



HAL
open science

Mean field optimal stopping and approximations of partial differential equations on Wasserstein space

Mehdi Talbi

► **To cite this version:**

Mehdi Talbi. Mean field optimal stopping and approximations of partial differential equations on Wasserstein space. General Mathematics [math.GM]. Institut Polytechnique de Paris, 2022. English. NNT : 2022IPPAX067 . tel-04094607

HAL Id: tel-04094607

<https://theses.hal.science/tel-04094607>

Submitted on 11 May 2023

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

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



INSTITUT
POLYTECHNIQUE
DE PARIS

NNT : 2022IPPAX067

Thèse de doctorat



Arrêt optimal champ-moyen et approximations d'équations aux dérivées partielles sur l'espace de Wasserstein

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à École polytechnique

École doctorale n°574 École doctorale de mathématiques Hadamard (EDMH)
Spécialité de doctorat : Mathématiques appliquées

Thèse présentée et soutenue à Palaiseau, le 30 septembre 2022, par

MEHDI TALBI

Composition du Jury :

Dylan Possamaï Professeur, Département de mathématiques, ETH Zurich	Président du jury
Erhan Bayraktar Professeur, Département de mathématiques, Université du Michigan	Rapporteur
Pierre Cardaliaguet Professeur, CEREMADE, Université Paris-Dauphine	Rapporteur
Charles Bertucci Chargé de recherche CNRS, CMAP, École polytechnique	Examineur
François Delarue Professeur, Laboratoire J.A. Dieudonné, Université Côte d'Azur	Examineur
Astrid Hilbert Professeur, Département de mathématiques, Université Linnaeus	Examineur
Nizar Touzi Professeur, CMAP, École polytechnique	Directeur de thèse
Jianfeng Zhang Professeur, Département de mathématiques, Université de Californie du Sud	Co-directeur de thèse

Contents

I	Introduction (Français)	9
I.1	Généralités sur le contrôle optimal champ-moyen	9
I.1.1	Motivations	9
I.1.2	Quelques éléments d'analyse sur l'espace de Wasserstein	11
I.1.3	Principe de la programmation dynamique	12
I.2	Le cