu(\mathrm{d}x)\mu(\mathrm{d}y).$$

  In particular, $\langle \mu \otimes \mu, f\ f\rangle = \int\int f(x)f(y)\mu(\mathrm{d}x)\mu(\mathrm{d}y)$;

- Every time $\mu(\mathrm{d}x) \otimes \mu(\mathrm{d}y)$ integrates a product of functions over $\mathbb{R}^d$ and a function over $\mathbb{R}^d \times \mathbb{R}^d$, it is meant that the first one is integrated with respect to $\mu(\mathrm{d}x)$, the second one with respect to $\mu(\mathrm{d}y)$ and the third with respect to both, that is

$$\langle \mu \otimes \mu, fgh\rangle := \int\int f(x)g(y)h(x,y)\mu(\mathrm{d}x)\mu(\mathrm{d}y).$$

The extension to vector-valued and matrix-valued functions is straightforward.

Proposition 3.1.1. *Let $\rho = \{\rho_t, t \in [0,T]\}$ be a solution to the Zakai equation starting at $\mu \in \mathcal{M}^+_2(\mathbb{R}^d)$ and let u be in $\mathrm{C}^2_\mathrm{L}(\mathcal{M}^+_2(\mathbb{R}^d))$. Let also Assumption 2 holds.*

Then the following Itô formula holds:

$$
\begin{aligned}
u(\rho_t) = u(\mu) &+ \int_0^t \langle \rho_s, \mathrm{D}_\mu u(\rho_s) \cdot f \rangle \, \mathrm{d}s \\
&+ \int_0^t \frac{1}{2} \langle \rho_s, \mathrm{tr}\left\{ \mathrm{D}_x \mathrm{D}_\mu u(\rho_s) \sigma \sigma^\top \right\} \rangle \, \mathrm{d}s + \int_0^t \frac{1}{2} \langle \rho_t, \mathrm{tr}\left\{ \mathrm{D}_x \mathrm{D}_\mu u(\rho_s) \bar{\sigma} \bar{\sigma}^\top \right\} \rangle \, \mathrm{d}s \\
&+ \int_0^t \langle \rho_s, h \delta_\mu u(\rho_s) \rangle \cdot \mathrm{d}Y_s + \int_0^t \langle \rho_s, \bar{\sigma}^\top \mathrm{D}_\mu u(\rho_s) \rangle \cdot \mathrm{d}Y_s \qquad (3.3) \\
&+ \int_0^t \frac{1}{2} \langle \rho_s \otimes \rho_s, \delta^2_\mu u(\rho_s) h \cdot h \rangle \, \mathrm{d}s + \int_0^t \langle \rho_s \otimes \rho_s, h \cdot \bar{\sigma}^\top \delta_\mu \mathrm{D}_\mu u(\rho_s) \rangle \, \mathrm{d}s \\
&+ \int_0^t \frac{1}{2} \langle \rho_s \otimes \rho_s, \mathrm{tr}\left\{ \mathrm{D}^2_\mu u(\rho_s) \bar{\sigma} \bar{\sigma}^\top \right\} \rangle \, \mathrm{d}s, \quad t \in [0, T],
\end{aligned}
$$

almost surely in Ω.

Proof. The proof in divided into five steps: the idea is to show the formula for cylindrical function over a compact subset of $\mathcal{M}^+_2(\mathbb{R}^d)$, which is a direct consequence of the classical Itô formula, and then achieve the result by approximation and localization. In particular, for the first four steps we assume that, for a fixed $k > 1$, $\rho_t \in \mathcal{M}^+_{2,k}(\mathbb{R}^d)$ for every $t \in [0, T]$, that is $\rho_t(\mathbb{R}^d) \in [\frac{1}{k}, k]$ for every $t \in [0, T]$, almost surely. In the last step we get rid of this condition by a localization argument.

First step. We prove the formula for $u \in \mathcal{C}_2(\mathcal{M}^+(\mathbb{R}^d))$. More precisely, $u(\mu) = g(\langle \mu, \psi_1 \rangle, \dots, \langle \mu, \psi_n \rangle)$, $n \geq 1$, $g \in \mathrm{C}^2_\mathrm{b}(\mathbb{R}^\mathrm{n})$, $\{\psi_i\}_{i=1}^n \subset \mathrm{C}^2_\mathrm{b}(\mathbb{R}^\mathrm{d})$. Without loss of generality, we discuss the case $n = 1$, that is $u(\mu) = g(\langle \mu, \psi \rangle)$. The result with $n \geq 1$ is obtained with the same procedure. By the classical Itô formula and (3.2), we get that

$$
\begin{aligned}
\mathrm{d}u(\rho_t) = \ & \mathrm{d}g(\langle \rho_t, \psi \rangle) \\
&= g'(\langle \rho_t, \psi \rangle) \langle \rho_t, A\psi \rangle \, \mathrm{d}t + g'(\langle \rho_t, \psi \rangle) \langle \rho_t, h\psi + B\psi \rangle \cdot \mathrm{d}Y_t \\
&\quad + \frac{1}{2} g''(\langle \rho_t, \psi \rangle) \langle \rho_t, h\psi + B\psi \rangle \cdot \langle \rho_t, h\psi + B\psi \rangle \, \mathrm{d}t \\
&= g'(\langle \rho_t, \psi \rangle) \langle \rho_t, \mathrm{D}_x \psi \cdot f \rangle \, \mathrm{d}t + \frac{1}{2} g'(\langle \rho_t, \psi \rangle) \langle \rho_t, \mathrm{tr}\left\{ \mathrm{D}^2_x \psi \sigma \sigma^\top \right\} \rangle \, \mathrm{d}t \\
&\quad + \frac{1}{2} g'(\langle \rho_t, \psi \rangle) \langle \rho_t, \mathrm{tr}\left\{ \mathrm{D}^2_x \psi \bar{\sigma} \bar{\sigma}^\top \right\} \rangle \, \mathrm{d}t + g'(\langle \rho_t, \psi \rangle) \langle \rho_t, h\psi + \bar{\sigma}^\top \mathrm{D}_x \psi \rangle \cdot \mathrm{d}Y_t \\
&\quad + \frac{1}{2} g''(\langle \rho_t, \psi \rangle) \langle \rho_t, h\psi + \bar{\sigma}^\top \mathrm{D}_x \psi \rangle \cdot \langle \rho_t, h\psi + \bar{\sigma}^\top \mathrm{D}_x \psi \rangle) \, \mathrm{d}t.
\end{aligned}
$$

Then, recalling Example 2.1.11, we get that

$$
\begin{aligned}
&\mathrm{d}u(\rho_t) \\
&= \langle \rho_t, \mathrm{D}_\mu u(\rho_t) \cdot f \rangle \, \mathrm{d}t + \frac{1}{2} \langle \rho_t, \mathrm{tr}\left\{ \mathrm{D}_x \mathrm{D}_\mu u(\rho_t) \sigma \sigma^\top \right\} \rangle \, \mathrm{d}t + \frac{1}{2} \langle \rho_t, \mathrm{tr}\left\{ \mathrm{D}_x \mathrm{D}_\mu u(\rho_t) \bar{\sigma} \bar{\sigma}^\top \right\} \rangle \, \mathrm{d}t \\
&\quad + \langle \rho_t, h \delta_\mu u(\rho_t) \rangle \cdot \mathrm{d}Y_t + \langle \rho_t, \bar{\sigma}^\top \mathrm{D}_\mu u(\rho_t) \rangle \cdot \mathrm{d}Y_t + \frac{1}{2} \langle \rho_t \otimes \rho_t, \delta^2_\mu u(\rho_t) h \cdot h \rangle \, \mathrm{d}t \\
&\quad + \langle \rho_t \otimes \rho_t, h \cdot \bar{\sigma}^\top \delta_\mu \mathrm{D}_\mu u(\rho_t) \rangle \, \mathrm{d}t + \frac{1}{2} \langle \rho_t \otimes \rho_t, \mathrm{tr}\left\{ \mathrm{D}^2_\mu u(\rho_t) \bar{\sigma} \bar{\sigma}^\top \right\} \rangle \, \mathrm{d}t, \qquad (3.4)
\end{aligned}
$$

for every $t \in [0,T]$, almost surely in Ω.

Second step. Let us fix $N \geq 1$ and $k > 1$. Now we show the formula for functions of the form $u(\mu) = \langle \frac{\mu^r}{\mu(\mathbb{R}^d)^r}, \varphi(\cdot, \dots, \cdots, \mu(\mathbb{R}^d)) \rangle$, with $\mu \in \mathcal{H}_N^k$, $r \in \mathbb{N}$, $\varphi \in \mathrm{C}_\mathrm{b}^2(K_N^r \times \left[\frac{1}{k}, k\right])$ and φ symmetrical in the first r arguments. Thanks to Lemma 2.2.7, there exists $\{u^n\}_{n\geq 1} \subset \mathcal{C}_2(\mathcal{M}^+(\mathbb{R}^d))$ such that $\|u - u^n\|_{\mathrm{C}_\mathrm{L}^2(\mathcal{H}_N^k)} \to 0$ as $n \to +\infty$, where the norm has been introduced in (2.13). Thus, thanks to the first step, we get the formula (3.4) with u^n in place of u.

We study now the convergence of the terms in the expression we obtained. Since u^n converges to u for every $\mu \in \mathcal{H}_N^k$, we have that $u^n(\rho_t) \to u(\rho_t)$ and $u^n(\mu) \to u(\mu)$ almost surely in Ω, as $n \to +\infty$.

For the integrals in time, we can proceed by dominated convergence thanks to the convergence in $\|\cdot\|_{\mathrm{C}_\mathrm{L}^2, \mathcal{H}_\mathrm{N}^\mathrm{k}}$ norm of $\{u^n\}_{n\geq 1}$ and the boundedness of the coefficients $b, \sigma, \bar{\sigma}, h$. Here we discuss the convergence of $\int_0^t \frac{1}{2}\langle \rho_s \otimes \rho_s, \mathrm{tr}\left\{\mathrm{D}_\mu^2 u^n(\rho_s)\bar{\sigma}\bar{\sigma}^\top\right\}\rangle \,\mathrm{d}s$, but the other terms can be studied analogously. Thanks to Lemma 2.2.7, we have

$$\left|\mathrm{tr}\left\{D_\mu^2 u^n(\rho_s)\bar{\sigma}\bar{\sigma}^\top\right\} - \mathrm{tr}\left\{D_\mu^2 u(\rho_s)\bar{\sigma}\bar{\sigma}^\top\right\}\right| \leq 3\|\sigma\|_\infty^2 \|\mathrm{D}_\mu^2 u\|_\infty, \tag{3.5}$$

so by dominated convergence

$$\left|\langle \rho_s \otimes \rho_s, \mathrm{tr}\left\{D_\mu^2 u^n(\rho_s)\bar{\sigma}\bar{\sigma}^\top\right\} - \mathrm{tr}\left\{D_\mu^2 u(\rho_s)\bar{\sigma}\bar{\sigma}^\top\right\}\rangle\right| \to 0,$$

as $n \to +\infty$. Moreover, we have that for every $t \in [0,T]$

$$\left|\langle \rho_s \otimes \rho_s, \mathrm{tr}\left\{D_\mu^2 u^n(\rho_s)\bar{\sigma}\bar{\sigma}^\top\right\} - \mathrm{tr}\left\{D_\mu^2 u(\rho_s)\bar{\sigma}\bar{\sigma}^\top\right\}\rangle\right| \leq 3k\|\sigma\|_\infty^2 \|\mathrm{D}_\mu^2 u\|_\infty \in L^\infty([0,t]),$$

since $\rho_t(\mathbb{R}^d) \in \left[\frac{1}{k}, k\right]$. Thus, again by dominated convergence we can conclude that for any $t \in [0,T]$

$$\left|\int_0^t \frac{1}{2}\langle \rho_s \otimes \rho_s, \mathrm{tr}\left\{D_\mu^2 u^n(\rho_s)\bar{\sigma}\bar{\sigma}^\top\right\}\rangle \,\mathrm{d}s - \frac{1}{2}\int_0^t \langle \rho_s \otimes \rho_s, \mathrm{tr}\left\{D_\mu^2 u(\rho_s)\bar{\sigma}\bar{\sigma}^\top\right\}\rangle \,\mathrm{d}s\right| \to 0,$$

almost surely in Ω, as $n \to +\infty$.

For the stochastic integrals, we prove the convergence in $L^2(\Omega)$. Since the technique is the same for both the terms, let us focus on $\int_0^t \rho_s \left(h\delta_\mu u^n(\rho_s)\right) \cdot \mathrm{d}Y_s$. By Itô isometry, we have that for any $t \in [0,T]$ it holds

$$\mathbb{E}\left[\left(\int_0^t \left\{\langle \rho_s, h\delta_\mu u^n(\rho_s)\rangle - \langle \rho_s, h\delta_\mu u(\rho_s)\rangle\right\} \cdot \mathrm{d}Y_s\right)^2\right]$$
$$= \mathbb{E}\left[\int_0^t \left|\langle \rho_s, h\delta_\mu u^n(\rho_s)\rangle - \langle \rho_s, h\delta_\mu u(\rho_s)\rangle\right|^2 \mathrm{d}s\right] \to 0,$$

where the convergence is obtained thanks to the uniform convergence over $\mathcal{H}_N^k \times K_N$ of $\{\delta_\mu u^n\}_{n\geq 1}$ and the boundedness of h, combined with the dominated convergence argument we used for the deterministic integral.

Every convergence we proved implies the convergence in probability, so the relation

(3.3) holds almost everywhere in Ω, for every $t \in [0,T]$. Since both the right hand side and the left hand side of (3.3) are continuous , the relation holds for every $t \in [0,T]$ almost everywhere in Ω.

Third step. Let u be in $\mathrm{C}^2_{\mathrm{L}}(\mathcal{H}^k_N)$. Then by Lemma 2.2.4, there exists a sequence $\{\phi^n\}_{m\geq 1}$ which converges pointwise to u, and the same holds for the derivatives needed in the Itô formula. Moreover, as we pointed out in Remark 2.2.5,

$$\phi^n(\mu) = \langle \frac{\mu^{\times n}}{\mu(\mathbb{R}^d)^n}, \varphi_n(\cdot, \dots, \cdot, \mu(\mathbb{R}^d))\rangle,$$

with $\varphi_n \in \mathrm{C}^2_{\mathrm{b}}(K^n_N \times \left[\frac{1}{k}, k\right]), n \geq 1$. Thus, by step two, (3.3) holds for every ϕ^n. To conclude, we can pass to the limit with the same argument we used in step two, exploiting the bounds on the norms given by Lemma 2.2.4, the boundedness of $b, \sigma, \bar{\sigma}, h$, the fact that $\rho_t(\mathbb{R}^d) \in \left[\frac{1}{k}, k\right]$ for every $t \in [0,T]$ and the dominated convergence theorem.

Fourth step. Let u be in $\mathrm{C}^2_{\mathrm{L}}(\mathcal{M}^+_{2,\mathrm{k}}(\mathbb{R}^{\mathrm{d}}))$. Thanks to Lemma 2.2.3, we can conclude that (3.3) holds also for this class of functions, by the same argument we use in the previous steps.

Fifth step. Let us introduce the sequence of random times $\tau_k \colon \Omega \to [0, +\infty]$,

$$\tau_k = \left\{t \geq 0 : \rho_t(\mathbb{R}^d) \in [1/k, k]^{\mathrm{c}}\right\}, \quad k > 1.$$

First, $\{\tau_k\}_{k>1}$ are stopping times since they are exit time from a Borel set and moreover, thanks to Remark 1.3.4, $\tau_k \to +\infty$, $\mathbb{P}$ almost surely as $k \to +\infty$. Then, we can consider the stopped process $\rho^k_t := \rho_{t\wedge\tau_k}$ for which, thanks to te previous steps, (3.3) holds. Indeed it still satisfies the Zakai equation (3.2) and $\rho^k_t(\mathbb{R}^d) \in [1/k, k]$, for every $t \in [0,T]$ and $k > 1$. To conclude, we can let $k \to +\infty$ in the Itô formula for ρ^k_t, recovering (3.3) for ρ_t and $u \in \mathrm{C}^2_{\mathrm{L}}(\mathcal{M}^+_2(\mathbb{R}^d))$, thanks to the continuity in time of all the terms involved in the equation.

□

Remark 3.1.2. We can rewrite the formula (3.3) in the following way

$$\begin{aligned} \mathrm{d}u(\rho_t) = \langle \rho_t, A\delta_\mu u(\rho_t)\rangle \,\mathrm{d}t + \langle \rho_t, (h+B)\delta_\mu u(\rho_t)\rangle \cdot \mathrm{d}Y_t \\ + \frac{1}{2}\langle \rho_t \otimes \rho_t, (h+B)\cdot(h+B)\delta^2_\mu u(\rho_t)\rangle \,\mathrm{d}t, \quad (3.6) \end{aligned}$$

where

$$\begin{aligned} &\langle \rho_t \otimes \rho_t, (h+B)\cdot(h+B)\delta^2_\mu u(\rho_t)\rangle \\ &\quad = \iint (h(x) + \bar{\sigma}^\top(x)\mathrm{D}_x)\cdot(h(y) + \bar{\sigma}^\top(y)\mathrm{D}_y)\delta^2_\mu u(\rho_t, x, y)\rho_t(\,\mathrm{d}x)\rho_t(\,\mathrm{d}y). \end{aligned}$$

Corollary 3.1.3. *Assume that Assumption 2 holds, let $\rho = \{\rho_t, t \in [0,T]\}$ be a solution to the Zakai equation and let $u \colon [0,T] \times \mathcal{M}^+_2(\mathbb{R}^d) \to \mathbb{R}$ be in $\mathrm{C}^2_{\mathrm{L}}(\mathcal{M}^+_s(\mathbb{R}^d))$*

for the measure argument, in $\mathrm{C}^1([0,T])$ *for the time argument and let* u *and all its derivatives be bounded in all their arguments. Then it holds*

$$\begin{aligned} u(t,\rho_t) &= u(0,\mu) \\ &+ \int_0^t \partial_s u(s,\rho_s)\,\mathrm{d}s + \int_0^t \langle \rho_s, A\delta_\mu u(s,\rho_s)\rangle\,\mathrm{d}s + \int_0^t \langle \rho_s, (h+B)\delta_\mu u(s,\rho_s)\rangle \cdot \mathrm{d}Y_s \\ &+ \frac{1}{2}\int_0^t \langle \rho_s\otimes\rho_s, (h+B)\cdot(h+B)\delta_\mu^2 u(s,\rho_s)\rangle\,\mathrm{d}s, \quad t\in[0,T], \end{aligned} \tag{3.7}$$

almost surely.

Proof. The proof is basically the same of Proposition 3.1.1, with a standard modification in order to deal with the time dependence. □

3.2 Itô formula for Kushner-Stratonovich equation

Let us consider a weak solution $\{(\Omega,\mathcal{F},\{\mathcal{F}_t\}_{t\in[0,T]},\mathbb{P}),(\Pi,I)\}$ to the Kushner-Stratonovich equation starting at $\pi\in\mathcal{P}_2(\mathbb{R}^d)$. In particular, for every $t\in[0,T]$ and $\psi\in\mathrm{C}^2_{\mathrm{b}}(\mathbb{R}^d)$ it holds

$$\langle \Pi_t,\psi\rangle = \langle\pi,\psi\rangle + \int_0^t \langle \Pi_s, A\psi\rangle\,\mathrm{d}s + \int_0^t (\langle \Pi_s, h\psi + B\psi\rangle - \langle \Pi_s,\psi\rangle\langle\Pi_s,h\rangle)\cdot\mathrm{d}I_s. \tag{3.8}$$

As for the Zakai equation, is interesting to study the Itô formula for the composition of the process Π with a function $u\in\mathrm{C}^2_{\mathrm{L}}(\mathcal{P}_2(\mathbb{R}^d))$.

Proposition 3.2.1. *Let* $\{(\Omega,\mathcal{F},\{\mathcal{F}_t\}_{t\in[0,T]},\mathbb{P}),(\Pi_t,I_t)\}$ *be a weak solution to the Kushner-Stratonovich equation starting from* $\pi\in\mathcal{P}_2(\mathbb{R}^d)$ *and let* $u\in\mathrm{C}^2_{\mathrm{L}}(\mathcal{P}_2(\mathbb{R}^d))$. *Moreover, let Assumption 2 be in force. Then for every* $t\in[0,T]$ *it holds:*

$$\begin{aligned} u(\Pi_t) &= u(\pi) + \int_0^t \langle \Pi_s, \mathrm{D}_\mu u(\Pi_s)\cdot f\rangle\,\mathrm{d}s \\ &+ \int_0^t \frac{1}{2}\langle \Pi_s, \mathrm{tr}\left\{\mathrm{D}_x\mathrm{D}_\mu u(\Pi_s)\sigma\sigma^\top\right\}\rangle\,\mathrm{d}s + \int_0^t \frac{1}{2}\langle \Pi_t, \mathrm{tr}\left\{\mathrm{D}_x\mathrm{D}_\mu u(\Pi_s)\bar\sigma\bar\sigma^\top\right\}\rangle\,\mathrm{d}s \\ &+ \int_0^t \frac{1}{2}\langle \Pi_s\otimes\Pi_s, \delta_\mu^2 u(\Pi_s)h\cdot h\rangle\,\mathrm{d}s + \int_0^t \frac{1}{2}\langle \Pi_s\otimes\Pi_s, \mathrm{tr}\left\{\mathrm{D}_\mu^2 u(\Pi_s)\bar\sigma\bar\sigma^\top\right\}\rangle\,\mathrm{d}s \\ &+ \int_0^t \frac{1}{2}|\langle\Pi_s,h\rangle|^2\langle \Pi_s\otimes\Pi_s, \delta_\mu^2 u(\Pi_s)\rangle\,\mathrm{d}s + \int_0^t \langle \Pi_s\otimes\Pi_s, h\cdot\bar\sigma^\top\delta_\mu\mathrm{D}_\mu u(\Pi_s)\rangle\,\mathrm{d}s \\ &- \int_0^t \langle \Pi_s\otimes\Pi_s, \delta_\mu^2 u(\Pi_s)h\rangle\cdot\langle\Pi_s,h\rangle\,\mathrm{d}s - \int_0^t \langle \Pi_s\otimes\Pi_s, \bar\sigma^\top\delta_\mu\mathrm{D}_\mu u(\Pi_s)\rangle\cdot\langle\Pi_s,h\rangle\,\mathrm{d}s \\ &+ \int_0^t \langle \Pi_s, h\delta_\mu u(\Pi_s)\rangle\cdot\mathrm{d}I_s + \int_0^t \langle \Pi_s, \bar\sigma^\top\mathrm{D}_\mu u(\Pi_s)\rangle\cdot\mathrm{d}I_s - \int_0^t \langle \Pi_s, \delta_\mu u(\Pi_s)\rangle\langle\Pi_s,h\rangle\cdot\mathrm{d}I_s, \end{aligned} \tag{3.9}$$

almost surely.

Proof. The proof use the same approximation technique used in Proposition 3.1.1 proof. One has only to notice that the localization of the mass in the proof of Proposition 3.1.1 can be avoided, since $\Pi_t(\mathbb{R}^d) = 1$ for every $t \in [0,T]$, and the steps from one to four have to be done keeping in mind Remark 2.2.8. □

Remark 3.2.2. We can rewrite (3.9) as

$$\begin{aligned}\mathrm{d}u(\Pi_t) &= \langle \Pi_t, A\delta_\mu u(\Pi_t)\rangle\,\mathrm{d}t + \langle \Pi_t, (h - \langle \Pi_t, h\rangle + B)\delta_\mu u(\Pi_t)\rangle \cdot \mathrm{d}I_t \\ &\quad + \frac{1}{2}\langle \Pi_t \otimes \Pi_t, (h - \langle \Pi_t, h\rangle + B)\cdot(h - \langle \Pi_t, h\rangle + B)\delta^2_\mu u(\Pi_t)\rangle\,\mathrm{d}t. \quad (3.10)\end{aligned}$$

Remark 3.2.3. As for Corollary 3.1.3, an Itô formula for u depending also on time easily follows from Proposition 3.2.1.

Remark 3.2.4. In the literature, in particular in the mean field games context, some Itô formulas have been proved for the composition of $\mathcal{P}_2(\mathbb{R}^d)$-valued processes and real-valued functions over $\mathcal{P}_2(\mathbb{R}^d)$. A remarkable result can be found [32, Section 4.3], in which the $\mathcal{P}_2(\mathbb{R}^d)$-valued process is the law of a diffusion process of the form

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}s + \sigma(X_t)\,\mathrm{d}W_t + \bar{\sigma}(X_t)\,\mathrm{d}B_t,$$

conditioned to B, where B and W are two independent Brownian motions. The main difference with our technique is that we have an explicit equation for the measure-valued process and we use it to deduce the Itô formula, whilst in the approach of [32] the result is obtained combining the classical Itô formula, the empirical projection of the function u and the equation for the process X. In particular, a key tool in that approach are some formulas that relate the partial derivatives of the empirical projection $u(n^{-1}\sum_i^n \delta_{x_i})$ with the L-derivatives of u. We can also notice that, heuristically, if we set h equal to zero in the filtering problem, the Kushner-Stratonovich equation describe the law of X given the filtration generated by B up to a certain time. In this case, we can see that (3.9) coincides with the formula in [32] (see also Remark 3.2.5). Moreover, our technique also allows to deal with $\mathcal{M}_2^+(\mathbb{R}^d)$-valued processes, as we did in Section 3.1, thanks to the fact that it is based directly on the equation for the measure-valued process and not on the fact that the measure-valued process has to be a conditional law of a finite-dimensional process.

Remark 3.2.5. Our method can be used to obtain Itô formulas in other contexts, provided that a stochastic differential equation for a measure-valued process is available.

A first example is the filtering equations for a signal which is a pure-jump Markov process. In this case, the Zakai and the Kushner-Stratonovich equations have the same form as the ones considered above, with the only difference being that the signal's generator A is an integral operator. Under suitable assumptions on the signal, one can obtain Itô formulas that read as (3.6) and (3.10), where A is the integral generator mentioned above.

Another example, which is not related to stochastic filtering, is the SPDE associated to a McKean-Vlasov equation with a common noise B (see [64]). This SPDE is solved by a probability measure-valued process $\nu = \{\nu_t, t \in [0,T]\}$, which is the conditional law of the McKean-Vlasov process given the common noise. The equation reads as the Kushner-Stratonovich equation (3.8) where h has been taken equal to zero and with the coefficients $f, \sigma, \bar{\sigma}$ depending also on the measure. More precisely,

$$\mathrm{d}\langle \nu_t, \psi \rangle = \langle \nu_t, A(\nu_t)\psi \rangle \,\mathrm{d}t + \langle \nu_t, B(\nu_t)\psi \rangle \,\mathrm{d}B_t. \tag{3.11}$$

Thus, under proper assumptions (for instance $f, \sigma, \bar{\sigma}$ bounded and Lipschitz both on the space and the measure argument), one can obtain an Itô formula for (3.11). The formula will read as (3.10), with h equal to zero and with $A(\nu_t)$ instead of A.

Chapter 4

Classical solutions

In this chapter, we present most of the main results of this thesis, which are devoted to the introduction of the backward Kolmogorov equations associated to the filtering equations, and then to study the existence and uniqueness of their classical solutions. First, we focus on the Zakai equation, since its linearity makes some results easy to achieve. We investigate some regularity properties of its measure-valued solution. After this preliminary step, we write the backward Kolmogorov equation associated to the Zakai equation, which is a partial differential equation of parabolic type on a space of positive measures, and we prove the existence and uniqueness of classical solutions. In the last part of this chapter, we present analog results for the Kushner-Stratonovich equation, and in particular, we will study the existence and uniqueness of classical solution to a partial differential equation of parabolic type over a space of probability measures.

In the following sections we will ask repeatedly that Assuption 1 holds. We re-state them below.

Assumption 1. *All the mappings $f, \sigma, \bar{\sigma}, h$ are taken Borel-measurable. Moreover we assume:*

a. the mappings $f\colon \mathbb{R}^d \to \mathbb{R}^d$, $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ and $\bar{\sigma}\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are Lipschitz continuous;

b. the mapping $a := \sigma\sigma^\top\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is uniformly elliptic, that is there exists $\lambda > 0$ such that $\sum_{i,j=1}^d a_{ij}\xi_i\xi_j \geq \lambda|\xi|^2$ for every $x, \xi \in \mathbb{R}^d$;

c. the mappings $f, \sigma, \bar{\sigma}, h$ are bounded.

We also recall that the operators $A, B\colon \mathrm{C}^2_\mathrm{b}(\mathbb{R}^d) \to \mathrm{C}_\mathrm{b}(\mathbb{R}^d)$, introduced in Section 1.2, are defined for every $\psi \in \mathrm{C}^2_\mathrm{b}(\mathbb{R}^d)$, $x \in \mathbb{R}^d$, by

$$\begin{aligned} A\psi(x) &\coloneqq \sum_{i=1}^d f_i(x)\partial_i\psi(x) + \frac{1}{2}\sum_{i,j=1}^d \left\{ \left(\sigma\sigma^\top\right)_{ij}(x)\partial_{ij}\psi(x) + \left(\bar{\sigma}\bar{\sigma}^\top\right)_{ij}(x)\partial_{ij}\psi(x) \right\}, \\ B_k\psi(x) &\coloneqq \sum_{i=1}^d \bar{\sigma}_{ik}(x)\partial_i\psi(x), \quad k = 1, \dots d. \end{aligned} \tag{4.1}$$

4.1 Regularity with respect to the initial condition for the Zakai equation

In this section we investigate the differentiability of a solution to the Zakai equation with respect to the initial condition. We will consider a complete filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{Q})$ endowed with a $\{\mathcal{F}_t\}$-Brownian motion Y, and a solution to the Zakai equation $\rho^{s,\mu}$ on that probabilistic setup (see also Remark 1.3.10). We use the superscript s, μ to highlight the initial value $\mu \in \mathcal{M}_2^+(\mathbb{R}^d)$ and the initial time $s \in [0, T]$. In particular

$$\langle \rho_t^{s,\mu}, \psi\rangle = \langle \mu, \psi\rangle + \int_s^t \langle \rho_\tau^{s,\mu}, A\psi\rangle \, \mathrm{d}\tau + \int_s^t \langle \rho_\tau^{s,\mu}, h\psi + B\psi\rangle \cdot \mathrm{d}Y_\tau, \quad \psi \in \mathrm{C}_\mathrm{b}^2(\mathbb{R}^d), \tag{4.2}$$

where the coefficients are related to the filtering problem (1.1)-(1.2) and the operators A and B are defined by (4.1). We will also assume that Assumption 1 holds, so from Theorem 1.3.9 we have the pathwise uniqueness property. Finally, we will consider only initial conditions μ in $\mathcal{M}_2^+(\mathbb{R}^d)$, so ρ is a $\mathcal{M}_2^+(\mathbb{R}^d)$-valued process.

By computing formally the linear functional derivative of the equation (4.2), for every $x \in \mathbb{R}^d$ we get the following equation, defined over $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{Q}, Y)$, for an $\mathcal{M}(\mathbb{R}^d)$-valued process $Z^s(x) = \{Z_t^s(x), t \in [s, T]\}$:

$$\langle Z_t^s(x), \psi\rangle = \langle \delta_x, \psi\rangle + \int_s^t \langle Z_\tau^s(x), A\psi\rangle \, \mathrm{d}\tau + \int_s^t \langle Z_\tau^s(x), B\psi + h\psi\rangle \cdot \mathrm{d}Y_\tau. \tag{4.3}$$

We can look for solutions to (4.3) that are in $\mathcal{M}_2^+(\mathbb{R}^d)$ for every fixed $x \in \mathbb{R}^d$ and since it is a Zakai equation with initial condition in $\mathcal{P}_2(\mathbb{R}^d)$ and with the same coefficients of (4.2), we have that there exists a unique $\mathcal{M}_2^+(\mathbb{R}^d)$-valued solutions $Z_t(x)$ for every x.

Remark 4.1.1. The solution of (4.3) will play the role of linear functional derivative of the mapping $\mathcal{M}_2^+(\mathbb{R}^d) \ni \mu \mapsto \rho_t^{s,\mu}$, $t \in [s, T]$, see also Remark 4.1.5. We can notice that $Z^s(x)$ does not depend on μ, as expected since the Zakai equation is linear.

Before presenting the main results regarding the properties of the process $Z^s(x)$ introduced above, we provide an explicit estimate for the mass of a solution of the Zakai equation. We report the proof for completeness, even if the result is well known (see for instance Fact 3.2 in [80]).

Lemma 4.1.2. *Let $\rho^{s,\mu}$ be a solution of the Zakai equation* (4.2) *and let Assumption 1 be satisfied. Then it holds*

$$\mathbb{E}\left[|\langle \rho_t^{s,\mu}, \mathbf{1}\rangle|^2\right] \le 2\langle \mu, \mathbf{1}\rangle^2 e^{2T\|h\|_\infty^2},$$

where $\mathbf{1}$ is the function equal to 1 for every $x \in \mathbb{R}^d$ and $\mu \in \mathcal{M}_2^+(\mathbb{R}^d)$.

Proof. Since ρ^μ solves (4.2), we can write the Zakai equation for $\psi \equiv \mathbf{1}$ and then take the expected value. Thus, for every $t \in [0, T]$, we get

$$\mathbb{E}\left[|\langle \rho_t^{s,\mu}, \mathbf{1}\rangle|^2\right] \leq 2\langle \mu, \mathbf{1}\rangle^2 + 2\|h\|_\infty^2 \int_s^t \mathbb{E}\left[|\langle \rho_\tau^{s,\mu}, \mathbf{1}\rangle|^2\right] \mathrm{d}\tau,$$

and the thesis follows by Gronwall's lemma. □

Proposition 4.1.3. *Let $s \in [0, T)$. Let $\rho^{s,\mu}$ be the solution of* (4.2) *starting at $\mu \in \mathcal{M}_2^+(\mathbb{R}^d)$ and let $Z^s(x)$, $x \in \mathbb{R}^d$, be the solution of* (4.3). *Then*

i. for every $m, m' \in \mathcal{M}_2^+(\mathbb{R}^d)$ it holds

$$\langle \rho_t^{s,m'}, \psi\rangle - \langle \rho_t^{s,m}, \psi\rangle = \int_0^1 \int_{\mathbb{R}^d} \langle Z_t^s(x), \psi\rangle\, [m' - m]\,(\,\mathrm{d}x)\,\mathrm{d}\theta, \quad \forall t \in [s, T],$$

almost surely;

ii. for every $t \in [s, T]$ and $x \in \mathbb{R}^d$, there exists a constant $C = C(T, h) > 0$ such that

$$\mathbb{E}\left[|\langle Z_t^s(x), \mathbf{1}\rangle|^2\right] \leq C(T, h).$$

Proof. In this proof we hide the dependence on s in $\rho^{s,\mu}$ and $Z^s(x)$. Let us define, for every $t \in [s, T]$ and every $m, m' \in \mathcal{M}_2^+(\mathbb{R}^d)$, the mapping

$$\tilde{Z}_t^{m,m'} \colon \mathrm{C}_\mathrm{b}^2(\mathbb{R}^d) \ni \psi \mapsto \langle \tilde{Z}_t^{m,m'}, \psi\rangle := \int_{\mathbb{R}^d} \langle Z_t(x), \psi\rangle [m' - m](\,\mathrm{d}x).$$

It is easy to check that $\tilde{Z}_t^{m,m'} \in \mathcal{M}(\mathbb{R}^d)$ and then we can define the measure-valued process $\Delta^{m,m'} = \{\Delta_t^{m,m'} := \rho_t^{m'} - \rho_t^m - \tilde{Z}_t^{m,m'}, t \in [s, T]\}$. Recalling that $\rho^{m'}$ and ρ^m solve (4.2) and $Z(x)$ solves (4.3), by linearity we obtain that, for every $m, m' \in \mathcal{M}_2^+(\mathbb{R}^d)$, $\Delta^{m,m'}$ solves a Zakai equation with null initial condition and same coefficients as (4.2). Then, it holds that $\Delta^{m,m'}$ is the process equal to the null measure for every $t \in [s, T]$. Thus, since $Z_t(x)$ does not depend on μ, we can say that for every $t \in [s, T]$

$$\langle \rho_t^{m'}, \psi\rangle - \langle \rho_t^m, \psi\rangle = \int_0^1 \int_{\mathbb{R}^d} \langle Z_t(x), \psi\rangle\, [m' - m]\,(\,\mathrm{d}x)\,\mathrm{d}\theta,$$

almost surely. Regarding *ii*, it follows directly from Lemma 4.1.2 □

We also need to study the differentiability of the mapping $\mathbb{R}^d \ni x \mapsto \mathbb{E}\left[\langle Z_t^s(x), \psi\rangle\right] \in \mathbb{R}$, for every fixed $\psi \in \mathrm{C}_\mathrm{b}^2(\mathbb{R}^d)$ and $t \in [s, T]$.

Proposition 4.1.4. *Let $Z^s(x)$ be the solution of the equation* (4.3) *and let Assumption 1 holds. Then, for every $\psi \in \mathrm{C}_\mathrm{b}^2(\mathbb{R}^d)$ and $t \in [s, T]$, the mapping $\mathbb{R}^d \ni x \mapsto \mathbb{E}\left[\langle Z_t^s(x), \psi\rangle\right] \in \mathbb{R}$ is twice continuously differentiable, with bounded derivatives.*

Before proving Proposition 4.1.4, let us introduce some auxiliary tools. Let us denote by $I^s(x)$ the intensity measure associated to $Z^s(x)$, that is the measure such that $\mathbb{E}\left[\langle Z_t^s(x), \psi\rangle\right] = \langle I_t^s(x), \psi\rangle$. From (4.3) we have that

$$\mathbb{E}\left[\langle Z_t^s(x), \psi\rangle\right] = \langle \delta_x, \psi\rangle + \int_s^t \mathbb{E}\left[\langle Z_\tau^s(x), A\psi\rangle\right] \mathrm{d}\tau, \quad t \in [s, T], \quad \psi \in \mathrm{C}_\mathrm{b}^2(\mathbb{R}^\mathrm{d}), \quad (4.4)$$

and so $I^s(x)$ solves the following Fokker-Planck equation:

$$\langle I_t^s(x), \psi\rangle = \langle \delta_x, \psi\rangle + \int_s^t \langle I_\tau^s(x), A\psi\rangle \,\mathrm{d}\tau, \quad t \in [s,T], \quad \psi \in \mathrm{C}_\mathrm{b}^2(\mathbb{R}^\mathrm{d}). \tag{4.5}$$

Following the argument used for instance to prove Proposition 6.1.2 in [19] or Lemma 4.8 in [8], from (4.5) one can deduce that for every $\varphi \in \mathrm{C}_\mathrm{b}^{1,2}(\mathbb{R}^d \times [s,t])$ it holds that

$$\langle I_t^s(x), \varphi_t\rangle = \langle \delta_x, \varphi_s\rangle + \int_s^t \langle I_\tau^s(x), (\partial_\tau + A)\varphi_\tau\rangle \,\mathrm{d}\tau. \tag{4.6}$$

Indeed, for $i = 0, 1, \ldots, n-1$ we have (we set $s = 0$ for simplicity)

$$\langle I_{(i+1)t/n}(x), \varphi_{it/n}\rangle = \langle I_{it/n}(x), \varphi_{it/n}\rangle + \int_{it/n}^{(i+1)t/n} \langle I_\tau(x), A\varphi_{it/n}\rangle \,\mathrm{d}\tau,$$

and from Fubini's theorem it holds

$$\langle I_{(i+1)t/n}(x), \varphi_{(i+1)t/n} - \varphi_{it/n}\rangle = \int_{it/n}^{(i+1)t/n} \langle I_{(i+1)t/n}(x), \partial_\tau\varphi_\tau\rangle \,\mathrm{d}\tau.$$

Then, it follows that

$$\begin{aligned}\langle I_{(i+1)t/n}(x), \varphi_{(i+1)t/n}\rangle &= \langle I_{(i+1)t/n}(x), \varphi_{(i+1)t/n} - \varphi_{it/n}\rangle + \langle I_{(i+1)t/n}(x), \varphi_{it/n}\rangle \\ &= \langle I_{it/n}(x), \varphi_{it/n}\rangle + \int_{it/n}^{(i+1)t/n} \langle I_{(i+1)t/n}(x), \partial_\tau\varphi_\tau\rangle \,\mathrm{d}\tau + \int_{it/n}^{(i+1)t/n} \langle I_\tau(x), A\varphi_{it/n}\rangle \,\mathrm{d}\tau.\end{aligned}$$

Summing over the intervals $[it/n, (i+1)t/n]$ from $i = 0$ to $n-1$ we obtain

$$\langle I_t(x), \varphi_t\rangle \;=\; \langle \delta_x, \varphi_0\rangle + \int_0^t \langle I_{([ns/t]+1)t/n}(x), \partial_\tau\varphi_\tau\rangle \,\mathrm{d}\tau + \int_0^t \langle I_\tau(x), A\varphi_{[ns/t]t/n}\rangle \,\mathrm{d}\tau,$$

and (4.6) follows by taking the limit as n tends to infinity and using dominated convergence theorem.
Starting from (4.6), we can prove Proposition 4.1.4. The argument exploits the regularity of the solution of a suitable auxiliary backward partial differential equation.

Proof of Proposition 4.1.4. Let us fix $t \in [s,T]$, $\psi \in \mathrm{C}_\mathrm{b}^2(\mathbb{R}^d)$ and let us introduce the backward equation

$$\begin{cases}\partial_\tau v(y,\tau) + Av(y,\tau) = 0, & (y,\tau) \in \mathbb{R}^d \times [s,t], \\ v(y,t) = \psi(y), & y \in \mathbb{R}^d.\end{cases} \tag{4.7}$$

Thanks to Assumption 1 (see for instance Theorem 4.6 in [60] and more precisely the discussion after Theorem 5.1, page 147), we have that there exists a unique classical solution $v \in \mathrm{C}_\mathrm{b}^{2,1}(\mathbb{R}^d \times [s,t])$ to (4.7). Then, if we choose v as a test function in (4.6), we obtain

$$\langle I_t^s(x), v_t\rangle = \langle \delta_x, v_s\rangle + \int_s^t \langle I_\tau^s(x), (\partial_\tau + A)v_\tau\rangle \,\mathrm{d}\tau = v(x,s), \tag{4.8}$$

where the last equality follows from the fact that v solves (4.7). Thus, recalling that $v(x,t) = \psi(x)$ and the definition of $I^s(x)$, we obtain that

$$\mathbb{E}\left[\langle Z_t^s(x), \psi\rangle\right] = \langle I_t^s(x), \psi\rangle = v(x,s), \quad x \in \mathbb{R}^d,$$

and so the mapping $\mathbb{R}^d \ni x \mapsto \mathbb{E}\left[\langle Z_t^s(x), \psi\rangle\right]$ is in $\mathrm{C}^2_{\mathrm{b}}(\mathbb{R}^d)$ thanks to the regularity of v. □

Remark 4.1.5. In Proposition 4.1.3, we showed that for $t \in [s,T]$ and $\psi \in \mathrm{C}^2_{\mathrm{b}}(\mathbb{R}^d)$ fixed, the mapping $\mu \mapsto \langle \rho_t^{s,\mu}, \psi\rangle \in L^1(\Omega)$ is in $\mathrm{C}^1(\mathcal{M}_2^2(\mathbb{R}^d); L^1(\Omega))$, with derivative given by $\langle Z_t^s(x), \psi\rangle$ and independent of μ. Note that this does not imply that, almost surely, Z_t^s is the linear functional derivative of $\mu \mapsto \rho_t^{s,\mu}$, since the continuity of $x \mapsto \langle Z_t^s(x), \psi\rangle$ holds only under expectation.

With the same procedure used in Proposition 4.1.3 we can find a process, that we denote with $U^s(x,y)$, $x, y \in \mathbb{R}^d$, which is symmetrical with respect to x and y and which satisfies properties analogue to *i, ii,* in Proposition 4.1.3 where $\rho^{s,\mu}$ is substituted with $Z^s(x)$ and $Z^s(x)$ with $U^s(x,y)$. Moreover, it turns out that $U^s(x,y)$ coincides with the null measure for every $t \in [s,T]$ and $x, y \in \mathbb{R}^d$.

To conclude, we summarize in a proposition all the properties we showed in this section and which will be useful in the following discussions.

Proposition 4.1.6. *Let $s \in [0,T)$, let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{Q}, Y)$ be fixed and let Assumption 1 holds. If $\rho^{s,\mu}$ is the solution of the Zakai equation (4.2), then, for every $\psi \in \mathrm{C}^2_{\mathrm{b}}(\mathbb{R}^d)$ and $t \in [s,T]$, the mapping $\mu \mapsto \langle \rho_t^{s,\mu}, \psi\rangle$ is in $\mathrm{C}^2(\mathcal{M}_2^+(\mathbb{R}^d); L^1(\Omega))$. Moreover, the mapping $\mu \mapsto \mathbb{E}\left[\langle \rho_t^{s,\mu}, \psi\rangle\right]$ is in $\mathrm{C}^2_{\mathrm{L}}(\mathcal{M}_2^+(\mathbb{R}^d))$.*

4.2 The backward Kolmogorov equation associated to the Zakai equation

In this section we write and study the backward Kolmogorov equation associated to the Zakai equation, that is a parabolic partial differential equation on a space of positive measures. Let us denote with $\mathcal{L}$ the infinitesimal generator of the Zakai process, namely the operator $\mathcal{L}\colon \mathrm{C}^2_{\mathrm{L}}(\mathcal{M}_2^+(\mathbb{R}^d)) \to \mathrm{C}_{\mathrm{b}}(\mathcal{M}_2^+(\mathbb{R}^d))$ defined by

$$\mathcal{L}u(\mu) = \langle \mu, A\delta_\mu u(\mu)\rangle + \frac{1}{2}\langle \mu \otimes \mu, (h+B)\cdot(h+B)\delta_\mu^2 u(\mu)\rangle,$$

where A and B are defined by (4.1) and $h\colon \mathbb{R}^d \to \mathbb{R}^d$ is Borel measurable and bounded. The backward Kolmogorov equation we want to study is:

$$\begin{cases} \partial_s u(\mu,s) + \mathcal{L}u(\mu,s) = 0 & (\mu,s) \in \mathcal{M}_2^+(\mathbb{R}^d) \times [0,T], \\ u(\mu,T) = \Phi(\mu) & \mu \in \mathcal{M}_2^+(\mathbb{R}^d), \end{cases} \tag{4.9}$$

where Φ is in $\mathrm{C}^2_{\mathrm{L}}(\mathcal{M}_2^+(\mathbb{R}^d))$. Our aim is to study existence and uniqueness of classical solutions, in the sense given by the following definition:

Definition 4.2.1. *We say that* $u\colon \mathcal{M}_2^+(\mathbb{R}^d) \times [0,T] \to \mathbb{R}$ *is a classical solution to the backward Kolmogorov equation associated to the Zakai equation if it is of class* $\mathrm{C}^2_{\mathrm{L}}(\mathcal{M}_2^+(\mathbb{R}^d))$ *in the measure argument and* $\mathrm{C}^1([0,T])$ *in the time argument (where in* $t = 0$ *and* $t = T$ *the derivatives are understood in unilateral sense), if it is bounded and all its derivatives are bounded in all their arguments and if it satisfies the backward equation* (4.9)*.*

4.2.1 Existence and uniqueness of a classical solution

In order to show existence and uniqueness, we follow the classical approach to these kind of problems. First, we assume that a solution exists and we prove a representation formula which guarantees the uniqueness. In the following, when we refer to a solution $\rho^{s,\mu}$ to the Zakai equation, it is understood as the solution defined over a fixed complete filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{Q})$ endowed with a $\{\mathcal{F}_t\}$-Brownian motion Y (see Remark 1.3.10), which solves the equation starting from $\mu \in \mathcal{M}_2^+(\mathbb{R}^d)$ at $s \in [0,T)$.

Proposition 4.2.2. *Let* $u = u(\mu, s)$ *be a classical solution to* (4.9) *and let* $\rho^{s,\mu}$ *be a solution to the Zakai equation* (4.2)*. Then the following representation formula holds*

$$u(\mu, s) = \mathbb{E}^{\mathbb{Q}}\left[\Phi\left(\rho_T^{s,\mu}\right)\right], \quad (\mu, s) \in \mathcal{M}_2^+(\mathbb{R}^d) \times [0,T], \tag{4.10}$$

and so the solution u *is uniquely characterized.*

Proof. Let us consider the composition $u(\rho_T^{s,\mu}, T)$, where $\rho_T^{s,\mu}$ is the solution to (4.2) starting at time s with value μ. Then, by the Itô formula (3.7) we get

$$u(\rho_T^{s,\mu}, T) - u(\rho_s^{s,\mu}, s) = \int_s^T \{\partial_s u(\rho_\tau^{s,\mu}, \tau) + \mathcal{L}u(\rho_\tau^{s,\mu}, \tau)\}\,\mathrm{d}\tau + \int_s^T \mathcal{G}u(\rho_\tau^{s,\mu}, \tau) \cdot \mathrm{d}Y_\tau,$$

where $\mathcal{G}u(\mu, t) = \langle \mu, h\delta_\mu u(\mu, t) + \bar{\sigma}^\top \mathrm{D}_\mu u(\mu, t)\rangle$. First, we notice that since u is a solution of (4.9), the time integral is zero, and by taking the expectation we get

$$\mathbb{E}^{\mathbb{Q}}\left[\Phi\left(\rho_T^{s,\mu}\right)\right] - u(\mu, s) = \mathbb{E}^{\mathbb{Q}}\left[\int_s^T \mathcal{G}u(\rho_\tau^{s,\mu}, \tau) \cdot \mathrm{d}Y_\tau\right].$$

The right hand side is equal to zero since the stochastic integral in the expected value is a martingale. Then the thesis follows, since for every $(\mu, s) \in \mathcal{M}_2^+(\mathbb{R}^d) \times [0,T]$,

$$\mathbb{E}^{\mathbb{Q}}\left[\Phi\left(\rho_T^{s,\mu}\right)\right] = u(\mu, s).$$

□

In order to prove the existence of a solution, we will show that a function u defined by (4.10) is regular enough and satisfies (4.9). To this aim, we need some auxiliary results on the differentiability of u with respect to the measure argument that we collect in the following proposition:

Proposition 4.2.3. *Let* $u(\mu, s) = \mathbb{E}^{\mathbb{Q}}[\Phi(\rho_T^{s,\mu})]$*, where* $\rho_T^{s,\mu}$ *is the solution to the Zakai equation starting at time* s *from* $\mu \in \mathcal{M}_2^+(\mathbb{R}^d)$ *and* $\Phi \in \mathrm{C}^2_{\mathrm{L}}(\mathcal{M}_2^+(\mathbb{R}^d))$*. Then, for every* $s \in [0,T]$*,* u *is in* $\mathrm{C}^2_{\mathrm{L}}(\mathcal{M}_2^+(\mathbb{R}^d))$*.*

Proof. Let us first deal with the differentiability in linear functional sense. Proceeding as in Proposition 2.1.22, by a simple chain rule argument we conclude that $u(\mu, s)$ is in $\mathrm{C}^2(\mathcal{M}_2^+(\mathbb{R}^\mathrm{d}))$, thanks to the fact $\Phi \in \mathrm{C}^2_\mathrm{L}(\mathcal{M}_2^+(\mathbb{R}^d))$ combined with Proposition 4.1.6. In particular we get the following formulas:

$$\begin{aligned}\delta_\mu u(\mu, x, s) &= \mathbb{E}^\mathbb{Q}\left[\langle Z^s_T(x), \delta_\mu \Phi(\rho^{s,\mu}_T)\rangle\right],\\ \delta^2_\mu u(\mu, x, y, s) &= \mathbb{E}^\mathbb{Q}\left[\langle Z^s_T(x) \otimes Z^s_t(y), \delta^2_\mu \Phi(\rho^{s,\mu}_T)\rangle\right],\end{aligned}$$

where $\delta_\mu\Phi(\rho^{s,\mu}_T, x)$ means $\delta_\mu\Phi(\mu, x)$ evaluated in $\rho^{s,\mu}_T$, and Z^s is the process introduced in (4.3).

Regarding the differentiability of the first-order linear functional derivative with respect the additional space variable, we have that the mapping $x \mapsto \delta_\mu u(\mu, x, s)$ is twice continuously differentiable with bounded derivatives thanks to Proposition 4.1.6 and to the fact that $\delta_\mu\Phi(\rho^{s,\mu}_T) \in \mathrm{C}^2_\mathrm{b}(\mathbb{R}^d)$ for μ fixed. In a similar way we can show that the mapping $(x, y) \mapsto \delta^2_\mu u(\mu, x, y)$ is twice continuously differentiable with bounded derivatives. Indeed, if we take a symmetrical function $\psi \in \mathrm{C}^2_\mathrm{b}(\mathbb{R}^d \times \mathbb{R}^d)$, we can obtain a version of Proposition 4.1.6 for the mapping $(x, y) \mapsto \mathbb{E}^\mathbb{Q}\left[\langle Z^s_T(x) \otimes Z^s_t(y), \psi\rangle\right]$, by combining the technique used in the proof of Proposition 4.1.6 and the ideas in the proof of Theorem 4.26 in [8]. □

Now that the object $\mathcal{L}u$ is well defined, we need to investigate its regularity with respect to the time.

Lemma 4.2.4. *Let u be defined by (4.10). Then, for every $\mu \in \mathcal{M}_2^+(\mathbb{R}^d)$, the mappings $[0, T] \ni s \mapsto \mathcal{L}u(\mu, s)$ and $[s, T]\times[0, T] \ni (\tau, \sigma) \mapsto \mathcal{L}u(\rho^{s,\mu}_\tau, \sigma) \in L^2(\Omega)$ are continuous.*

Proof. Let us fix $t \in [s, T]$. Then, by classical estimates on (4.2), it follows that for every $\psi \in \mathrm{C}^2_\mathrm{b}(\mathbb{R}^d)$ the mapping $[0, T] \ni s \mapsto \langle\rho^{s,\mu}_t, \psi\rangle \in L^2(\Omega)$ is continuous. Thus, combining this with the expression for the derivatives of u in Proposition 4.2.3 and the boundedness of Φ with its derivatives, we get that $[0, T] \ni s \mapsto \mathcal{L}u(\mu, s)$ is continuous. Regarding $[s, T] \times [0, T] \ni (\tau, \sigma) \mapsto \mathcal{L}u(\rho^{s,\mu}_\tau, \sigma)$, again we can conclude recalling that $\Phi \in \mathrm{C}^2_\mathrm{L}(\mathcal{M}_2^+(\mathbb{R}^d))$ combined with Proposition 4.2.3 and the continuity of $\rho^{s,\mu}$. □

Finally we can show the main result, that is the existence of a solution to (4.9) via representation formula:

Theorem 4.2.5. *Let $u(\mu, s) = \mathbb{E}^\mathbb{Q}\left[\Phi(\rho^{s,\mu}_T)\right]$, where $\rho^{s,\mu}_T$ is the solution to the Zakai equation starting at time s from $\mu \in \mathcal{M}_2^+(\mathbb{R}^d)$, $\Phi \in \mathrm{C}^2_\mathrm{L}(\mathcal{M}_2^+(\mathbb{R}^d))$ and let Assumption 1 holds. Then it is the unique classical solution to the backward Kolmogorov equation (4.9).*

Proof. Let us fix h small and positive. We want to show that

$$\lim_{h\to 0} \frac{1}{h}\left[u(\mu, s + h) - u(\mu, s)\right] = -\mathcal{L}u(\mu, s). \tag{4.11}$$

If this is true, the mapping $g\colon s \mapsto g(s) := u(\mu, s)$ has right derivative in $[0, T)$. Moreover, by Lemma 4.2.4 the right-hand term in (4.11) is continuous, so $g \in \mathrm{C}^1([0, T))$ and by a standard argument it can be shown that it is continuously differentiable in $[0, T]$.

Let us show (4.11). First, thanks to the Markov property of the process $\rho^{s,\mu}$ it holds that $u(\mu, s) = \mathbb{E}^{\mathbb{Q}}\left[u\left(\rho^{s,\mu}_{s+h}, s+h\right)\right]$. Then, we can proceed by applying Itô formula and taking the expectation:

$$\begin{aligned} &u(\mu, s+h) - u(\mu, s) \\ &\qquad = \mathbb{E}^{\mathbb{Q}}\left[u(\mu, s+h) - u\left(\rho^{s,\mu}_{s+h}, s+h\right)\right] = -\,\mathbb{E}^{\mathbb{Q}}\left[\int_s^{s+h} \mathcal{L}u(\rho^{s,\mu}_\tau, s+h)\,\mathrm{d}\tau\right]. \end{aligned}$$

To conclude, it remains to show that

$$\lim_{h\to 0}\frac{1}{h}\,\mathbb{E}^{\mathbb{Q}}\left[\int_s^{s+h} \mathcal{L}u(\rho^{s,\mu}_\tau, s+h)\,\mathrm{d}\tau\right] = \mathcal{L}u(\mu, s),$$

but this follows from Lemma 4.2.4 and mean-value theorem. □

Remark 4.2.6. All the previous results can be extended to the time inhomogeneous case, that is when the coefficients $b, \sigma, \bar\sigma$ depend also on time, by assuming that Assumption 1 holds with uniform in time constants.

4.3 The backward Kolmogorov equation associated to the Kushner-Stratonovich equation

Our goal in this last section is to prove existence and uniqueness for the backward Kolmogorov equation associated to the Kushner-Stratonovich equation. We will proceed by exploiting the relation with the Zakai equation, pointed out at the end of Section 1.3. Let us fix $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{Q}, Y)$ and let $\rho^{s,\pi}$ be a solution to (4.2) starting at $\pi \in \mathcal{P}_2(\mathbb{R}^d)$. Let us define the couple

$$I^\pi_t = Y_t - \int_s^t \frac{\langle \rho^{s,\pi}_\tau, h\rangle}{\rho^{s,\pi}_\tau(\mathbb{R}^d)}\,\mathrm{d}\tau, \quad \xi^\pi_t = \exp\left\{\int_s^t \frac{\langle \rho^{s,\pi}_\tau, h\rangle}{\rho^{s,\pi}_\tau(\mathbb{R}^d)}\,\mathrm{d}Y_\tau - \frac{1}{2}\int_s^t \left|\frac{\langle \rho^{s,\pi}_\tau, h\rangle}{\rho^{s,\pi}_\tau(\mathbb{R}^d)}\right|^2 \mathrm{d}\tau\right\}, \tag{4.12}$$

and set

$$\mathrm{d}\mathbb{P}^\pi = \xi^\pi_T\,\mathrm{d}\mathbb{Q}, \quad \Pi^{s,\pi} = \rho^{s,\pi}/\rho^{s,\pi}(\mathbb{R}^d). \tag{4.13}$$

As remarked in Section 1.3, under Assumption 1 $\{(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P}^\pi), (\Pi^{s,\pi}, I^\pi)\}$ is the unique in law weak solution to the Kushner-Stratonovich equation starting at $\pi \in \mathcal{P}_2(\mathbb{R}^d)$. In particular, for every $t \in [s, T]$ and $\psi \in \mathrm{C}^2_{\mathrm{b}}(\mathbb{R}^d)$ it holds

$$\langle \Pi^{s,\pi}_t, \psi\rangle = \langle \pi, \psi\rangle + \int_s^t \langle \Pi^{s,\pi}_\tau, A\psi\rangle\,\mathrm{d}\tau + \int_s^t \left(\langle \Pi^{s,\pi}_\tau, h\psi + B\psi\rangle - \langle \Pi^{s,\pi}_\tau, \psi\rangle\langle \Pi^{s,\pi}_\tau, h\rangle\right)\cdot \mathrm{d}I^{s,\pi}_\tau. \tag{4.14}$$

4.3.1 The backward Kolmogorov equation

As we did for the Zakai equation, we want to discuss the existence and uniqueness of classical solutions to the backward Kolmogorov equation associated to the Kushner-Stratonovich equation (5.2). Such partial differential equation reads as

$$\begin{cases} \partial_s u(\pi, s) + \mathcal{L}^{KS} u(\pi, s) = 0 & (\pi, s) \in \mathcal{P}_2(\mathbb{R}^d) \times [0, T], \\ u(\pi, T) = \Phi(\pi) & \pi \in \mathcal{P}_2(\mathbb{R}^d), \end{cases} \tag{4.15}$$

where $\Phi \in \mathrm{C}^2_{\mathrm{L}}(\mathcal{P}_2(\mathbb{R}^d))$ and the operator $\mathcal{L}^{KS}\colon \mathrm{C}^2_{\mathrm{L}}(\mathcal{P}_2(\mathbb{R}^d)) \to \mathrm{C}_{\mathrm{b}}(\mathcal{P}_2(\mathbb{R}^d))$ is defined by

$$\mathcal{L}^{KS}u(\pi) = \langle \pi, A\delta_\mu u(\pi)\rangle + \frac{1}{2}\langle \pi\otimes\pi, (h + B - \langle \pi, h\rangle)\cdot(h + B - \langle \pi, h\rangle)\delta^2_\mu u(\pi)\rangle,$$

where A and B are defined by (4.1) and h is Borel measurable and bounded.

Definition 4.3.1. *We say that $u\colon \mathcal{P}_2(\mathbb{R}^d)\times[0,T] \to \mathbb{R}$ is a classical solution to* (4.15) *if it of class $\mathrm{C}^2_{\mathrm{L}}(\mathcal{P}_2(\mathrm{R}^{\mathrm{d}}))$ in the measure argument and $\mathrm{C}^1([0,T])$ (where in $t = 0$ and $t = T$ the derivatives are understood in unilateral sense) in the time argument, if it is bounded and all its derivatives are bounded in all their arguments, and if it satisfies the backward equation* (4.15)*.*

As we did for the Kolmogorov equation associated to the Zakai equation, we want to show existence and uniqueness via a representation formula. Let $\rho^{s,\pi}$ be a solution to the Zakai equation defined over $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{Q}, Y)$. Following (4.12)-(4.13), the couple $\{(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P}^\pi), (\Pi^{s,\pi}, I^\pi)\}$ solves weakly the Kushner-Stratonovich equation. We can notice that the probability space $(\Omega, \mathcal{F}, \mathbb{Q})$ is fixed for every $\pi \in \mathcal{P}_2(\mathbb{R}^d)$, but since I^π and ξ^π depend on $\rho^{s,\pi}$, the probability space $(\Omega, \mathcal{F}, \mathbb{P}^\pi)$ depends on the initial point $\pi \in \mathcal{P}_2(\mathbb{R}^d)$. Our claim is that

$$u(\pi, s) \coloneqq \mathbb{E}^{\mathbb{P}^\pi}\left[\Phi(\Pi^{s,\pi}_T)\right] = \mathbb{E}^{\mathbb{Q}}\left[\Phi(\rho^{s,\pi}_T/\rho^{s,\pi}_T(\mathbb{R}^d))\xi^\pi_T\right]$$

is the unique weak solution to (4.15). In order to study its regularity, we rely on the relations (4.13) and the regularity results obtained in Section 4.1 for the Zakai context.

Proposition 4.3.2. *Let $u(\pi, s) \coloneqq \mathbb{E}^{\mathbb{P}^\pi}[\Phi(\Pi^{s,\pi}_T)] = \mathbb{E}^{\mathbb{Q}}[\Phi(\Pi^{s,\pi}_T)\xi^\pi_T]$ be defined as above and let Assumption 1 holds. Then for every $s \in [0,T]$ the mapping $u(\cdot, s) \in \mathrm{C}^2_{\mathrm{L}}(\mathcal{P}_2(\mathbb{R}^d))$. Moreover, the mappings $[0,T] \ni s \mapsto \mathcal{L}^{KS}u(\pi, s)$ and $[s,T]\times[0,T] \ni (\tau, \sigma) \mapsto \mathcal{L}^{KS}u(\Pi^{s,\pi}_\tau, \sigma) \in L^2(\Omega, \mathbb{P}^\pi)$ are continuous.*

Proof. First, since $\mathrm{d}\rho^{s,\pi}(\mathbb{R}^d) = \rho^{s,\pi}(\mathbb{R}^d)\langle \Pi^{s,\pi}_t, h\rangle \cdot \mathrm{d}Y_t$, the process $1/\rho^{s,\pi}(\mathbb{R}^d)$ has uniformly in time bounded $\mathbb{Q}$ moments of any order $p \in [1, +\infty)$. Then, let us compute the linear functional derivative of u for a fixed $s \in [0, T)$:

$$\begin{aligned}
\delta_\pi u(\pi, x, s) &= \mathbb{E}^{\mathbb{Q}}\left[\xi^\pi_T \delta_\pi\left(\Phi\left(\frac{\rho^{s,\cdot}_T}{\rho^{s,\cdot}_T(\mathbb{R}^d)}\right)\right)(\pi, x) + \Phi\left(\frac{\rho^{s,\cdot}_T}{\rho^{s,\cdot}_T(\mathbb{R}^d)}\right)\delta_\pi \xi^\cdot_T(\pi, x)\right] \\
&= \mathbb{E}^{\mathbb{Q}}\left[\frac{\xi^\pi_T}{\rho^{s,\pi}_T(\mathbb{R}^d)}\langle Z^s_T(x), \delta_\mu\Phi\left(\frac{\rho^{s,\pi}_T}{\rho^{s,\pi}_T(\mathbb{R}^d)}\right)\rangle\right] \\
&\quad - \mathbb{E}^{\mathbb{Q}}\left[\frac{\xi^\pi_T}{\rho^{s,\pi}_T(\mathbb{R}^d)}\langle\frac{\rho^{s,\pi}_T}{\rho^{s,\pi}_T(\mathbb{R}^d)}, \delta_\mu\Phi\left(\frac{\rho^{s,\pi}_T}{\rho^{s,\pi}_T(\mathbb{R}^d)}\right)\rangle\langle Z^s_T(x), \mathbf{1}\rangle\right] \\
&\quad + \mathbb{E}^{\mathbb{Q}}\left[\Phi\left(\frac{\rho^{s,\pi}_T}{\rho^{s,\pi}_T(\mathbb{R}^d)}\right)\xi^\pi_T\int_s^T \frac{1}{\rho^{s,\pi}_\tau(\mathbb{R}^d)}\left(\langle Z^s_\tau(x), h\rangle - \frac{\langle \rho^{s,\pi}_\tau, h\rangle}{\rho^{s,\pi}_\tau(\mathbb{R}^d)}\langle Z^s_\tau(x), \mathbf{1}\rangle\right)\cdot \mathrm{d}Y_\tau\right] \\
&\quad - \mathbb{E}^{\mathbb{Q}}\left[\Phi\left(\frac{\rho^{s,\cdot}_T}{\rho^{s,\cdot}_T(\mathbb{R}^d)}\right)\xi^\pi_T\int_s^T \frac{\langle \rho^{s,\pi}_\tau, h\rangle}{\rho^{s,\pi}_\tau(\mathbb{R}^d)^2}\cdot\left(\langle Z^s_\tau(x), h\rangle - \frac{\langle \rho^{s,\pi}_\tau, h\rangle}{\rho^{s,\pi}_\tau(\mathbb{R}^d)}\langle Z^s_\tau(x), \mathbf{1}\rangle\right)\mathrm{d}\tau\right],
\end{aligned}$$

where we computed $\delta_\pi \xi_T^\cdot(\pi, x)$ thanks to stochastic and deterministic Fubini's theorem. Continuity and boundedness are guaranteed by the regularity and boundedness of the processes involved under the $\mathbb{Q}$-expectation. In the same way one can show the second-order differentiability in linear functional sense of u. Regarding the differentiability in space, again we can bring the derivative in space inside the expectation and exploit the regularity results in Proposition 4.1.6. To conclude, the continuity of the mappings $s \mapsto \mathcal{L}^{KS}u(\pi, s)$ and $(\tau, \sigma) \mapsto \mathcal{L}^{KS}u(\Pi_\tau^{s,\pi}, \sigma) \in L^2(\Omega, \mathbb{P}^\pi)$ follows as in Lemma 4.2.4. □

Finally, we can state the existence and uniqueness result for the Kolmogorov equation associated to the Kushner-Stratonovich equation (5.2):

Theorem 4.3.3. *Let Assumption 1 holds and let* $\{(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P}^\pi), (\Pi^{s,\pi}, I^\pi)\}$ *be the weak solution to the Kushner-Stratonovich equation obtained through* (4.12)-(4.13)*. There exists a unique classical solution to the backward Kolmogorov equation* (4.15) *starting at* $\Phi \in \mathrm{C}^2_{\mathrm{L}}(\mathbb{R}^d)$*, given by*

$$u(\pi, s) = \mathbb{E}^{\mathbb{P}^\pi}\left[\Phi(\Pi_T^{s,\pi})\right], \quad (\pi, s) \in \mathcal{P}_2(\mathbb{R}^d) \times [0, T].$$

Proof. The proof follows exactly the one we did in Section 4.2.1 for the backward Kolmogorov equation associated to the Zakai equation, exploiting the regularity results in Proposition 4.3.2 and the Markov property for the existence, and the Itô formula in Proposition 3.2.1 for the uniqueness. □

Chapter 5

Viscosity solutions in the Kushner-Stratonovich case

In this chapter, we continue the analysis of the backward Kolmorogorov equations introduced in Chapter 4, by trying to reduce the regularity assumptions given on the final condition. Despite the small effort put into the formulation of this new problem, it turns out that the approach must be completely different from the one adopted before. First, we have to change the notion of solution. Indeed, if we consider the candidate solution given by the usual representation formula, there is no hope to obtain the regularity required by the notion of classical solution from a non-smooth final condition. Thus, we decide to look for viscosity solutions, since they are widely and successfully used to deal with the nonlinear version of Kolmogorov equations, namely the Hamilton-Jacobi-Bellman equations. However, even if this notion of solution allows us to obtain some results, we are forced to add some assumptions. In particular, we will focus on the Kolmogorov equation associated to the Kushner-Stratonovich equation, and we will consider the partial differential equation over a space of probability measures with compact support. We point out that, differently from the previous chapter, the techniques we used here can not be immediately extended to the case of positive measures with compact support, and so different ideas should be necessary to study the Zakai equation case.

5.1 Statement of the problem

In this section we state the problem we are interested in and we collect the necessary assumptions. Let K be a compact subset of $\mathbb{R}^d$ and let $\mathcal{P}(K)$ be the space of probability measures over K. Our aim is to study the following backward Kolmogorov equation with final condition $\Phi \in \mathrm{C}(\mathcal{P}(K))$:

$$\begin{cases} -\partial_t u(\pi,t) - \mathcal{L}^{KS} u(\pi,t) = 0, & (\pi,t) \in \mathcal{P}(K) \times [0,T), \\ u(\pi,T) = \Phi(\pi), & \pi \in \mathcal{P}(K). \end{cases} \tag{5.1}$$

This equation has been introduced in Section 4.3 and it is associated to the Kushner-Stratonovich equation (see Section 1.2 and Section 1.3 for the details)

$$\mathrm{d}\langle \Pi_t, \psi\rangle = \langle \Pi_t, A\psi\rangle \,\mathrm{d}t + (\langle \Pi_t, h\psi + B\psi\rangle - \langle \Pi_t, \psi\rangle\langle \Pi_t, h\rangle) \cdot \mathrm{d}I_t, \quad \psi \in \mathrm{C}^2(K). \tag{5.2}$$

In particular, the differential operator $\mathcal{L}^{KS}\colon \mathrm{C}^2_{\mathrm{L}}(\mathcal{P}(K)) \to \mathrm{C}(\mathcal{P}(K))$ is given by

$$\mathcal{L}^{KS}u(\pi) = \langle \pi, A\delta_\mu u(\pi)\rangle + \frac{1}{2}\langle \pi\otimes\pi, (h + B - \langle\pi, h\rangle)\cdot(h + B - \langle\pi, h\rangle)\delta^2_\mu u(\pi)\rangle,$$

where the operators $A, B\colon \mathrm{C}^2_{\mathrm{b}}(\mathbb{R}^d) \to \mathrm{C}_{\mathrm{b}}(\mathbb{R}^d)$, introduced in Section 1.2, are defined for every $\psi \in \mathrm{C}^2_{\mathrm{b}}(\mathbb{R}^d)$, $x \in \mathbb{R}^d$, by

$$\begin{aligned} A\psi(x) &:= \sum_{i=1}^d f_i(x)\partial_i\psi(x) + \frac{1}{2}\sum_{i,j=1}^d \left\{ \left(\sigma\sigma^\top\right)_{ij}(x)\partial_{ij}\psi(x) + \left(\bar\sigma\bar\sigma^\top\right)_{ij}(x)\partial_{ij}\psi(x) \right\}, \\ B_k\psi(x) &:= \sum_{i=1}^d \bar\sigma_{ik}(x)\partial_i\psi(x), \quad k = 1,\dots d. \end{aligned} \tag{5.3}$$

Thus, the operator $\mathcal{L}^{KS}$ can be written in a more explicit way:

$$\begin{aligned} \mathcal{L}^{KS}u(\pi) &= \langle \pi, A\delta_\mu u(\pi)\rangle + \frac{1}{2}\langle \pi\otimes\pi, (h + B - \langle\pi, h\rangle)\cdot(h + B - \langle\pi, h\rangle)\delta^2_\mu u(\pi)\rangle \\ &= \langle \pi, f\cdot \mathrm{D}_\pi u(\pi)\rangle + \frac{1}{2}\langle \pi, \mathrm{tr}\{\mathrm{D}_x\mathrm{D}_\pi u(\pi)\sigma\sigma^\top\}\rangle + \frac{1}{2}\langle \pi, \mathrm{tr}\{\mathrm{D}_x\mathrm{D}_\pi u(\pi)\bar\sigma\bar\sigma^\top\}\rangle \\ &+ \frac{1}{2}\langle \pi\otimes\pi, \delta^2_\pi u(\pi)h\cdot h\rangle + \frac{1}{2}\langle \pi\otimes\pi, \mathrm{tr}\{\mathrm{D}^2_\pi u(\pi)\bar\sigma\bar\sigma^\top\}\rangle + \frac{1}{2}\langle \pi\otimes\pi, \delta^2_\pi u(\pi)\rangle|\langle\pi, h\rangle|^2 \\ &+ \langle \pi\otimes\pi, h\cdot\bar\sigma^\top\delta_\pi\mathrm{D}_\pi u(\pi)\rangle - \langle \pi\otimes\pi, \delta^2_\pi u(\pi)h\rangle\cdot\langle\pi, h\rangle - \langle \pi\otimes\pi, \bar\sigma^\top\delta_\pi\mathrm{D}_\pi u(\pi)\rangle\cdot\langle\pi, h\rangle. \end{aligned}$$

We recall that the mappings $b, h\colon \mathbb{R}^d \to \mathbb{R}^d, \sigma, \bar\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are related to the filtering problem for a diffusion type signal introduced in Section 1.1.

As in Chapter 4, it will be necessary to assume the following hypotheses.

Assumption 1. *All the mappings $f, \sigma, \bar\sigma, h$ are taken Borel-measurable. Moreover we assume:*

a. the mappings $f\colon \mathbb{R}^d \to \mathbb{R}^d$, $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ and $\bar\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are Lipschitz continuous;

b. the mapping $a := \sigma\sigma^\top\colon \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is uniformly elliptic, that is there exists $\lambda > 0$ such that $\sum_{i,j=1}^d a_{ij}\xi_i\xi_j \geq \lambda|\xi|^2$ for every $x, \xi \in \mathbb{R}^d$;

c. the mappings $f, \sigma, \bar\sigma, h$ are bounded.

Moreover, we will ask for an invariance property. This is a technical assumption necessary in the proof of Theorem 5.2.4 (see also Remark 5.2.6).

Assumption 3. *The probability measure-valued process Π has trajectories confined in the subset $\mathcal{P}(K)$, $K \subset \mathbb{R}^d$ compact.*

For instance, this is the case when Π is the filter associated to a signal process $X = \{X_t, t \in [0, T]\}$ with trajectories confined in a compact subset $K \subset \mathbb{R}^d$. Some conditions on $b, \sigma, \bar\sigma$ that guarantee this property are stated in Remark 5.1.1.

Remark 5.1.1. Let G be a subset of $\mathbb{R}^d$. The invariance of a stochastic process with respect to G has been intensively studied and many criteria that guarantee this property are available in the literature. For instance, following [61, Section 12.2], let $G \subset \mathbb{R}^d$ be closed and with C^3 connected boundary ∂G. Let ν be the outward normal to ∂G and let us define $\rho(x) := \inf_{y \in G} |x - y|$ the distance from $x \in \mathbb{R}^d$ to G. If the conditions

$$\sum_{i,j=1}^{d} a_{ij} \nu_i \nu_j = 0, \qquad \sum_{i=1}^{d} f_i \nu_i + \frac{1}{2} \sum_{i,j=1}^{d} a_{ij} \partial^2_{ij} \rho \geq 0 \quad \text{on } \partial G,$$

hold, where $a = \sigma\sigma^\top + \bar{\sigma}\bar{\sigma}^\top$, then G is invariant for X. Other conditions can be found in [7, 39, 40] and in the references therein.

Remark 5.1.2. All the following results do not change when K is a compact manifold as long as we consider the appropriate differential operators (see for instance [51, 84]). An example is when the signal process lives on the d-dimensional torus $\mathbb{T}^d$.

Remark 5.1.3. The results obtained in Section 4.3, namely the Itô formula in Proposition 3.2.1 and Theorem 4.3.3, hold without any modification in this compact framework. We have only to take in account the remarks about the derivatives given in Section 2.1.2.

5.1.1 Some analytical results

Here we introduce some analytical tools that will be useful in the discussion below, and in particular for the proof of the comparison principle (Theorem 5.2.4). For the entire chapter, we will deal with probability measures in $\mathcal{P}(K)$, where K is a compact subset of $\mathbb{R}^d$. Since K is compact, we have that the set $\mathcal{P}(K)$ coincides with the set of probability measures with finite moment of any order. Moreover, the compactness of K is inherited by $\mathcal{P}(K)$, which is also compact in the weak topology.

Let us consider a family $\{\xi_k\}_{k\in\mathbb{N}}$ of functions dense in $\mathrm{C}^\infty_{\mathrm{c}}(K)$, with $\sup_{x\in K}|\xi_k(x)| =: \|\xi_k\|_\infty \leq 1$, and let us assume that the function identically equal to 1 belongs to this family. Let us introduce a sequence $\{q_k\}_{k\in\mathbb{N}} \subset [1, +\infty)$ that we will make explicit later in this section. We define $\mathsf{d}_2 \colon \mathcal{P}(K) \times \mathcal{P}(K) \to [0, +\infty)$ as

$$\mathsf{d}_2(\mu, \nu) := \left(\sum_{k=1}^{+\infty} \frac{1}{2^k q_k} \langle \mu - \nu, \xi_k \rangle^2 \right)^{\frac{1}{2}}. \tag{5.4}$$

It is easy to show that d_2 is a distance over $\mathcal{P}(K)$ and, since K compact, one can also prove that it is complete (see [21, Theorem 2.4.1]). Moreover, d_2 metricize the weak convergence ([21, Theorem 2.1.1]).

An important property we need to study is the differentiability of the distance d_2 introduced in (5.4). For what follows, we need to choose

$$q_k := \max\left\{1, \|\mathrm{D}_x\xi_k\|_\infty, \|\mathrm{D}^2_x\xi_k\|_\infty, \|\mathrm{D}_x\xi_k\|^2_\infty\right\}, \quad k \in \mathbb{N}.$$

Let $\mu_0 \in \mathcal{P}(K)$ be fixed and let $\rho_{\mu_0} \colon \mathcal{P}(K) \to [0, +\infty)$ be defined as

$$\rho_{\mu_0}(\mu) := \mathsf{d}_2(\mu_0, \mu)^2, \quad \mu \in \mathcal{P}(K). \tag{5.5}$$

Lemma 5.1.4. *Let $\mu_0 \in \mathcal{P}(K)$ be fixed. Then, the mapping ρ_{μ_0} is in $\mathrm{C}^2_{\mathrm{L}}(\mathcal{P}(K))$ with derivatives uniformly bounded with respect to μ_0.*

Proof. First, let us fix $k \in \mathbb{N}$ and show that the mapping $\mu \mapsto \langle \mu - \mu_0, \xi_k \rangle^2 =: g_k(\mu)$, where ξ_k is in the family appearing in (5.4), is in $\mathrm{C}^2_{\mathrm{L}}(\mathcal{P}(K))$. By standard computations we get

$$\begin{aligned} \delta_\mu g_k(\mu, x) &= 2\langle \mu - \mu_0, \xi_k \rangle \xi_k(x), \qquad && \mu \in \mathcal{P}(K), x \in K, \\ \delta^2_\mu g_k(\mu, x, y) &= 2\xi_k(x)\xi_k(y), && \mu \in \mathcal{P}(K), x, y \in K, \end{aligned} \tag{5.6}$$

with $\|\delta_\mu g_k\|_\infty \leq 4$ and $\|\delta^2_\mu g_k\|_\infty \leq 2$ for every $k \in \mathbb{N}$. Moreover, we have that

$$\begin{aligned} \mathrm{D}_\mu g_k(\mu, x) &= 2\langle \mu - \mu_0, \xi_k \rangle \mathrm{D}_x \xi_k(x), \qquad && \mu \in \mathcal{P}(K), x \in K, \\ \mathrm{D}_x \mathrm{D}_\mu g_k(\mu, x) &= 2\langle \mu - \mu_0, \xi_k \rangle \mathrm{D}^2_x \xi_k(x), && \mu \in \mathcal{P}(K), x \in K, \\ \mathrm{D}^2_\mu g_k(\mu, x, y) &= 2\mathrm{D}_x \xi_k(x) \mathrm{D}^\top_y \xi_k(y), && \mu \in \mathcal{P}(K), x, y \in K, \\ \delta_\mu \mathrm{D}_\mu g_k(\mu, x, y) &= 2\mathrm{D}_x \xi_k(x) \xi_k(y), && \mu \in \mathcal{P}(K), x, y \in K, \end{aligned}$$

with

$$\begin{aligned} \|\mathrm{D}_\mu g_k\|_\infty &\leq 4\|\mathrm{D}_x \xi_k\|_\infty, \qquad & \|\mathrm{D}_x \mathrm{D}_\mu g_k\|_\infty &\leq 4\|\mathrm{D}^2_x \xi_k\|_\infty, \\ \|\mathrm{D}^2_\mu g\|_\infty &\leq 2\|\mathrm{D}_x \xi_k\|^2_\infty, & \|\delta_\mu \mathrm{D}_\mu g_k\|_\infty &\leq 2\|\mathrm{D}_x \xi_k\|_\infty. \end{aligned}$$

To conclude, by dominated convergence we can bring the derivative inside the series and get

$$\delta_\mu \rho_{\mu_0}(\mu, x) = \sum_{k=1}^{+\infty} \frac{1}{2^k q_k} \delta_\mu g_k(\mu, x), \quad \mu \in \mathcal{P}(K), x \in K,$$

which is also jointly continuous. Moreover, recalling the definition of $\{q_k\}_{k \in \mathbb{N}}$, we have

$$\|\delta_\mu \rho_{\mu_0}\|_\infty \leq \sum_{k=1}^{+\infty} \frac{1}{2^k q_k} \|\delta_\mu g_k\|_\infty \leq 4.$$

We can proceed in the same way for the other derivatives and the bounds follow easily. $\square$

Remark 5.1.5. Let $\mathcal{L}^{KS}$ be the differential operator introduced in Section 5.2. Under Assumption 1, we have that from Lemma 5.1.4 follows that there exists a constant (independent of μ_0) $C > 0$ such that $\|\mathcal{L}^{KS} \rho_{\mu_0}\|_\infty = \sup_{\mu \in \mathcal{P}(K)} |(\mathcal{L}^{KS} \rho_{\mu_0})(\mu)| \leq C$. Indeed

$$\begin{aligned} \|\mathcal{L}^{KS} \rho_{\mu_0}\|_\infty &\leq \|A \delta_\mu u\|_\infty + \frac{1}{2} \|(h + B - \langle \mu, h \rangle)^\top (h + B - \langle \mu, h \rangle) \delta^2_\mu u\|_\infty \\ &\leq 4\|b\|_\infty + 2\|\sigma\|^2_\infty + 3\|\bar{\sigma}\|^2_\infty + 4\|h\|^2_\infty + 4\|h\|_\infty \|\bar{\sigma}\|_\infty =: C. \end{aligned} \tag{5.7}$$

To conclude this auxiliary section, we provide a class of polynomials on the space of probability measures that can be used to approximate functions in $\mathrm{C}(\mathcal{P}(K))$, which is the set of continuous real-valued functions over $\mathcal{P}(K)$. This kind of approximation has been discussed for instance in [38] and allows us to approximate continuous functions over $\mathcal{P}(K)$ with functions in $\mathrm{C}^2_{\mathrm{L}}(\mathcal{P}(K))$.

Definition 5.1.6. *We define the set of polynomials over $\mathcal{P}(K)$:*

$$\mathscr{P}(\mathcal{P}(K)) \\ := \{\phi\colon \mathcal{P}(K) \to \mathbb{R} \text{ of the form } \phi(\mu) = g(\langle \mu, \psi_1\rangle, \dots, \langle \mu, \psi_n\rangle), \text{ for some } n \geq 1\},$$

with $g\colon \mathbb{R}^n \to \mathbb{R}$ and $\psi_i\colon \mathbb{R}^d \to \mathbb{R}$ polynomials, $i = 1, \dots, n$.

Remark 5.1.7. It holds that $\mathscr{P}(\mathcal{P}(K)) \subset \mathrm{C}^2_{\mathrm{L}}(\mathcal{P}(K))$. Indeed, every $\phi \in \mathscr{P}(\mathcal{P}(K))$ is a cylindrical function, which means that it is of the form discussed in Example 2.1.11.

One can prove (see for instance [38, Section 2.3]) that the family $\mathscr{P}(\mathcal{P}(K))$ is an algebra in $\mathcal{P}(K)$ that separates the points and which contains the constants. Thus, by Stone-Weiestrass theorem one can conclude that it is dense in $\mathrm{C}(\mathcal{P}(K))$:

Proposition 5.1.8. *The family $\mathscr{P}(\mathcal{P}(K))$ is dense in $\mathrm{C}(\mathcal{P}(K))$ with respect to the supremum norm over $\mathcal{P}(K)$.*

5.2 Viscosity solutions to the backward Kolmogorov equation

In this section we continue the study of the backward Kolmogorov equation (5.1) associated to the Kushner-Stratonovich equation (5.2). We proved in Section 4.3, Theorem 4.3.3, that under Assumption 1 there exists a unique classical solution to (5.1) when the final condition is chosen smooth enough (namely $\mathrm{C}^2_{\mathrm{L}}(\mathcal{P}(K))$). In particular, the solution is given by the usual formula

$$u(\pi, t) = \mathbb{E}^{\mathbb{P}}\left[\Phi(\Pi^{\pi;t}_T)\right], \quad (\pi, t) \in \mathcal{P}(K) \times [0, T], \tag{5.8}$$

where $\{(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P}), (\Pi^{\pi,s}, I)\}$ is the weak solution to the Kushner-Stratonovich equation obtained starting from a solution to the Zakai equation, as explained in Section 4.3.

Here we investigate the well-posedness of (5.1) when the final condition has lower regularity, namely $\Phi \in \mathrm{C}(\mathcal{P}(K))$. In this case classical solutions may not exist, and a natural way to face the problem is to consider a generalized notion of solution. In particular we will focus on the notion of viscosity solution:

Definition 5.2.1. *A upper semicontinuous (resp. lower semicontinuous) function $u\colon \mathcal{P}(K) \times [0, T] \to \mathbb{R}$ is a viscosity subsolution (resp. supersolution) to equation* (5.1) *if:*

i. $u(\pi, T) \leq$ (resp. $\geq$) $\Phi(\pi)$, for every $\pi \in \mathcal{P}(K)$;

ii. for every $(\pi, t) \in \mathcal{P}(K) \times [0, T)$ and any $\varphi \in \mathrm{C}^{2,1}_{\mathrm{L}}(\mathcal{P}(K) \times [0, T])$ such that $u - \varphi$ has a global maximum (resp. minimum) at (π, t) with value 0, then (5.1) *holds for φ with the inequality sign $\leq$ (resp. $\geq$) instead of the equality.*

We say that u is a viscosity solution if it is both a viscosity subsolution and a viscosity supersolution.

In the definition above, we denoted by $\mathrm{C}^{2,1}_{\mathrm{L}}(\mathcal{P}(K) \times [0, T])$ the space of continuous functions $u\colon \mathcal{P}(K) \times [0, T] \to \mathbb{R}$ of class $\mathrm{C}^2_{\mathrm{L}}(\mathcal{P}(K))$ with respect to the measure argument and in $\mathrm{C}^1([0, T])$ with respect to the time argument.

5.2.1 Existence of viscosity solutions

Let us introduce a function u by the usual probabilistic representation formula (5.8), where $\Phi \in \mathrm{C}(\mathcal{P}(K))$ is the final condition of (5.1) and $\{(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P}), (\Pi^{\pi,s}, I)\}$ is the solution to the Kushner-Stratonovich equation introduced in Section 4.3. To have a lighter notation, in what follows we will write $\mathbb{E}$ instead of $\mathbb{E}^{\mathbb{P}}$.

Proposition 5.2.2. *Let $\Phi \in \mathrm{C}(\mathcal{P}(K))$, let Assumption 1 holds and let $\Pi^{\pi,s}$ be defined as above and such that is satisfies Assumption 3. Then the function*

$$u(\pi, t) := \mathbb{E}\left[\Phi(\Pi_T^{\pi,t})\right], \quad (\pi, t) \in \mathcal{P}(K) \times [0, T],$$

is a viscosity solution to the backward Kolmogorov equation (5.1).

Proof. We only prove that u is a subsolution, since proving that it is a supersolution is analogue. First, thanks to the continuity of Φ and $\Pi^{\pi,s}$, we notice that u is continuous and bounded. Moreover $u(\pi, T) = \mathbb{E}\left[\Phi(\Pi_T^{\pi,T})\right] = \Phi(\pi)$ for every $\pi \in \mathcal{P}(K)$.

Now, let us fix $(\pi, t) \in \mathcal{P}(K) \times [0, T)$ and let us consider $\varphi \in \mathrm{C}^{2,1}_{\mathrm{L}}(\mathcal{P}(K) \times [0, T])$ such that $\varphi(\pi, t) = u(\pi, t)$ and $\varphi(\nu, s) \geq u(\nu, s)$ for every $(\nu, s) \in \mathcal{P}(K) \times [0, T)$. For $h > 0$ it holds that

$$0 \leq \frac{\varphi(\Pi_{t+h}^{\pi,t}, t+h) - u(\Pi_{t+h}^{\pi,t}, t+h)}{h}.$$

Moreover, by taking the expectation combined with Itô formula and Markov property of $\Pi^{\pi,t}$ (as in the proof of Theorem 4.2.5), we obtain

$$0 \leq \frac{1}{h}\mathbb{E}\left[\int_t^{t+h} \mathcal{L}^{KS}\varphi(\Pi_\tau^{\pi,t}, t+h)\,\mathrm{d}\tau\right] + \frac{\varphi(\pi, t+h) - \varphi(\pi, t)}{h},$$

where u disappeared thanks to the fact that

$$\mathbb{E}\left[u(\Pi_{t+h}^{\pi,t}, t+h)\right] = u(\pi, t) = \varphi(\pi, t). \tag{5.9}$$

Finally, thanks to the regularity of φ and its derivatives, we can pass to the limit as $h \to 0$ and obtain

$$0 \leq \mathcal{L}^{KS}\varphi(\pi, t) + \partial_t\varphi(\pi, t).$$

□

Remark 5.2.3. This existence result is still valid even if we drop Assumption 3 and we look at the equation over $\mathcal{P}_2(\mathbb{R}^d)$.

5.2.2 Comparison theorem and uniqueness

Here we state and prove a comparison principle for viscosity solutions to (5.1), which is a crucial tool to prove the uniqueness. The technique for the proof is inspired by the one used for Theorem II.1 in [78], and by its recent refinement used in [35]. In particular, we use the solution u introduced by the representation formula (5.8) as a reference to obtain partial comparison results.

Theorem 5.2.4. *Let $v_1, v_2 \colon \mathcal{P}(K) \times [0,T] \to \mathbb{R}$ be respectively a viscosity subsolution and supersolution of* (5.1)*. Moreover, let Assumption 1 and Assumption 3 hold. Then,* $v_1(\pi, t) \le v_2(\pi, t)$ *for every* $(\pi, t) \in \mathcal{P}(K) \times [0,T]$.

Proof. We want to prove that $v_1 \le u \le v_2$, where u is given by (5.8). The proof is divided into five steps and we show only that $v_1 \le u$. Indeed we can prove $u \le v_2$ by noticing that v_2 is a subsolution of (5.1) with final condition $-\Phi$ and by taking $-u$ instead of u.

By contradiction, let us assume that there exists $(\pi_0, t_0) \in \mathcal{P}(K) \times [0,T]$ such that

$$v_1(\pi_0, t_0) - u(\pi_0, t_0) > 0. \tag{5.10}$$

First, we notice that $t_0 < T$, indeed $v_1(\pi, T) \le \Phi(\pi) = u(\pi, T)$ for every $\pi \in \mathcal{P}(K)$, thanks to the subsolution property of v_1.

Step 1. Let $\{\Phi_n\}_{n \ge 1} \subset \mathrm{C}^2_{\mathrm{L}}(\mathcal{P}(K))$ be the sequence that converges uniformly to $\Phi \in \mathrm{C}(\mathcal{P}(K))$ given by Proposition 5.1.8. For every $n \ge 1$, setting

$$u_n(\pi, t) := \mathbb{E}\left[\Phi_n(\Pi_T^{\pi,t})\right], \quad (\pi, t) \in \mathcal{P}(K) \times [0,T],$$

we can say from Theorem 4.3.3 that u_n is the unique classical solution to equation (5.1) with final condition Φ_n. Since $\{\Phi_n\}_{n \ge 1}$ converges uniformly to Φ, the sequence is bounded by a positive constant $M = M(\Phi)$. Then, dominated convergence allows us to conclude that

$$\lim_{n \to +\infty} u_n(\pi, t) = u(\pi, t), \quad (\pi, t) \in \mathcal{P}(K) \times [0,T]. \tag{5.11}$$

Step 2. Let us introduce

$$\bar{v}_1(\pi, t) := \mathrm{e}^{t - t_0} v_1(\pi, t), \quad \bar{u}_n(\pi, t) := \mathrm{e}^{t - t_0} u_n(\pi, t), \quad (\pi, t) \in \mathcal{P}(K) \times [0,T), \tag{5.12}$$

and

$$\bar{\Phi}(\pi) := \mathrm{e}^{T - t_0} \Phi(\pi), \quad \bar{\Phi}_n(\pi) := \mathrm{e}^{T - t_0} \Phi_n(\pi), \quad \pi \in \mathcal{P}(K).$$

Then, it is easy to see that $\bar{u}_n$ is a classical solution of the equation

$$\begin{cases} -\partial_t v(\pi, t) - \mathcal{L}^{KS} v(\pi, t) = -v(\pi, t), & (\pi, t) \in \mathcal{P}(K) \times [0,T), \\ v(\pi, T) = \bar{\Phi}_n(\pi), & \pi \in \mathcal{P}(K), \end{cases} \tag{5.13}$$

and, analogously, that $\bar{v}_1$ is a viscosity subsolution of

$$\begin{cases} -\partial_t v(\pi, t) - \mathcal{L}^{KS} v(\pi, t) = -v(\pi, t), & (\pi, t) \in \mathcal{P}(K) \times [0,T), \\ v(\pi, T) = \bar{\Phi}(\pi), & \pi \in \mathcal{P}(K). \end{cases}$$

Step 3. For every $\lambda > 0$, let us define the mapping $\Psi_\lambda \colon \mathcal{P}(K) \times [0,T] \to \mathbb{R}$ as

$$\Psi_\lambda(\pi, t) := \bar{v}_1(\pi, t) - \bar{u}_n(\pi, t) - \lambda (t - t_0)^2 - \lambda \mathsf{d}_2(\pi_0, \pi)^2. \tag{5.14}$$

Since $\bar{v}_1$ is upper semicontinuous over a compact space, and $\bar{u}_n$ is continuous and bounded uniformly in n, there exists $(\pi_\lambda, t_\lambda) \in \mathcal{P}(K) \times [0,T]$ global maximum for Ψ_λ. Moreover, from (5.14) it follows that

$$\bar{v}_1(\pi_0, t_0) - \bar{u}_n(\pi_0, t_0) = \Psi_\lambda(\pi_0, t_0) \leq \Psi(\pi_\lambda, t_\lambda) \leq \bar{v}_1(\pi_\lambda, t_\lambda) - \bar{u}_n(\pi_\lambda, t_\lambda), \tag{5.15}$$

and so

$$\lambda(t_\lambda - t_0)^2 + \lambda \mathsf{d}_2(\pi_0, \pi_\lambda)^2 \leq (\bar{v}_1(\pi_\lambda, t_\lambda) - \bar{u}_n(\pi_\lambda, t_\lambda)) - (\bar{v}_1(\pi_0, t_0) - \bar{u}_n(\pi_0, t_0)) \,. \tag{5.16}$$

Thanks to (5.10) and (5.11), for n large enough we have that

$$(\bar{v}_1(\pi_0, t_0) - \bar{u}_n(\pi_0, t_0)) > 0.$$

Combining this with the fact that $\bar{v}_1$ is bounded from above and $\bar{u}_n$ is bounded uniformly in n, from (5.16) we obtain that there exists a constant $M > 0$ such that

$$|t_\lambda - t_0| \leq \sqrt{\frac{M}{\lambda}}, \quad \mathsf{d}_2(\pi_0, \pi_\lambda) \leq \sqrt{\frac{M}{\lambda}}. \tag{5.17}$$

Step 4. Now we show that $t_\lambda < T$. By contradiction, let $t_\lambda = T$. From (5.15),

$$v_1(\pi_0, t_0) - u_n(\pi_0, t_0) = \bar{v}_1(\pi_0, t_0) - \bar{u}_n(\pi_0, t_0) \leq \bar{v}_1(\pi_\lambda, T) - \bar{u}_n(\pi_\lambda, T),$$

and then

$$v_1(\pi_0, t_0) - u_n(\pi_0, t_0) \leq \mathrm{e}^{T - t_0} \left(\Phi(\pi_\lambda) - \Phi_n(\pi_\lambda) \right). \tag{5.18}$$

Letting $n \to \infty$, since $\Phi_n \to \Phi$ uniformly (see Proposition 5.1.8) and $u_n(\pi_0, t_0) \to u(\pi_0, t_0)$, we obtain

$$v_1(\pi_0, t_0) - u(\pi_0, t_0) \leq 0,$$

which contradicts (5.10).

Step 5. Let us define $\mathcal{P}(K) \times [0,T] \ni (\pi, t) \mapsto \varphi_\lambda(\pi, t) := \lambda(t - t_0)^2 + \lambda \mathsf{d}_2(\pi_0, \pi)^2$ and let us notice that $\varphi_\lambda \in \mathrm{C}^{2,1}_{\mathrm{L}}(\mathcal{P}(K) \times [0,T])$ thanks to Lemma 5.1.4. We want to use $\bar{u}_n + \varphi_\lambda$ as test function for the viscosity subsolution property of $\bar{v}_1$. First, $\bar{u}_n + \varphi_\lambda$ is in $\mathrm{C}^{2,1}_{\mathrm{L}}(\mathcal{P}(K) \times [0,T])$ since $\bar{u}_n, \varphi_\lambda \in \mathrm{C}^{2,1}_{\mathrm{L}}(\mathcal{P}(K) \times [0,T])$. Moreover, from *Step 3* we have that $\bar{v}_1 - (\bar{u}_n + \varphi_\lambda)$ has a maximum in (π_λ, t_λ) and from *Step 4* it holds that $t_\lambda < T$.

Thus, we can actually use $\bar{u}_n + \varphi_\lambda$ as a test function in (π_λ, t_λ) and, by the viscosity subsolution property of $\bar{v}_1$, obtain

$$\bar{v}_1(\pi_\lambda, t_\lambda) \leq \partial_t(\bar{u}_n + \varphi_\lambda)(\pi_\lambda, t_\lambda) + \mathcal{L}^{KS}(\bar{u}_n + \varphi_\lambda)(\pi_\lambda, t_\lambda).$$

Since $\bar{u}_n$ is a classical solution of (5.13), this entails

$$\bar{v}_1(\pi_\lambda, t_\lambda) - \bar{u}_n(\pi_\lambda, t_\lambda) \leq \partial_t \varphi_\lambda(\pi_\lambda, t_\lambda) + \mathcal{L}^{KS} \varphi_\lambda(\pi_\lambda, t_\lambda).$$

Now we can use the estimates (5.17) and the bound in Remark 5.1.5 to obtain, for a positive constant $C > 0$,

$$\bar{v}_1(\pi_\lambda, t_\lambda) - \bar{u}_n(\pi_\lambda, t_\lambda) \leq 2\lambda(t_\lambda - t_0) + \lambda(\mathcal{L}\rho_{\pi_0})(\pi_\lambda) \leq 2\sqrt{\lambda M} + \lambda C,$$

where ρ_{π_0} is defined as in (5.5). Finally, by (5.15) and (5.12), it follows that

$$v_1(\pi_0, t_0) - u_n(\pi_0, t_0) = \bar{v}_1(\pi_0, t_0) - \bar{u}_n(\pi_0, t_0) \leq 2\sqrt{\lambda M} + \lambda C(\pi_0), \tag{5.19}$$

which contradicts (5.10) if we let $\lambda \to 0$ and $n \to \infty$. □

Corollary 5.2.5. *Let* $\Phi \in \mathrm{C}(\mathcal{P}(K))$ *and let Assumption 1 and Assumption 3 hold. Then*

$$u(\pi, t) = \mathbb{E}\left[\Phi(\Pi_T^{\pi,t})\right], \quad (\pi, t) \in \mathcal{P}(K) \times [0, T],$$

is the unique viscosity solution to the backward Kolmogorov equation (5.1)*.*

Proof. Let u be the solution to (5.1) defined by (5.8) and let v be another viscosity solution. Since in particular u is a subsolution and v is a supersolution, Theorem 5.2.4 tells us that $u \leq v$. Changing the role of u and v we obtain that $v \leq u$ and so the two solutions coincide. □

Remark 5.2.6. The technical Assumption 3 and in particular the necessity of considering the equation (5.1) over $\mathcal{P}(K)$ is due to a property of the distance d_2. In particular, given $S \subseteq \mathbb{R}^d$ and defined d_2 over $\mathcal{P}(S)$ as in (5.4), it holds that d_2 is complete if and only if S is compact (see [21, Theorem 2.4.1]). This fact prevent us to reproduce the argument presented in [36, 35] to extend our results to $\mathcal{P}(\mathbb{R}^d)$. Indeed that approach is based on a smooth variational principle (see for instance [22, Theorem 2.5.2]) which requires a complete metric on the space.

Chapter 6

Mean Field Optimal Stopping: a limit approach

In this last chapter, we present a problem that is completely different from the one presented before, even if in both cases the main difficulty is due to the fact that some relevant variables take values in a space space of measures. We study a class of finite horizon time-inconsistent optimal stopping problems of mean field type, adapted to the Brownian filtration, including those related to mean field diffusion processes and recursive utility functions. In particular, the mean field interaction is due to the fact that we consider a terminal cost depending not only on the stopped process but also on its law. We highlight the fact that this differs from the case where the law of the process is evaluated at the stopping time, and, due to this, the problem becomes time-inconsistent.

Optimal stopping problems (OSPs) with mean field interaction arise for instance when the goal is to minimize the variance or, more generally, when the cost depends also on the mean of the stopped process. In the work by Pedersen and Peskir [88, 89]), the problems of optimal variance stopping and optimal mean-variance stopping have been investigated in the case where the underlying process is for instance a Geometric Brownian motion, highlighting connections with portfolio choice. Even though the OSP of the variance is *time-inconsistent*, which means that the value process is not a supermartingale, they succeeded to solve the problem and they derived a variational inequality for the value function with an explicit stopping region. More recently, the interest in optimal stopping problems with a more general mean field type interaction such as the dependence on the law of the stopped process increased significantly, due mainly to the connection with the theory of mean field games and mean field optimal control (see for instance [31] for a systematic presentation of the topic). The first contribution in this direction is the work by Bertucci [15], where an optimal stopping problem for a mean field game is studied using mainly PDE techniques. Then, other results and extensions have been obtained with different techniques in [33, 85, 23, 47] and in the recent papers [100, 48, 1].

Our goal in this chapter is to provide a characterization of an optimal stopping time as a hitting time, and more precisely we show that it is optimal to stop when the value-process hits the reward process for the first time, as in the case for the standard time-consistent optimal stopping problem. We solve the problem by approximating the corresponding value-process with a sequence of Snell envelope of processes, for which a sequence of optimal stopping times is constituted of hitting times of each of the reward processes by the associated value-process.

Throughout this chapter, we only consider the one-dimensional Brownian motion and diffusion processes. The generalization to the multidimensional case is straightforward.

6.1 Optimal stopping of a recursive utility function

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and $T > 0$ a finite time horizon. Let $W = \{W_t, t \in [0,T]\}$ be a standard one-dimensional Brownian motion. We denote by $\mathbb{F} = \{\mathcal{F}_t\}_{t\in[0,T]}$ the (completed) natural filtration of the Brownian motion W, with $\mathcal{F}_0 = \{\emptyset, \Omega\}$. In particular, $\mathbb{F}$ is continuous, that is for each $t \geq 0$ it holds $\mathcal{F}_{t^-} = \mathcal{F}_t$. We denote by $\mathcal{T}_t$ the set of $\mathbb{F}$-stopping times τ such that $\tau \in [t,T]$ almost surely.

Let us introduce the recursive value-process

$$Y_t = \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_t} \mathbb{E}\left[h(Y_\tau, \mathbb{E}[Y_\tau])\mathbb{1}_{\{\tau<T\}} + \xi\mathbb{1}_{\{\tau=T\}} \,|\, \mathcal{F}_t\right], \tag{6.1}$$

where h is a sufficiently smooth cost (see Assumption 4 below) and $\xi \in L^2(\mathcal{F}_T)$. The (simplified) finite horizon optimal stopping problem (OSP) of mean field type associated to (6.1) reads as:

$$Y_0 = \sup_{\tau\in\mathcal{T}_0} \mathbb{E}\left[h(Y_\tau, \mathbb{E}[Y_\tau])\mathbb{1}_{\{\tau<T\}} + \xi\mathbb{1}_{\{\tau=T\}}\right]. \tag{6.2}$$

The OSP (6.2) is time-inconsistent, that is the associated value-process is no longer a supermartingale, due to the presence of expected value (law) of the stopped random variable Y_τ.

Recursive utilities have been introduced in [45, 46] and they have been employed in various frameworks, such as economic, mathematical finance, and insurance (see, for instance, [55, 56] and the references therein for the classical theory and some relevant examples and applications, or [44] for a more recent discussion concerning a mean field model that can be related to life insurance problems).

We would like to investigate whether the value-process Y is well-defined, that is whether there exists a unique solution to (6.1) and whether there exists a optimal stopping time τ^* to the OSP (6.2):

$$\tau^* = \operatorname*{arg\,max}_{\tau\in\mathcal{T}_0} \mathbb{E}\left[h(Y_\tau, \mathbb{E}[Y_\tau])\mathbb{1}_{\{\tau<T\}} + \xi\mathbb{1}_{\{\tau=T\}}\right]. \tag{6.3}$$

We suggest to solve this problem by using a limit approach based on approximating $\mathbb{E}[Y_\cdot]$ by its empirical mean $\frac{1}{n}\sum_{j=1}^n Y_\cdot^{j,n}$ for some suitable sample $Y^{i,n}$, $i = 1, 2, \ldots, n$ of 'interacting' value-processes which solve a system of standard OSPs.

To this end we set $W^1 = W$ and let $\{W^i\}_{i\geq 1}$ be independent Brownian motions and denote by $\mathbb{G}^n := \{\mathcal{G}^n_t\}_{t\in[0,T]}$ the completion of the filtration generated by $\{W^i\}_{i=1}^n$. Thus, for each $1 \leq i \leq n$, the filtration generated by W^i is a sub-filtration of $\mathbb{G}^n$, and we denote by $\mathbb{F}^i := \{\mathcal{F}^i_t\}_{t\in[0,T]}$ its completion. Let $\mathcal{T}^n_t$ (resp. $\mathcal{T}^i_t$) be the set of $\mathbb{G}^n$ (resp. $\mathbb{F}^i$) stopping times with values in $[t, T]$.

Consider the following family of finite horizon stopping problems.

$$Y^{i,n}_t = \operatorname*{ess\,sup}_{\tau\in\mathcal{T}^i_t} \mathbb{E}\left[h(Y^{i,n}_\tau, \frac{1}{n}\sum_{j=1}^n Y^{j,n}_\tau)1\!\!1_{\{\tau<T\}} + \xi^i 1\!\!1_{\{\tau=T\}} \,|\, \mathcal{F}^i_t\right], \quad i = 1, 2, \ldots, n, \tag{6.4}$$

and

$$Y^i_t = \operatorname*{ess\,sup}_{\tau\in\mathcal{T}^i_t} \mathbb{E}\left[h(Y^i_\tau, \mathbb{E}[Y^i_\tau])1\!\!1_{\{\tau<T\}} + \xi^i 1\!\!1_{\{\tau=T\}} \,|\, \mathcal{F}^i_t\right], \quad i \geq 1, \tag{6.5}$$

where h and $\{\xi^i\}_{i\geq 1}$ satisfies the following conditions.

Assumption 4. *The coefficients h and $\{\xi^i\}_{i\geq 1}$ satisfy the following conditions:*

(i) For each $i \geq 1$, $\xi^i \in L^2(\mathcal{F}^i_T)$. Moreover, the ξ^i's are independent copies of ξ, with $\xi^1 = \xi$;

(ii) the function $h\colon \Omega \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$

(ii-a) is Lipschitz continuous with respect to (y, z): there exist two positive constants γ_1 and γ_2 such that

$$|h(\omega, y_1, z_1) - h(\omega, y_2, z_2)| \leq \gamma_1|y_1 - y_2| + \gamma_2|z_1 - z_2|,$$

for any $\omega \in \Omega$ and any $y_1, y_2, z_1, z_2 \in \mathbb{R}$.

(ii-b) $h(\cdot, 0, 0) \in L^2(\mathbb{P})$.

6.1.1 Existence and uniqueness of the value-processes

In this section we show the well-posedness of the systems (6.5) and (6.4), since in both cases the performance function depends also on the value-process.

Theorem 6.1.1. *Suppose that Assumption 4 is in force. Assume further that γ_1 and γ_2 satisfy*

$$\gamma_1^2 + \gamma_2^2 < \frac{1}{2}. \tag{6.6}$$

Then there exists a unique solution in $\mathcal{S}^2_c$ to each of the systems (6.4) *and* (6.5)*.*

The proof of Theorem (6.1.1) is based on the fixed point argument used in the proof of Theorem 2.1 in [43]. We omit to reproduce it here.

Corollary 6.1.2. *For each $i = 1, \dots n$, the $\mathcal{F}^i$-stopping time*

$$\hat{\tau}^{i,n} = \inf\left\{ t \geq 0,\ Y_t^{i,n} = \mathbb{E}[h(Y_t^{i,n}, \frac{1}{n}\sum_{j=1}^{n} Y_t^{j,n}) \,|\, \mathcal{F}_t^i] \right\} \wedge T \tag{6.7}$$

is optimal for $Y_0^{i,n}$.

Proof. Since each of the processes $Y^{i,n}$ is in $\mathcal{S}_c^2$, by Assumption 4 (ii) and Doob's inequality, it follows that the obstacle process $\mathcal{X}_t^{i,n} := \mathbb{E}[h(Y_t^{i,n}, \frac{1}{n}\sum_{j=1}^{n} Y_t^{j,n}) \,|\, \mathcal{F}_t^i]$ is also in $\mathcal{S}^2$ and thus is càdlàg. Moreover, since for each $t \in [0,T]$, both $h(Y_t^{i,n}, \frac{1}{n}\sum_{j=1}^{n} Y_t^{j,n})$ and $\mathcal{F}_t^i$ are continuous, the optional and predictable projections of $h(Y_\cdot^{i,n}, \frac{1}{n}\sum_{j=1}^{n} Y_\cdot^{j,n})$ with respect to $\mathbb{F}^i$ coincide. This implies that the obstacle process $\mathcal{X}^{i,n}$ is almost surely continuous. Therefore, for instance by Theorem D.19 in Appendix D in [68] or Proposition 1.1.8 in [90] (see also [17] and [53] for a more general set up), for each $i = 1, 2, \dots, n$, the stopping time $\tau^{i,n}$ given by (6.7) is optimal for $Y_0^{i,n}$. □

6.2 Convergence results

The main aim of this section is to characterize an optimal stopping time for the time inconsistent problem (6.2). The idea is to exploit the time consistency of the system of interacting optimal stopping problems

$$Y_0^{i,n} = \sup_{\tau \in \mathcal{T}_0^i} \mathbb{E}\left[h(Y_\tau^{i,n}, \frac{1}{n}\sum_{j=1}^{n} Y_\tau^{j,n}) \mathbb{1}_{\{\tau < T\}} + \xi^i \mathbb{1}_{\{\tau = T\}} \right], \quad i = 1, 2, \dots, n,$$

to obtain an explicit sequence of optimal stopping times, and then to show that this sequence converges to an optimal stopping time for the OSP (6.2).

6.2.1 Convergence of the particle system

This subsection is concerned with the convergence of the process $Y^{i,n}$, solution of the particle system (6.4), to the solution Y^i of the mean field system (6.5).

Definition 6.2.1 (Exchangeable r.v.). *The random variables $X^1, X^2, \dots, X^n$ are said to be exchangeable if the law of the random vector $(X^1, X^2, \dots, X^n)$ is the same as that of the random vector $(X^{\sigma(1)}, X^{\sigma(2)}, \dots, X^{\sigma(n)})$ for every permutation σ of the set $\{1, 2, \dots, n\}$. We write*

$$\mathfrak{L}(X^1, X^2, \dots, X^n) = \mathfrak{L}(X^{\sigma(1)}, X^{\sigma(2)}, \dots, X^{\sigma(n)}).$$

In the following proposition, we show that the exchangeability property of the final conditions $\{\xi^i\}_{i\geq 1}$, entailed by Assumption 4 (i), transfers to the solutions of the systems (6.4) and (6.5).

Proposition 6.2.2 (Exchangeability property). *Let Assumption 4 hold. Then the processes $\{Y^{i,n}\}_{i=1}^n$ solution of the system* (6.4) *are exchangeable. Moreover, the processes $\{Y^i\}_{i\geq 1}$ solution of the system* (6.5) *are independent and equally distributed and hence exchangeable.*

Proof. First let us focus on $\{Y^{i,n}\}_{i=1}^n$. For any permutation σ of the set $\{1,2,\ldots,n\}$, we have for any $t \in [0,T]$

$$\frac{1}{n}\sum_{j=1}^n Y_t^{j,n} = \frac{1}{n}\sum_{j=1}^n Y_t^{\sigma(j),n},$$

and thanks to the uniqueness result in Theorem 6.1.1 we have

$$(Y^{1,n}, Y^{2,n}, \ldots, Y^{n,n}) = (Y^{\sigma(1),n}, Y^{\sigma(2),n}, \ldots, Y^{\sigma(n),n}), \quad \text{a.s. },$$

so the processes $\{Y^{i,n}\}_{i=1}^n$ are exchangeable. Regarding the processes $\{Y^i\}_{i\geq 1}$, from (6.5) we may write $Y_t^i = \varphi(\xi^i, (W_s^i)_{0\leq s\leq t})$ for some Borel measurable function φ. But, the pairs (W^i, ξ^i) are independent and equally distributed. Therefore, the Y^i's are independent and equally distributed and thus exchangeable. □

Proposition 6.2.3. *Assume that γ_1 and γ_2 satisfy*

$$\gamma_1^2 + \gamma_2^2 < \frac{1}{16}. \tag{6.8}$$

Then, under Assumption 4 we have

$$\lim_{n\to\infty} \sup_{1\leq i\leq n} \mathbb{E}\left[\sup_{t\in[0,T]} |Y_t^{i,n} - Y_t^i|^2\right] = 0.$$

Proof. For any $t \leq T$, we have

$$\begin{aligned}
|Y_t^{i,n} - Y_t^i| &= \Bigg| \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_t^i} \mathbb{E}\left[h(Y_\tau^{i,n}, \frac{1}{n}\sum_{j=1}^n Y_\tau^{j,n})\mathbb{1}_{\{\tau<T\}} + \xi^i \mathbb{1}_{\{\tau=T\}} \,|\, \mathcal{F}_t^i\right] \\
&\qquad - \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_t^i} \mathbb{E}\left[h(Y_\tau^i, \mathbb{E}[Y_\tau^i])\mathbb{1}_{\{\tau<T\}} + \xi^i \mathbb{1}_{\{\tau=T\}} \,|\, \mathcal{F}_t^i\right] \Bigg| \\
&\leq \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_t^i} \Bigg| \mathbb{E}\left[h(Y_\tau^{i,n}, \frac{1}{n}\sum_{j=1}^n Y_\tau^{j,n})\mathbb{1}_{\{\tau<T\}} + \xi^i \mathbb{1}_{\{\tau=T\}} \,|\, \mathcal{F}_t^i\right] \\
&\qquad - \mathbb{E}\left[h(Y_\tau^i, \mathbb{E}[Y_\tau^i])\mathbb{1}_{\{\tau<T\}} + \xi^i \mathbb{1}_{\{\tau=T\}} \,|\, \mathcal{F}_t^i\right] \Bigg| \\
&\leq \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_t^i} \mathbb{E}\left[\left|h(Y_\tau^{i,n}, \frac{1}{n}\sum_{j=1}^n Y_\tau^{j,n}) - h(Y_\tau^i, \mathbb{E}[Y_\tau^i])\right| \,|\, \mathcal{F}_t^i\right] \\
&\leq \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_t^i} \mathbb{E}\left[\left(\gamma_1|Y_\tau^{i,n} - Y_\tau^i| + \gamma_2|\frac{1}{n}\sum_{j=1}^n Y_\tau^{j,n} - \mathbb{E}[Y_\tau^i]|\right) \,|\, \mathcal{F}_t^i\right] \\
&\leq \mathbb{E}\left[\gamma_1 \sup_{s\in[0,T]} |Y_s^{i,n} - Y_s^i| + \gamma_2 \sup_{s\in[0,T]} |\frac{1}{n}\sum_{j=1}^n Y_s^{j,n} - \mathbb{E}[Y_s^i]| \,|\, \mathcal{F}_t^i\right] \\
&\leq \mathbb{E}\left[G_T^{i,n} + \gamma_2 \sup_{s\in[0,T]} |\frac{1}{n}\sum_{j=1}^n Y_s^j - \mathbb{E}[Y_s^i]| \,|\, \mathcal{F}_t^i\right],
\end{aligned} \tag{6.9}$$

where in passage to the fifth inequality we used Lemma 6.2.4, which is stated below. Moreover,

$$G_T^{i,n} := \gamma_1 \sup_{s\in[0,T]} |Y_s^{i,n} - Y_s^i| + \frac{\gamma_2}{n}\sum_{j=1}^n \sup_{s\in[0,T]} |Y_s^{j,n} - Y_s^j|.$$

By Doob's inequality, we have

$$\mathbb{E}\left[\sup_{t\in[0,T]} |Y_t^{i,n} - Y_t^i|^2\right] \le 4\mathbb{E}\left[\left(G_T^{i,n} + \gamma_2 \sup_{s\in[0,T]} |\frac{1}{n}\sum_{j=1}^n Y_s^j - \mathbb{E}[Y_s^i]|\right)^2\right].$$

Therefore,

$$\begin{aligned}&\mathbb{E}\left[\sup_{t\in[0,T]} |Y_t^{i,n} - Y_t^i|^2\right]\\ &\quad\le 8\gamma_2^2\mathbb{E}\left[\Lambda_n\right] + 16\mathbb{E}\left[\gamma_1^2 \sup_{s\in[0,T]} |Y_s^{i,n} - Y_s^i|^2 + \frac{\gamma_2^2}{n}\sum_{j=1}^n \sup_{s\in[0,T]} |Y_s^{j,n} - Y_s^j|^2\right],\end{aligned} \tag{6.10}$$

where, since the Y^i's are identically distributed, we have

$$\Lambda_n := \sup_{s\in[0,T]} |\frac{1}{n}\sum_{j=1}^n Y_s^j - \mathbb{E}[Y_s^i]|^2 = \sup_{s\in[0,T]} |\frac{1}{n}\sum_{j=1}^n (Y_s^j - \mathbb{E}[Y_s^j])|^2. \tag{6.11}$$

In view of the exchangeability of the processes $\{Y^{i,n}\}_{i=1}^n$ and $\{Y^i\}_{i\ge1}$ (see Proposition 6.2.2), we have

$$\mathbb{E}\left[\frac{1}{n}\sum_{j=1}^n \sup_{s\in[0,T]} |Y_s^{j,n} - Y_s^j|^2\right] = \mathbb{E}\left[\sup_{t\in[0,T]} |Y_t^{i,n} - Y_t^i|^2\right].$$

Thus, from (6.10) we obtain

$$\mathbb{E}\left[\sup_{t\in[0,T]} |Y_t^{i,n} - Y_t^i|^2\right] \le 16(\gamma_1^2+\gamma_2^2)\,\mathbb{E}\left[\sup_{s\in[0,T]} |Y_s^{i,n} - Y_s^i|^2\right] + 32\gamma_2^2\mathbb{E}[\Lambda_n],$$

where, by (6.8), $16(\gamma_1^2+\gamma_2^2) < 1$. So

$$\sup_{1\le i\le n} \mathbb{E}\left[\sup_{t\in[0,T]} |Y_t^{i,n} - Y_t^i|^2\right] \le 32(1-16(\gamma_1^2+\gamma_2^2))^{-1}\mathbb{E}[\Lambda_n].$$

Finally, since the processes $\{Y^j\}_{j\ge1}$ are i.i.d. $C([0,T];\mathbb{R})$-valued random variables with finite second moments (since they are in $\mathcal{S}^2$), by the strong law of large numbers for Banach-valued random variables (see Theorem 4.1.1 in [86]) and dominated convergence we have

$$\lim_{n\to\infty} \mathbb{E}[\Lambda_n] = 0, \tag{6.12}$$

which yields the desired result. □

Lemma 6.2.4. *Let X be an $\mathbb{F}$-adapted continuous process. Then*

(a) $\sup_{t\in[0,T]} X_t = \operatorname{ess\,sup}_{\tau\in\mathcal{T}_0} X_\tau$ a.s.

(b) Assume $X \in \mathcal{S}_c^2$. Then

$$\inf_{\tau\in\mathcal{T}_0} \mathbb{E}[X_\tau] = \inf_{t\in[0,T]} \mathbb{E}[X_t]. \tag{6.13}$$

Proof. (a) Since X is (right)-continuous and adapted, it holds that

$$\sup_{t\in[0,T]} X_t = \sup_{t\in[0,T]\cap\mathbb{Q}} X_t \text{ a.s. },$$

where $\mathbb{Q}$ denotes the set of rational numbers. Therefore, $\sup_{t\in[0,T]} X_t$ is a random variable. By the uniqueness of the essential supremum, we have $\sup_{t\in[0,T]} X_t = \operatorname{ess\,sup}_{t\in[0,T]} X_t$ almost surely. Furthermore, for every $\tau \in \mathcal{T}_0$, $X_\tau \le \sup_{t\in[0,T]} X_t$ almost surely. Hence, $\operatorname{ess\,sup}_{\tau\in\mathcal{T}_0} X_\tau \le \sup_{t\in[0,T]} X_t$ a.s. . The reverse inequality follows from the fact that each $t \in [0,T]$ is in $\mathcal{T}_0$ and so $X_t \le \operatorname{ess\,sup}_{\tau\in\mathcal{T}_0} X_\tau$ a.s. .

(b) Since $X \in \mathcal{S}_c^2$, both $\inf_{t\in[0,T]} \mathbb{E}[X_t]$ and $\inf_{\tau\in\mathcal{T}_0} \mathbb{E}[X_\tau]$ are lower bounded. Furthermore, for every $\tau \in \mathcal{T}_0$, the stopped process $(X_{\tau\wedge t})_{0\le t\le T}$ is also in $\mathcal{S}_c^2$.

For each $t \in [0,T]$, we have

$$\mathbb{E}[X_{\tau\wedge t}] \ge \inf_{0\le u\le \tau\wedge t} \mathbb{E}[X_u] \ge \inf_{v\in[0,T]} \mathbb{E}[X_v], \tag{6.14}$$

and by dominated convergence we obtain

$$\mathbb{E}[X_\tau] = \lim_{t\to T} \mathbb{E}[X_{\tau\wedge t}] \ge \inf_{v\in[0,T]} \mathbb{E}[X_v].$$

Thus, $\inf_{\tau\in\mathcal{T}_0} \mathbb{E}[X_\tau] \ge \inf_{v\in[0,T]} \mathbb{E}[X_v]$.

To prove the reverse inequality, we note that every $t \in [0,T]$ is in $\mathcal{T}_0$. Therefore, $\mathbb{E}[X_t] \ge \inf_{\tau\in\mathcal{T}_0} \mathbb{E}[X_\tau]$, which implies that $\inf_{v\in[0,T]} \mathbb{E}[X_v] \ge \inf_{\tau\in\mathcal{T}_0} \mathbb{E}[X_\tau]$. □

Remark 6.2.5.

- The condition $X \in \mathcal{S}_c^2$ is needed to enable the application of the dominated convergence theorem to show that for any $\tau \in \mathcal{T}_0$, $\mathbb{E}[X_\tau] = \lim_{t\to T} \mathbb{E}[X_{\tau\wedge t}]$. It can be relaxed depending on the problem at hand.
- Equation (6.13) is proved here only for the class of bounded stopping times. In general, it does not hold for the class of stopping times which contains a.s. finite stopping times but are not bounded, such as the first hitting time τ_b of the standard Brownian motion B to a point $b < 0$. Indeed, $\inf_{\tau\ge 0} \mathbb{E}[B_\tau] \le \mathbb{E}[B_{\tau_b}] = b < 0$ but $\inf_{t\ge 0} \mathbb{E}[B_t] = 0$.

6.2.2 Convergence of the optimal stopping times

In Section 6.2.1 we proved that $Y^{i,n}$ converges to Y^i as n goes to infinity in the $\mathcal{S}^2$ norm. This entails the convergence of the values of the OSPs $Y_0^{i,n}$ to Y_0^i, as $n \to \infty$, for every $i \geq 1$. Furthermore, by Corollary 6.1.2, for each $i = 1, \dots, n$, the stopping time $\hat{\tau}^{i,n}$ given by (6.7) is optimal for the OSP $Y_0^{i,n}$, that is

$$\hat{\tau}^{i,n} = \arg\max_{\tau \in \mathcal{T}_0^i} \mathbb{E}\left[h(Y_\tau^{i,n}, \frac{1}{n}\sum_{j=1}^n Y_\tau^{j,n}) \mathbb{1}_{\{\tau < T\}} + \xi^i \mathbb{1}_{\{\tau = T\}} \right]. \tag{6.15}$$

For every $i \geq 1$, let us introduce the stopping time

$$\hat{\tau}^i = \inf\{t \geq 0,\ Y_t^i = h(Y_t^i, \mathbb{E}[Y_t^i])\} \wedge T. \tag{6.16}$$

Due to the time-inconsistency of the problem (6.5), it is not immediate to conclude that $\hat{\tau}^i$ is optimal, i.e. that it coincides with the optimal stopping $\tau^{i,*}$ defined by

$$\tau^{i,*} = \arg\max_{\tau \in \mathcal{T}_0} \mathbb{E}\left[h(Y_\tau^i, \mathbb{E}[Y_\tau^i]) \mathbb{1}_{\{\tau < T\}} + \xi^i \mathbb{1}_{\{\tau = T\}} \right]. \tag{6.17}$$

The main result of this section is to show that (6.17) actually holds. In particular, since $W^1 = W$ and $\xi^1 = \xi$, we have $Y^1 = Y$ i.e. Y^1 is the value-process Y given by (6.1) and $\tau^{1,*}$ is the associated optimal stopping time given by (6.3) i.e. $\tau^* := \tau^{1,*}$.

Theorem 6.2.6. *Let Assumption 4 hold and assume that γ_1 and γ_2 satisfy* (6.8)*. Then for any $i \geq 1$ the stopping time $\hat{\tau}^i$ defined by* (6.16) *is optimal for the OSP* (6.17)*.*

To prove this statement, we rely on the next Proposition 6.2.8 about the convergence of optimal stopping times and its Corollary 6.2.9. We also need the following

Lemma 6.2.7. *Let us set, for every $t \in [0, T]$,*

$$Z_t^{i,n} := Y_t^{i,n} - \mathbb{E}\left[h(Y_t^{i,n}, \frac{1}{n}\sum_{j=1}^n Y_t^{j,n}) \,|\, \mathcal{F}_t^i \right], \quad Z_t^i = Y_t^i - h(Y_t^i, \mathbb{E}\left[Y_t^i\right]). \tag{6.18}$$

Then, we have

$$\lim_{n\to\infty} \sup_{1 \leq i \leq n} \mathbb{E}\left[\sup_{t \in [0,T]} |Z_t^{i,n} - Z_t^i|^2 \right] = 0. \tag{6.19}$$

Proof. By Assumption 4 (ii), for every $1 \leq i \leq n$, we have

$$\begin{aligned}\sup_{t\in[0,T]} |Z_t^{i,n} - Z_t^i| &\leq \sup_{t\in[0,T]} |Y_t^{i,n} - Y_t^i| + \sup_{t\in[0,T]} |\mathbb{E}[h(Y_t^{i,n}, \frac{1}{n}\sum_{j=1}^n Y_t^{j,n}) \,|\, \mathcal{F}_t^i] - h(Y_t^i, \mathbb{E}[Y_t^i])| \\ &\leq (1+\gamma_1) \sup_{t\in[0,T]} |Y_t^{i,n} - Y_t^i| + \gamma_2 \sup_{t\in[0,T]} \mathbb{E}[\sup_{s\in[0,T]} |\frac{1}{n}\sum_{j=1}^n Y_s^{j,n} - \mathbb{E}[Y_s^i]| \,|\, \mathcal{F}_t^i].\end{aligned}$$

Using Doob's inequality, we obtain

$$\mathbb{E}\left[\left(\sup_{t\in[0,T]}\mathbb{E}\left[\sup_{s\in[0,T]}|\frac{1}{n}\sum_{j=1}^{n}Y_s^{j,n}-\mathbb{E}\left[Y_s^i\right]|\,|\,\mathcal{F}_t^i\right]\right)^2\right]$$
$$\leq 4\,\mathbb{E}\left[\left(\sup_{t\in[0,T]}|\frac{1}{n}\sum_{j=1}^{n}Y_t^{j,n}-\mathbb{E}[Y_t^i]|\right)^2\right]$$
$$\leq 8\,\mathbb{E}\left[\left(\frac{1}{n}\sum_{j=1}^{n}\sup_{t\in[0,T]}|Y_t^{j,n}-Y_t^j|\right)^2\right]+8\,\mathbb{E}\left[\left(\sup_{t\in[0,T]}|\frac{1}{n}\sum_{j=1}^{n}|Y_t^j-\mathbb{E}[Y_t^j]|\right)^2\right]$$
$$\leq 8\left[\sup_{t\in[0,T]}|Y_t^{i,n}-Y_t^i|^2\right]+8\mathbb{E}[\Lambda_n],$$

where the first term of the last inequality follows from the Cauchy-Schwarz inequality and the exchangeability of the processes $\{Y^{i,n}\}_{i=1}^n$ and $\{Y^i\}_{i\geq 1}$ (by Proposition 6.2.2) and Λ_n is given by (6.11).

Therefore, we have

$$\sup_{1\leq i\leq n}\mathbb{E}\left[\sup_{t\in[0,T]}|Z_t^{i,n}-Z_t^i|^2\right]$$
$$\leq 2\left((1+\gamma_1)^2+8\gamma_2^2\right)\sup_{1\leq i\leq n}\mathbb{E}\left[\sup_{t\in[0,T]}|Y_t^{i,n}-Y_t^i|^2\right]+16\gamma_2^2\mathbb{E}[\Lambda_n].$$

Thus, thanks to Proposition 6.2.3 and (6.12), it holds that

$$\lim_{n\to\infty}\sup_{1\leq i\leq n}\mathbb{E}\left[\sup_{t\in[0,T]}|Z_t^{i,n}-Z_t^i|^2\right]=0, \tag{6.20}$$

□

Proposition 6.2.8. *For every $i\geq 1$, let $\{\hat{\tau}^{i,n}\}_{n\geq 1}$ be the sequence of optimal stopping times defined by* (6.7) *and let $\hat{\tau}^i$ be defined by* (6.16)*. Let Assumption 4 and* (6.8) *hold. Then, for every $\epsilon>0$,*

$$\lim_{n\to\infty}\sup_{1\leq i\leq n}\mathbb{P}\left(|\hat{\tau}^{i,n}-\hat{\tau}^i|>\epsilon\right)=0. \tag{6.21}$$

In particular, for every fixed $i\geq 1$, $\hat{\tau}^{i,n}$ converges to $\hat{\tau}^i$ in probability, as n goes to infinity.

Proof. Recall the notation

$$Z_t^{i,n}:=Y_t^{i,n}-\mathbb{E}\left[h(Y_t^{i,n},\frac{1}{n}\sum_{j=1}^{n}Y_t^{j,n})\,|\,\mathcal{F}_t^i\right],\quad Z_t^i=Y_t^i-h(Y_t^i,\mathbb{E}\left[Y_t^i\right]),\quad t\in[0,T].$$

We notice that, for every $t \in [0,T]$, $Z_t^{i,n} \geq 0$ a.s. and that $\hat{\tau}^{i,n} = \inf\{t \geq 0,\ Z_t^{i,n} = 0\} \wedge T$. The same holds for Z^i and $\hat{\tau}^i$.

For any $\epsilon > 0$, we have

$$\mathbb{P}\left(|\hat{\tau}^{i,n} - \hat{\tau}^i| > \epsilon\right) = \mathbb{P}\left(\hat{\tau}^{i,n} - \hat{\tau}^i > \epsilon\right) + \mathbb{P}\left(\hat{\tau}^i - \hat{\tau}^{i,n} > \epsilon\right). \tag{6.22}$$

We first show that $\mathbb{P}\left(\hat{\tau}^{i,n} - \hat{\tau} > \epsilon\right) \to 0$ as $n \to \infty$. The event

$$\left\{\hat{\tau}^{i,n} - \hat{\tau}^i > \epsilon\right\}$$

means that $Z^{i,n}$ attains 0 at a time which is larger than the time $\hat{\tau}^i$ at which Z^i attains the same level 0 with at least $\epsilon > 0$. In other words,

$$\mathbb{P}\left(\hat{\tau}^{i,n} - \hat{\tau}^i > \epsilon\right) = \mathbb{P}\left(\inf_{0 \leq t \leq \hat{\tau}^i + \epsilon} Z_t^{i,n} > 0, Z_{\hat{\tau}^i}^i = 0\right). \tag{6.23}$$

Given $\delta > 0$, let us consider the stopping time

$$\sigma^{i,n} = \inf\{t \geq \hat{\tau}^i,\ Z_t^{i,n} \leq \delta\} \wedge T.$$

We have $\inf_{\hat{\tau}^i \leq t < (\hat{\tau}^i + \epsilon) \wedge \sigma^{i,n}} Z_t^{i,n} > \delta$. So the following events coincide:

$$\left\{\inf_{\hat{\tau}^i \leq t < (\hat{\tau}^i + \epsilon) \wedge \sigma^{i,n}} Z_t^{i,n} > \delta\right\} = \left\{\inf_{\hat{\tau}^i \leq t < (\hat{\tau}^i + \epsilon) \wedge \sigma^{i,n}} Z_t^{i,n} > 0\right\}.$$

Furthermore, $\inf_{\hat{\tau}^i \leq t \leq \hat{\tau}^i + \epsilon} Z_t^{i,n} \leq \inf_{\hat{\tau}^i \leq t < (\hat{\tau}^i + \epsilon) \wedge \sigma^{i,n}} Z_t^{i,n}$, and so it follows that

$$\mathbb{P}\left(\inf_{\hat{\tau}^i \leq t \leq \hat{\tau}^i + \epsilon} Z_t^{i,n} > 0,\ Z_{\hat{\tau}^i}^i = 0\right) \leq \mathbb{P}\left(\inf_{\hat{\tau}^i \leq t < (\hat{\tau}^i + \epsilon) \wedge \sigma^{i,n}} Z_t^{i,n} > \delta,\ Z_{\hat{\tau}^i}^i = 0\right).$$

We have

$$\begin{aligned}
\sup_{1 \leq i \leq n} \mathbb{P}\left(\hat{\tau}^{i,n} - \hat{\tau}^i > \epsilon\right) &= \sup_{1 \leq i \leq n} \mathbb{P}\left(\inf_{0 \leq t \leq \hat{\tau}^i + \epsilon} Z_t^{i,n} > 0,\ Z_{\hat{\tau}^i}^i = 0\right) \\
&\leq \sup_{1 \leq i \leq n} \mathbb{P}\left(\inf_{\hat{\tau}^i \leq t < (\hat{\tau}^i + \epsilon) \wedge \sigma^{i,n}} Z_t^{i,n} > \delta,\ Z_{\hat{\tau}^i}^i = 0\right) \\
&\leq \sup_{1 \leq i \leq n} \mathbb{P}\left(\inf_{\hat{\tau}^i \leq t < (\hat{\tau}^i + \epsilon) \wedge \sigma^{i,n}} Z_t^{i,n} - Z_{\hat{\tau}^i}^i > \delta\right) \\
&\leq \sup_{1 \leq i \leq n} \mathbb{P}\left(\sup_{t \in [0,T]} |Z_t^{i,n} - Z_t^i| > \delta\right),
\end{aligned}$$

which entails that $\sup_{1 \leq i \leq n} \mathbb{P}\left(\hat{\tau}^{i,n} - \hat{\tau}^i > \epsilon\right) \to 0$ as $n \to \infty$, in view of (6.20).

Let us now consider the second term on the right hand side of (6.22). Similarly to the previous step, given $\rho > 0$, let us consider the stopping time

$$\alpha^{i,n} = \inf\{t \geq \hat{\tau}^{i,n},\ Z_t^i \leq \rho\} \wedge T.$$

We have $\inf_{\hat{\tau}^{i,n}\le t<(\hat{\tau}^{i,n}+\epsilon)\wedge\alpha^{i,n}} Z_t^{i,n} > \delta$. This in turn yields

$$\left\{\inf_{\hat{\tau}^{i,n}\le t<(\hat{\tau}^{i,n}+\epsilon)\wedge\alpha^{i,n}} Z_t^i > \rho\right\} = \left\{\inf_{\hat{\tau}^{i,n}\le t<(\hat{\tau}^{i,n}+\epsilon)\wedge\alpha^{i,n}} Z_t^i > 0\right\}.$$

Since, $\inf_{\hat{\tau}^{i,n}\le t\le\hat{\tau}^{i,n}+\epsilon} Z_t^i \le \inf_{\hat{\tau}^{i,n}\le t<(\hat{\tau}^{i,n}+\epsilon)\wedge\alpha^{i,n}} Z_t^i$, we have

$$\mathbb{P}\left(\inf_{\hat{\tau}^{i,n}\le t\le\hat{\tau}^{i,n}+\epsilon} Z_t^i > 0,\ Z_{\hat{\tau}^{i,n}}^{i,n}=0\right) \le \mathbb{P}\left(\inf_{\hat{\tau}^{i,n}\le t<(\hat{\tau}^{i,n}+\epsilon)\wedge\alpha^{i,n}} Z_t^i > \rho,\ Z_{\hat{\tau}^i}^i = 0\right).$$

Therefore,

$$\begin{aligned}\sup_{1\le i\le n}\mathbb{P}\left(\hat{\tau}^i-\hat{\tau}^{i,n}>\epsilon\right) &\le \sup_{1\le i\le n}\mathbb{P}\left(\inf_{\hat{\tau}^{i,n}\le t<(\hat{\tau}^{i,n}+\epsilon)\wedge\alpha^{i,n}} Z_t^i > \rho,\ Z_{\hat{\tau}^{i,n}}^{i,n}=0\right)\\ &\le \sup_{1\le i\le n}\mathbb{P}\left(\inf_{\hat{\tau}^{i,n}\le t<(\hat{\tau}^{i,n}+\epsilon)\wedge\alpha^{i,n}} Z_t^i - Z_{\hat{\tau}^{i,n}}^{i,n} > \rho\right)\\ &\le \sup_{1\le i\le n}\mathbb{P}\left(\sup_{t\in[0,T]}|Z_t^{i,n}-Z_t^i|>\rho\right).\end{aligned}$$

which entails that $\sup_{1\le i\le n}\mathbb{P}\left(\hat{\tau}^i-\hat{\tau}^{i,n}>\epsilon\right)\to 0$ as $n\to\infty$, in view again of (6.20).

□

Corollary 6.2.9. *We have, for every $i\ge 1$,*

$$\lim_{n\to\infty}\mathbb{E}[Y_{\hat{\tau}^{i,n}}^i]=\mathbb{E}[Y_{\hat{\tau}^i}^i]. \tag{6.24}$$

Moreover, up to a subsequence, it holds that

$$\begin{aligned}\lim_{n\to\infty}\mathbb{E}\left[h(Y_{\hat{\tau}^{i,n}}^{i,n},\frac{1}{n}\sum_{j=1}^n Y_{\hat{\tau}^{j,n}}^{j,n})\mathbb{1}_{\{\hat{\tau}^{i,n}<T\}}+\xi^i\mathbb{1}_{\{\hat{\tau}^{i,n}=T\}}\right]&\\ =\mathbb{E}\left[h(Y_{\hat{\tau}^i}^i,\mathbb{E}[Y_{\hat{\tau}^i}^i])\mathbb{1}_{\{\hat{\tau}^i<T\}}+\xi^i\mathbb{1}_{\{\hat{\tau}^i=T\}}\right].&\end{aligned} \tag{6.25}$$

Proof. We derive (6.24) by contradiction. Assume that $\hat{\tau}^{i,n}$ converges in probability to $\hat{\tau}^i$ with $|\mathbb{E}[Y_{\hat{\tau}^{i,n}}^i]-\mathbb{E}[Y_{\hat{\tau}^i}^i]|\ge\varepsilon>0$ for all n. But, then we can extract a subsequence $\hat{\tau}^{i,n_k}$ which converges to $\hat{\tau}^i$ a.s. Since the continuous process Y^i is in $\mathcal{S}^2$, by dominated convergence, we arrive at a contradiction.

To derive (6.25), we note that since the process Y^i is continuous and $\hat{\tau}^{i,n},\hat{\tau}^i$ are $\mathbb{F}^i$-stopping times, it holds that the sequence $(Y^{i,n},\hat{\tau}^{i,n})$ converges in probability to $(Y^i,\hat{\tau}^i)$. Therefore, in view of [3], Corollary 16.23, $(\hat{\tau}^{i,n},Y_{\hat{\tau}^{i,n}}^{i,n})$ converges in distribution to $(\hat{\tau}^i,Y_{\hat{\tau}^i}^i)$. For each $i\ge 1$, let $\{\hat{\tau}^{i,n_k}\}_{k\ge 1}$ be a subsequence of the sequence of stopping times $\{\hat{\tau}^{i,n}\}_{n\ge 1}$, which converges a.s. to $\hat{\tau}^i$. We claim that for every $i\ge 1$, $\frac{1}{n_k}\sum_{j=1}^{n_k}Y_{\hat{\tau}^{j,n_k}}^{j,n_k}\overset{L^1}{\to}Y_{\hat{\tau}^i}^i$ as $k\to\infty$. Indeed, noting that (Y^i,τ^i), $i=1,2,\dots$ are i.i.d., we have

$$\mathbb{E}\left[\left|\frac{1}{n_k}\sum_{j=1}^{n_k} Y^{j,n_k}_{\hat\tau^{j,n_k}} - \mathbb{E}[Y^i_{\hat\tau^i}]\right|\right]$$
$$\le \mathbb{E}\left[\left|\frac{1}{n_k}\sum_{j=1}^{n_k}(Y^{j,n_k}_{\hat\tau^{j,n_k}} - Y^j_{\hat\tau^{j,n_k}})\right|\right] + \frac{1}{n_k}\sum_{j=1}^{n_k}\mathbb{E}\left[\left|Y^j_{\hat\tau^{j,n_k}} - Y^j_{\hat\tau^j}\right|\right]$$
$$+ \mathbb{E}\left[\left|\frac{1}{n_k}\sum_{j=1}^{n_k}(Y^j_{\hat\tau^j} - \mathbb{E}[Y^j_{\hat\tau^j}])\right|\right].$$

We have

$$\mathbb{E}\left[\left|\frac{1}{n_k}\sum_{j=1}^{n_k}(Y^{j,n_k}_{\hat\tau^{j,n_k}} - Y^j_{\hat\tau^{j,n_k}})\right|\right] \le \frac{1}{n_k}\sum_{j=1}^{n_k}\mathbb{E}\left[\sup_{t\in[0,T]}|Y^{j,n_k}_t - Y^j_t|\right] \to 0, \quad k\to\infty,$$

by the Cesaro Mean Lemma. Moreover, in view of the dominated convergence theorem, for every $j \ge 1$, $\mathbb{E}\left[\left|Y^j_{\hat\tau^{j,n_k}} - Y^j_{\hat\tau^j}\right|\right] \to 0$ as $k\to\infty$. Thus, again by the Cesaro Mean Lemma, we have $\frac{1}{n_k}\sum_{j=1}^{n_k}\mathbb{E}\left[\left|Y^j_{\hat\tau^{j,n_k}} - Y^j_{\hat\tau^j}\right|\right] \to 0$ as $k\to\infty$.

Since (Y^i,τ^i), $i=1,2,\ldots,$ are i.i.d., the random variables $Y^i_{\hat\tau^i}$, $i=1,2,\ldots$ are i.i.d. By the strong law of large numbers and dominated convergence we have

$$\mathbb{E}\left[\left|\frac{1}{n_k}\sum_{j=1}^{n_k}(Y^j_{\hat\tau^j} - \mathbb{E}[Y^j_{\hat\tau^j}])\right|\right] \to 0, \quad k\to\infty.$$

Therefore, as $k\to\infty$, $h(Y^{i,n_k}_{\hat\tau^{i.n_k}}, \frac{1}{n_k}\sum_{j=1}^{n} Y^{j,n_k}_{\hat\tau^{j,n_k}})\mathbb{1}_{\{\hat\tau^{i,n_k}<T\}} + \xi^i\mathbb{1}_{\{\hat\tau^{i,n_k}=T\}}$ converges almost surely to $h(Y^i_{\hat\tau^i}, \mathbb{E}[Y^i_{\hat\tau^i}])\mathbb{1}_{\{\hat\tau^i<T\}} + \xi^i\mathbb{1}_{\{\hat\tau^i=T\}}$. The claim (6.25) follows by dominated convergence. □

Proof of Theorem 6.2.6. Let $\{\hat\tau^{i,n_k}\}_{k\ge1}$ be a subsequence of the sequence of stopping times $\{\hat\tau^{i,n}\}_{n\ge1}$, which converges a.s. to $\hat\tau^i$. In view of (6.25) and the optimality of $\{\hat\tau^{i,n_k}\}_{k\ge1}$, we have

$$Y^i_0 = \lim_{k\to\infty} Y^{i,n_k}_0 = \lim_{k\to\infty}\mathbb{E}\left[h(Y^{i,n_k}_{\hat\tau^{i.n_k}}, \frac{1}{n_k}\sum_{j=1}^{n_k} Y^{j,n_k}_{\hat\tau^{j,n_k}})\mathbb{1}_{\{\hat\tau^{i,n_k}<T\}} + \xi^i\mathbb{1}_{\{\hat\tau^{i,n_k}=T\}}\right]$$
$$= \mathbb{E}\left[h(Y^i_{\hat\tau^i}, \mathbb{E}[Y^i_{\hat\tau^i}])\mathbb{1}_{\{\hat\tau^i<T\}} + \xi^i\mathbb{1}_{\{\hat\tau^i=T\}}\right].$$

□

6.2.3 Extension to the general mean field case

The aim of this section is to extend the results we proved in Sections 6.1.1 and 6.2 to the more general case where the performance function depends on the law of the stopped process and not only on its mean. More precisely, we want to solve the OSP

$$Y_0 = \sup_{\tau\in\mathcal{T}_0}\mathbb{E}\left[h(Y_\tau, \mathbb{P}_{Y_\tau})\mathbb{1}_{\{\tau<T\}} + \xi\mathbb{1}_{\{\tau=T\}}\right], \tag{6.26}$$

where $\mathbb{P}_\zeta$ denotes the law of the random variable ζ under $\mathbb{P}$. To this end, as in Section 6.1, we introduce two families of value-processes:

$$Y_t^{i,n} = \operatorname*{ess\,sup}_{\tau \in \mathcal{T}_t^i} \mathbb{E}\left[h(Y_\tau^{i,n}, \frac{1}{n}\sum_{j=1}^n \delta_{Y_\tau^{j,n}}) \mathbb{1}_{\{\tau < T\}} + \xi^i \mathbb{1}_{\{\tau = T\}} \,|\, \mathcal{F}_t^i \right], \quad i = 1, 2, \ldots, n, \tag{6.27}$$

where δ_x is the Dirac measure centered at $x \in \mathbb{R}$, and

$$Y_t^i = \operatorname*{ess\,sup}_{\tau \in \mathcal{T}_t^i} \mathbb{E}\left[h(Y_\tau^i, \mathbb{P}_{Y_\tau}) \mathbb{1}_{\{\tau < T\}} + \xi^i \mathbb{1}_{\{\tau = T\}} \,|\, \mathcal{F}_t^i \right], \quad i = 1, 2, \ldots. \tag{6.28}$$

Since now h is also a function of probability measures, we have to modify Assumption 4(ii) but keep 4(i).

Assumption 5. *The coefficients h and $\{\xi^i\}_{i \geq 1}$ satisfy*

(i) for each $i \geq 1$, $\xi^i \in L^2(\mathcal{F}_T^i)$. Moreover, the ξ^i's are independent copies of ξ, with $\xi^1 = \xi$;

(ii) $h \colon \Omega \times \mathbb{R} \times \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}$

(ii-a) is Lipschitz with respect to (y, μ), that is there exist two positive constants γ_1 and γ_2 such that

$$|h(\omega, y_1, \mu_1) - h(\omega, y_2, \mu_2)| \leq \gamma_1 |y_1 - y_2| + \gamma_2 \mathcal{W}_2(\mu_1, \mu_2),$$

for any $\omega \in \Omega$, any $y_1, y_2 \in \mathbb{R}$ and any $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R})$.

(ii-b) $h(\cdot, 0, \delta_0) \in L^2(\mathbb{P})$.

With Assumption 5 instead of 4, we can recover Theorem 6.1.1, Corollary 6.1.2 and Theorem 6.2.6, as well as all the intermediate results, for this more general framework by using the following inequalities satisfied by the 2-Wasserstein distance:

1. Given two random variables $X, Y \in L^2(\mathbb{P})$,

$$\mathcal{W}_2^2(\mathbb{P}_X, \mathbb{P}_Y) \leq \mathbb{E}[|X - Y|^2],$$

and in particular $\mathcal{W}_2(\mathbb{P}_X, \delta_0) \leq \mathbb{E}[|X|^2]^{\frac{1}{2}}$;

2. for any $x, y \in \mathbb{R}^n$,

$$\mathcal{W}_2^2\left(\frac{1}{n}\sum_{i=1}^n \delta_{x^i}, \frac{1}{n}\sum_{i=1}^n \delta_{y^i} \right) \leq \frac{1}{n}\sum_{i=1}^n |x^i - y^i|^2.$$

In particular, $\mathcal{W}_2^2\left(\frac{1}{n}\sum_{i=1}^n \delta_{x^i}, \delta_0\right) \leq \frac{1}{n}\sum_{i=1}^n |x^i|^2$.

This inequality allows to show that the process $t \mapsto h(Y_t^{i,n}, \frac{1}{n}\sum_{j=1}^n \delta_{Y_t^{j,n}})$ is continuous a.s., a condition needed to derive a similar result as Corollary 6.1.2.

For the convergence results we also need the following law of large numbers (see for instance Lemma 4.1 in [43] for a proof):

Lemma 6.2.10. *Let $Y^1, Y^2, \ldots, Y^n$ be solutions to the system* (6.28) *(notice that they are independent and equally distributed with common probability law Q). Then,*

$$\lim_{n\to\infty} \mathbb{E}\left[\widetilde{\mathcal{W}}_2^2\left(\frac{1}{n}\sum_{i=1}^n \delta_{Y^i}, Q\right)\right] = 0,$$

where $\widetilde{\mathcal{W}}_2$ is the Wasserstein metric on the space of probability measure on the space $C([0,T];\mathbb{R})$ endowed with the supremum norm.

6.3 Optimal stopping of mean field SDEs

Let us consider the following mean field extension of the standard optimal stopping problem of one-dimensional a diffusion process X. Find a stopping time τ^* such that

$$\tau^* = \arg\max_{\tau\in\mathcal{T}_0} \mathbb{E}\left[h(X_\tau, \mathbb{P}_{X_\tau})1\!\!1_{\{\tau<T\}} + \xi 1\!\!1_{\{\tau=T\}}\right], \tag{6.29}$$

where X is a diffusion process of mean field type

$$X_t = X_0 + \int_0^t b(s, X_s, \mathbb{P}_{X_s})\,\mathrm{d}s + \int_0^t \sigma(s, X_s, \mathbb{P}_{X_s})\,\mathrm{d}W_s, \quad t\in[0,T], \tag{6.30}$$

where X_0 is square-integrable and independent of W. Here, $\{\mathcal{F}_t\}_{t\in[0,T]}$ is the $\mathbb{P}$-completion of $\sigma(X_0, W_s, s\le t)$.

The OSP associated with the MF-SDE (6.30) is

$$Y_0 = \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_0} \mathbb{E}\left[h(X_\tau, \mathbb{P}_{X_\tau})1\!\!1_{\{\tau<T\}} + \xi 1\!\!1_{\{\tau=T\}}\right]. \tag{6.31}$$

The particle system to use to solve this OSP is simply the i.i.d. processes $\{X^i\}_{i\ge1}$ which solve

$$X_t^i = X_0^i + \int_0^t b(s, X_s^i, \mathbb{P}_{X_s^i})\,\mathrm{d}s + \int_0^t \sigma(s, X_s^i, \mathbb{P}_{X_s^i})\,\mathrm{d}W_s^i, \quad t\in[0,T], \tag{6.32}$$

and vector $(X^{1,n}, \ldots, X^{n,n})$ of n weakly interacting diffusions defined by

$$X_t^{i,n} = X_0^i + \int_0^t b(s, X_s^{i,n}, \frac{1}{n}\sum_{j=1}^n \delta_{X_s^{j,n}})\,\mathrm{d}s + \int_0^t \sigma(s, X_s^{i,n}, \frac{1}{n}\sum_{j=1}^n \delta_{X_s^{j,n}})\,\mathrm{d}W_s^i, \quad t\in[0,T], \tag{6.33}$$

where $(X_0^1, W^1) = (X_0, W)$, which implies that $X^1 = X$, and (X_0^i, W^i) are independent and equally distributed.

To the system (6.32) we associate the OSP

$$Y_0^i = \operatorname*{ess\,sup}_{\tau\in\mathcal{T}_0^i} \mathbb{E}\left[h(X_\tau^i, \mathbb{P}_{X_\tau^i})1\!\!1_{\{\tau<T\}} + \xi^i 1\!\!1_{\{\tau=T\}}\right], \quad i\ge1, \tag{6.34}$$

and the associated family of optimal stopping times

$$\hat{\tau}^i = \inf\left\{t\ge0,\ Y_t^{i,n} = h(X_t^i, \mathbb{P}_{X_t^i})\right\}\wedge T. \tag{6.35}$$

Moreover, to the system (6.33) we associate the family of OSPs

$$Y_0^{i,n} = \operatorname*{ess\,sup}_{\tau \in \mathcal{T}_0^i} \mathbb{E}\left[h(X_\tau^{i,n}, \frac{1}{n}\sum_{j=1}^n \delta_{X_\tau^{j,n}}) \mathbb{1}_{\{\tau<T\}} + \xi^i \mathbb{1}_{\{\tau=T\}} \right], \quad i = 1, 2, \ldots, n, \tag{6.36}$$

and the associated family of optimal stopping times

$$\hat{\tau}^{i,n} = \inf\left\{ t \geq 0,\ Y_t^{i,n} = \mathbb{E}[h(X_t^{i,n}, \frac{1}{n}\sum_{j=1}^n \delta_{X_t^{j,n}}) \,|\, \mathcal{F}_t^i] \right\} \wedge T. \tag{6.37}$$

Provided that b and σ satisfy conditions which yield (see for instance Theorem 3 in [65]):

Assumption 6.

(1) Each of the X^i's and $X^{i,n}$'s is in $\mathcal{S}_c^2$,

(2) $\displaystyle \lim_{n\to\infty} \sup_{1\leq i\leq n} \mathbb{E}\left[\sup_{t\in[0,T]} |X_t^{i,n} - X_t^i|^2 \right] = 0, \quad i \geq 1,$

based on the above results, we obtain the following

Theorem 6.3.1. *Under Assumptions 5 and 6, the hitting time*

$$\tau^* = \inf\{s \geq 0;\ Y_s = h(X_s, \mathbb{P}_{X_s})\} \wedge T$$

satisfies

$$\tau^* = \arg\max_{\tau\in\mathcal{T}_0} \mathbb{E}\left[h(X_\tau, \mathbb{P}_{X_\tau}) \mathbb{1}_{\{\tau<T\}} + \xi \mathbb{1}_{\{\tau=T\}} \right].$$

6.3.1 Optimal stopping of the variance of a mean field diffusion

For $h(\omega, x, \mu) := (x - \int_{\mathbb{R}^d} y\mu(dy))^2$ and $\xi \geq 0$, we obtain an OSP of the variance:

$$Y_0 = \sup_{\tau\in\mathcal{T}_0} \mathbb{E}\left[(X_\tau - \mathbb{E}[X_\tau])^2 \mathbb{1}_{\{\tau<T\}} + \xi \mathbb{1}_{\{\tau=T\}} \right]. \tag{6.38}$$

Since h is not Lipschitz continuous we cannot directly apply the above results to claim that the hitting time

$$\tau^* = \inf\{s \geq 0;\ Y_s = (X_s - \mathbb{E}[X_s])^2\} \wedge T \tag{6.39}$$

satisfies

$$\tau^* = \arg\max_{\tau\in\mathcal{T}_0} \mathbb{E}\left[(X_\tau - \mathbb{E}[X_\tau])^2 \mathbb{1}_{\{\tau<T\}} + \xi \mathbb{1}_{\{\tau=T\}} \right]. \tag{6.40}$$

We note that when X is one of the Markov processes considered in [88], the stopping region defined by τ^* is the same as the one given in Theorems 3.2, 4.1 and 5.1 in [88].

Below, we provide a proof that the hitting time τ^* defined by (6.39) satisfies (6.40). To this end, using the notation above, we need to show similar results as Proposition 6.2.8 and Corollary 6.2.9 which follow provided that

$$\lim_{n\to\infty}\mathbb{E}\left[\sup_{t\in[0,T]}|Y_t^{1,n}-Y_t|\right]=0,\qquad \lim_{n\to\infty}\mathbb{E}\left[\sup_{t\in[0,T]}|Z_t^{1,n}-Z_t^1|\right]=0, \tag{6.41}$$

where $Z^{1,n}$ and Z^1 are defined as in (6.18) and

$$Y_t^{1,n}=\operatorname*{ess\,sup}_{\tau\in\mathcal{T}_t^1}\mathbb{E}\left[(X_\tau^{1,n}-\frac{1}{n}\sum_{j=1}^n X_\tau^{j,n})^2\,|\,\mathcal{F}_t\right],\quad Y_t=\operatorname*{ess\,sup}_{\tau\in\mathcal{T}_t}\mathbb{E}\left[(X_\tau-\mathbb{E}[X_\tau])^2\,|\,\mathcal{F}_t\right]. \tag{6.42}$$

In view of the proofs of Propositions 6.2.3 and 6.2.8, the limits (6.41) hold provided that

$$\lim_{n\to\infty}\mathbb{E}\left[\sup_{t\in[0,T]}|\mathbb{E}[(X_t^{1,n}-\frac{1}{n}\sum_{j=1}^n X_t^{j,n})^2\,|\,\mathcal{F}_t]-(X_t-\mathbb{E}[X_t])^2|\right]=0. \tag{6.43}$$

Let us show (6.43). We have

$$\begin{aligned}
&\sup_{t\in[0,T]}|\mathbb{E}[(X_t^{1,n}-\frac{1}{n}\sum_{j=1}^n X_t^{j,n})^2\,|\,\mathcal{F}_t]-(X_t-\mathbb{E}[X_t])^2|\\
&=\sup_{t\in[0,T]}|\mathbb{E}[((X_t^{1,n}-\frac{1}{n}\sum_{j=1}^n X_t^{j,n})-(X_t-\mathbb{E}[X_t]))((X_t^{1,n}-\frac{1}{n}\sum_{j=1}^n X_t^{j,n})+(X_t-\mathbb{E}[X_t]))|\,\mathcal{F}_t]|\\
&\le\sup_{t\in[0,T]}\mathbb{E}[\sup_{s\in[0,T]}|(X_s^{1,n}-\frac{1}{n}\sum_{j=1}^n X_s^{j,n})-(X_s-\mathbb{E}[X_s])(X_s^{1,n}-\frac{1}{n}\sum_{j=1}^n X_s^{j,n})+(X_s-\mathbb{E}[X_s])||\,\mathcal{F}_t]\\
&\qquad\le\sup_{t\in[0,T]}\left(\mathbb{E}\left[\sup_{s\in[0,T]}|(X_s^{1,n}-\frac{1}{n}\sum_{j=1}^n X_s^{j,n})-(X_s-\mathbb{E}[X_s])|^2|\,\mathcal{F}_t\right]\right)^{\frac{1}{2}}\\
&\qquad\qquad\sup_{t\in[0,T]}\left(\mathbb{E}\left[\sup_{s\in[0,T]}|(X_s^{1,n}-\frac{1}{n}\sum_{j=1}^n X_s^{j,n})+(X_s-\mathbb{E}[X_s])|^2|\,\mathcal{F}_t\right]\right)^{\frac{1}{2}},
\end{aligned}$$

where in the last inequality we used the Cauchy-Schwarz inequality for the conditional expectation. Then, again by Doob's inequalities, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\in[0,T]} |\mathbb{E}[(X_t^{1,n} - \frac{1}{n}\sum_{j=1}^n X_t^{j,n})^2 \,|\, \mathcal{F}_t] - (X_t - \mathbb{E}[X_t])^2|\right] \\
&\qquad \le 4\left(\mathbb{E}\left[\sup_{t\in[0,T]} |(X_t^{1,n} - \frac{1}{n}\sum_{j=1}^n X_t^{j,n}) - (X_t - \mathbb{E}[X_t])|^2\right]\right)^{\frac{1}{2}} \\
&\qquad\quad \left(\mathbb{E}\left[\sup_{t\in[0,T]} |(X_t^{1,n} - \frac{1}{n}\sum_{j=1}^n X_t^{j,n}) + (X_t - \mathbb{E}[X_t])|^2\right]\right)^{\frac{1}{2}} \\
&\qquad \le C\left(\mathbb{E}\left[\sup_{t\in[0,T]} |(X_t^{1,n} - \frac{1}{n}\sum_{j=1}^n X_t^{j,n}) - (X_t - \mathbb{E}[X_t])|^2\right]\right)^{\frac{1}{2}},
\end{aligned}$$

where in the last inequality we have used the fact that the term

$$\mathbb{E}\left[\sup_{t\in[0,T]} |(X_t^{1,n} - \frac{1}{n}\sum_{j=1}^n X_t^{j,n}) + (X_t - \mathbb{E}[X_t])|^2\right]$$

is bounded by a constant C which only depends on the $\mathcal{S}^2$-norm of X, due to Assumption 6 (ii) and the exchangeability of the sequence $\{X^{j,n}\}_{j=1}^n$. Furthermore, we have

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\in[0,T]} |(X_t^{1,n} - \frac{1}{n}\sum_{j=1}^n X_t^{j,n}) - (X_t - \mathbb{E}[X_t])|^2\right] \\
&\qquad \le 2\,\mathbb{E}\left[\sup_{t\in[0,T]} |X_t^{1,n} - X_t|^2\right] + 2\,\mathbb{E}\left[\sup_{t\in[0,T]} |\frac{1}{n}\sum_{j=1}^n X_t^{j,n} - \mathbb{E}[X_t])|^2\right] \\
&\qquad \le 2\,\mathbb{E}\left[\sup_{t\in[0,T]} |X_t^{1,n} - X_t|^2\right] + 4\,\mathbb{E}\left[\sup_{t\in[0,T]} |\frac{1}{n}\sum_{j=1}^n (X_t^{j,n} - X_t^j)|^2\right] \\
&\qquad\qquad + 4\,\mathbb{E}\left[\sup_{t\in[0,T]} |\frac{1}{n}\sum_{j=1}^n (X_t^j - \mathbb{E}[X_t])|^2\right].
\end{aligned}$$

Again, by Assumption 6 (ii) and the exchangeability of the process $\{X^{j,n} - X^j\}_{j=1}^n$, the first two terms in the last inequality go 0 as n goes to infinity. Now, since the processes $\{X^j\}_{j\ge1}$ are i.i.d. $C([0,T];\mathbb{R})$-valued random variables with finite second moments (since they are in $\mathcal{S}^2$), by the strong law of large numbers for Banach-valued r.v. (see Theorem 4.1.1 in [86]) and dominated convergence, it holds that

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{t\in[0,T]} \left|\frac{1}{n}\sum_{j=1}^n (X_t^j - \mathbb{E}[X_t])\right|^2\right] = 0.$$

This finishes the proof of (6.43).

Bibliography

[1] N. Agram and B. Øksendal. Optimal stopping of conditional mckean-vlasov jump diffusions. *preprint: arXiv:2207.13994v1 [math.OC]*, 2022.

[2] S. Albeverio, Y. Kondratiev, and M. Rockner. Differential geometry of Poisson spaces. *Comptes rendus de l'Académie des sciences. Série I. Mathématique*, 323(10):1129–1134, 1996.

[3] D. J. Aldous. *Weak convergence and the general theory of processes.* Preprint: `https://www.stat.berkeley.edu/~aldous/Papers/weak-gtp.pdf`, 1981.

[4] L. Ambrosio and W. Gangbo. Hamiltonian ODEs in the Wasserstein space of probability measures. *Comm. Pure Appl. Math.*, 61(1):18–53, 2008.

[5] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows In Metric Spaces and in the Space of Probability Measures.* Lectures in Mathematics. ETH Zürich. Birkhäuser Basel, 2005.

[6] C. Arenas. On randomized stopping times. *Trabajos de Estadística*, 5(2):3–7, 1990.

[7] J.-P. Aubin and G. Da Prato. Stochastic viability and invariance. *Annali della Scuola Normale Superiore di Pisa - Classe di Scienze*, Ser. 4, 17(4):595–613, 1990.

[8] A. Bain and D. Crisan. *Fundamentals of Stochastic Filtering.* Number 60 in Stochastic Modelling and Applied Probability. Springer-Verlag New York, 2009.

[9] E. Bandini, A. Cosso, M. Fuhrman, and H. Pham. Backward SDEs for optimal control of partially observed path-dependent stochastic systems: A control randomization approach. *The Annals of Applied Probability*, 28(3):1634–1678, 2018.

[10] E. Bandini, A. Cosso, M. Fuhrman, and H. Pham. Randomized filtering and Bellman equation in Wasserstein space for partial observation control problem. *Stochastic Processes and their Applications*, 129(2):674–711, 2019.

[11] J. R. Baxter and R. V. Chacon. Compactness of stopping times. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 40(3):169–181, 1977.

[12] E. Bayraktar, A. Cosso, and H. Pham. Randomized dynamic programming principle and Feynman-Kac representation for optimal control of McKean-Vlasov dynamics. *Trans. Amer. Math. Soc.*, 370:2115–2160, 2018.

[13] A. Bensoussan. *Stochastic Control of Partially Observable Systems.* Cambridge University Press, 1992.

[14] A. Bensoussan and S. C. P. Yam. Mean field approach to stochastic control with partial information. *ESAIM: COCV*, 27:89, 2021.

[15] C. Bertucci. Optimal stopping in mean field games, an obstacle problem approach. *J. Math. Pures Appl. (9)*, 120:165–194, 2018.

[16] A. G. Bhatt and G. Kallianpur. Uniqueness and robustness of solution of measure-valued equations of nonlinear filtering. *The Annals of Applied Probability*, 23(4):1895–1938, 1995.

[17] J.-M. Bismut and B. Skalli. Temps d'arrêt optimal, théorie générale des processus et processus de Markov. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 39(4):301–313, 1977.

[18] V. I. Bogachev. *Measure Theory*, volume 2. Springer-Verlag Berlin Heidelberg, 2007.

[19] V. I. Bogachev, N. V. Krylov, M. Rockner, and S. V. Shaposhnikov. *Fokker-Planck-Kolmogorov Equations.* American Mathematical Society, 2015.

[20] V. S. Borkar. The Nisio semigroup for controlled diffusions with partial observations. *Stochastics*, 17:145 – 160, 1986.

[21] V. S. Borkar. *Probability Theory: An Advanced Course.* Springer, 1995.

[22] J. M. Borwein and Q. Zhu. *Techniques of variational analysis.* CMS books in mathematics 20. Springer, 2005.

[23] G. Bouveret, R. Dumitrescu, and P. Tankov. Mean-Field Games of Optimal Stopping: A Relaxed Solution Approach. *SIAM Journal on Control and Optimization*, 58(4):1795–1821, 2020.

[24] R. Buckdahn, J. Li, S. Peng, and C. Rainer. Mean-field stochastic differential equations and associated PDEs. *The Annals of Probability*, 45(2):824 – 878, 2017.

[25] I. Bucur and G. Paltineanu. *Topics in Uniform Approximation of Continuous Functions.* Birkhäuser Basel, 2020.

[26] R. S. Bucy. Optimum finite time filters for a special non-stationary class of inputs. Technical report, Johns Hopkins Univ Silver Spring Md Applied Physics Lab, 1959.

[27] M. Burzoni, V. Ignazio, M. Reppen, and M. Soner. Viscosity solutions for controlled McKean–Vlasov jump-diffusions. *SIAM J. Control Optim.*, 58(3):1676–1699, 2020.

[28] P. Cardaliaguet. Notes on mean field games. 2013.

[29] P. Cardaliaguet, M. Cirant, and A. Porretta. Splitting methods and short time existence for the master equations in mean field games. *J. Eur. Math. Soc.*, 2022.

[30] P. Cardaliaguet, F. Delarue, J.-M. Lasry, and P.-L. Lions. *The Master Equation and the Convergence Problem in Mean Field Games.* Princeton University Press, 2019.

[31] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications*, volume 1. Springer International Publishing, 2018.

[32] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications*, volume 2. Springer International Publishing, 2018.

[33] R. Carmona, F. Delarue, and D. Lacker. Mean field games of timing and models for bank runs. *Applied Mathematics & Optimization*, 76(1):217–260, 2017.

[34] A. Cosso, F. Gozzi, I. Kharroubi, H. Pham, and M. Rosestolato. Optimal control of path-dependent McKean-Vlasov SDEs in infinite dimension. *arxiv: 2012.14772*, 2020.

[35] A. Cosso, F. Gozzi, I. Kharroubi, H. Pham, and M. Rosestolato. Master Bellman equation in the Wasserstein space: Uniqueness of viscosity solutions. *arXiv: 2107.10535*, 2021.

[36] A. Cosso and F. Russo. Crandall–Lions viscosity solutions for path-dependent PDEs: The case of heat equation. *Bernoulli*, 28(1):481–503, 2022.

[37] M. G. Crandall, H. Ishii, and P.-L. Lions. User's guide to viscosity solutions of second-order partial differential equations. *Bulletin of the AMS*, 27:1–67, 1992.

[38] C. Cuchiero, M. Larsson, and S. Svaluto-Ferro. Probability measure-valued polynomial diffusions. *Electronic Journal of Probability*, 24(30):1–32, 2019.

[39] G. Da Prato and H. Frankowska. Invariance of stochastic control systems with deterministic arguments. *Journal of Differential Equations*, 200(1):18–52, 2004.

[40] G. Da Prato and H. Frankowska. Stochastic viability of convex sets. *Journal of Mathematical Analysis and Applications*, 333(1):151–163, 2007.

[41] G. Da Prato and J. Zabczyk. *Second Order Partial Differential Equations in Hilbert Spaces.* London Mathematical Society Lecture Note Series. Cambridge University Press, 2002.

[42] D. Dawson. Measure-valued Markov processes. In *Ecole d'Ete de Probabilites de Saint-Flour XXI - 1991*, number 1541 in Lectur Notes in Mathematics. Springer-Verlag Berlin Heidelberg, 1991.

[43] B. Djehiche, R. Dumitrescu, and J. Zeng. A propagation of chaos result for weakly interacting nonlinear snell envelopes. preprint, 2021.

[44] B. Djehiche, R. Elie, and S. Hamadène. Mean-field reflected backward stochastic differential equations. preprint, 2019.

[45] D. Duffie and L. G. Epstein. Asset pricing with stochastic differential utility. *The Review of Financial Studies*, 5(3):411–436, 1992.

[46] D. Duffie and L. G. Epstein. Stochastic Differential Utility. *Econometrica*, 60(2):353–394, 1992.

[47] R. Dumitrescu, M. Leutscher, and P. Tankov. Control and optimal stopping mean field games: a linear programming approach. *Electron. J. Probab.*, 26:Paper No. 157, 1–49, 2021.

[48] R. Dumitrescu, M. Leutscher, and P. Tankov. Linear programming fictitious play algorithm for mean field games with optimal stopping and absorption. preprint, 2022.

[49] T. E. Duncan. Likelihood functions for stochastic signals in white noise. *Information and Control*, 16:303–310, 1970.

[50] T. E. Duncan. On the absolute continuity of measures. *Ann. Math. Statist.*, 41:30–38, 1970.

[51] T. E. Duncan. Stochastic filtering in manifolds. *IFAC Proceedings Volumes*, 14(2):553–556, 1981. 8th IFAC World Congress on Control Science and Technology for the Progress of Society, Kyoto, Japan, 24-28 August 1981.

[52] G. A. Edgar, A. Millet, and L. Sucheston. On compactness and optimality of stopping times. In *Martingale theory in harmonic analysis and Banach spaces*, pages 36–61. Springer, 1982.

[53] N. El Karoui. Les aspects probabilistes du contrôle stochastique. In *Ninth Saint Flour Probability Summer School—1979 (Saint Flour, 1979)*, volume 876 of *Lecture Notes in Math.*, pages 73–238. Springer, Berlin-New York, 1981.

[54] N. El Karoui, J. Lepeltier, and A. Millet. A probabilistic approach to the reduite in optimal stopping. *Probab. Math. Statist*, 13(1):97–121, 1992.

[55] N. El Karoui, S. Peng, and M. C. Quenez. Backward stochastic differential equations in finance. *Math. Finance*, 7(1):1–71, 1997.

[56] N. El Karoui, S. Peng, and M. C. Quenez. A dynamic maximum principle for the optimization of recursive utilities under constraints. *Ann. Appl. Probab.*, 11(3):664–693, 2001.

[57] G. Fabbri, F. Gozzi, and A. Swiech. *Stochastic optimal control in infinite dimension.* Springer, 2017.

[58] F. Flandoli and G. Zanco. An infinite-dimensional approach to path-dependent Kolmogorov equations. *The Annals of Probability*, 44(4):2643–2693, 2016.

[59] W. H. Fleming. Nonlinear semigroup for controlled partially observed diffusions. *SIAM J. Control Optim.*, 20(2):286 – 301, 1982.

[60] A. Friedman. *Stochastic Differential Equations and Applications*, volume 1. Academic Press, 1975.

[61] A. Friedman. *Stochastic Differential Equations and Applications*, volume 2. Academic Press, 1975.

[62] M. Fujisaki, G. Kallianpur, and H. Kunita. Stochastic differential equations for the non linear filtering problem. *Osaka J. Math.*, 9:19–40, 1972.

[63] W. Gangbo and A. Tudorascu. On differentiability in the Wasserstein space and well-posedness for Hamilton–Jacobi equations. *Journal de Mathématiques Pures et Appliquées*, 125:119–174, 2019.

[64] W. R. P. Hammersley, Š. D., and L. Szpruch. Weak existence and uniqueness for McKean–Vlasov SDEs with common noise. *The Annals of Probability*, 49(2):527–555, 2021.

[65] B. Jourdain, S. Méléard, and W. Woyczynski. Nonlinear SDEs driven by Lévy processes and related PDEs. *Alea*, 4:1–29, 2008.

[66] T. Kailath. An innovations approach to least-squares estimation. i. linear filtering in additive white noise. *IEEE Trans. Automat. Control*, 13:646–655, 1968.

[67] R. E. Kalman. A new approach to linear filtering and prediction problems. *J. Basic Eng.*, 82:35–45, 1960.

[68] I. Karatzas and S. E. Shreve. *Methods of mathematical finance*, volume 39 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1998.

[69] A. N. Kolmogorov. Sur l'interpolation et extrapolation des suites stationnaires. *C. R. Acad. Sci.*, 208:2043, 1939.

[70] A. N. Kolmogorov. Interpolation and extrapolation. *Bulletin de l'académie des sciences de U.S.S.R.*, 5:3–14, 1941.

[71] V. N. Kolokoltsov. *Nonlinear Markov Processes and Kinetic Equations.* Cambridge Tracts in Mathematics. Cambridge University Press, 2010.

[72] M. G. Krein. On a generalization of some investigations of G. Szegö, W. M. Smirnov, and A. N. Kolmogorov. *Dokl. Adad. Nauk SSSR*, 44(1-2):65–102, 1945.

[73] M. G. Krein. On a problem of extrapolation of A. N. Kolmogorov. *Dokl. Adad. Nauk SSSR*, 46:306–309, 1945.

[74] N. V. Krylov, M. Röckner, and J. Zabczyk. *Stochastic PDE's and Kolmogorov Equations in Infinite Dimensions*, volume 1715 of *Lectur Notes in Mathematics.* Springer, 1999.

[75] T. G. Kurtz and D. L. Ocone. Unique characterization of conditional distributions in nonlinear filtering. *The Annals of Probability*, 16(1):80–107, 1988.

[76] H. Kushner. On the differential equations satisfied by conditional densities of Markov processes, with applications. *SIAM J. Control*, 2:106–119, 1964.

[77] H. Kushner. Approximations of nonlinear filters. *IEEE Trans. Automat. Control*, 12:546–556, 1967.

[78] 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.

[79] P.-L. Lions. Cours au Collège de France, 2007-2013.

[80] V. M. Lucic and A. J. Heunis. On uniqueness of solutions for the stochastic differential equations of nonlinear filtering. *Ann. Appl. Probab.*, 11(1):182–209, 2001.

[81] M. Martini. Kolmogorov equations on spaces of measures associated to nonlinear filtering processes. *arXiv: 2107.11865*, 2021.

[82] R. E. Mortensen. Stochastic optimal control with noisy observations. *Internat. J. Control*, 1:455–464, 1966.

[83] C. Mou and J. Zhang. Mean field games of controls: Propagation of monotonicities. *Probability, Uncertainty and Quantitative Risk*, 7(3):247–274, 2022.

[84] S. Ng and P. Caines. Nonlinear filtering in riemannian manifolds. *IFAC Proceedings Volumes*, 17(2):817 – 821, 1984. 9th IFAC World Congress: A Bridge Between Control Science and Technology, Budapest, Hungary, 2-6 July 1984.

[85] M. Nutz. A mean field game of optimal stopping. *SIAM Journal on Control and Optimization*, 56(2):1206–1221, 2018.

[86] W. J. Padgett and R. L. Taylor. *Laws of large numbers for normed linear spaces and certain Fréchet spaces*, volume 360. Springer, 2006.

[87] E. Pardoux. Filtrage non lineaire et equations aux derivees partielles stochastiques associees. In *Ecole d'Eté de Probabilités de Saint-Flour XIX - 1989*, volume 1464 of *Lectur Notes in Mathematics*. Springer, 1991.

[88] J. L. Pedersen. Explicit solutions to some optimal variance stopping problems. *Stochastics*, 83(4-6):505–518, 2011.

[89] J. L. Pedersen and G. Peskir. Optimal mean-variance selling strategies. *Math. Financ. Econ.*, 10(2):203–220, 2016.

[90] H. Pham. *Continuous-time Stochastic Control and Optimization with Financial Applications*. Number 61 in Stochastic Modelling and Applied Probability. Springer, 2009.

[91] H. Pham and X. Wei. Dynamic programming for optimal control of stochastic mckean-vlasov dynamics. *SIAM J. Control Optim.*, 55(2):1069–1101, 2017.

[92] H. Pham and X. Wei. Bellman equation and viscosity solutions for mean-field stochastic control problem. *ESAIM Control Optim. Calc. Var.*, 24(1):437–461, 2018.

[93] P. Ren and F. Y. Wang. Derivative formulas in measure on Riemannian manifolds. *Bulletin of the London Mathematical Society*, 53(6):1786 – 1800, 2021.

[94] B. L. Rozovsky. *Stochastic Evolution Systems*. Number 35 in Mathematics and its Applications. Springer Netherlands, 1990.

[95] H. M. Soner and Q. Yan. Viscosity solutions for mckean-vlasov control on a torus, 2022.

[96] R. L. Stratonovich. Application of the theory of Markov processes for optimum filtration of signals. *Radio Eng. Electron. Phys*, 1:1–19, 1960.

[97] R. L. Stratonovich. Conditional Markov processes. *Theory of Probability Applications*, 5:156–178, 1960.

[98] A. Swiech. Unbounded second order partial differential equations in infinite dimensional Hilbert spaces. *Communications in Partial Differential Equations*, 19:1999–2036, 1994.

[99] J. Szpirglas. Sur l'équivalence d'équations différentielles stochastiques à valeurs mesures intervenant dans le filtrage Markovien non linéaire. *Ann. Inst. H. Poincaré Sect. B (N.S.)*, 14(1):33–59, 1978.

[100] M. Talbi, N. Touzi, and J. Zhang. Dynamic programming equation for the mean field optimal stopping problem, 2021.

[101] C. Villani. *Optimal transport - Old and new*. Springer, 2008.

[102] N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications*. MIT Press, 1949.

[103] C. Wu and J. Zhang. Viscosity solutions to parabolic master equations and McKean–Vlasov SDEs with closed-loop controls. *The Annals of Applied Probability*, 30(2):936–986, 2020.

[104] J. Xiong. *An Introduction to Stochastic Filtering Theory.* Oxford University Press, 2008.

[105] M. Zakai. On the optimal filtering of diffusion processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 11:230–243, 1969.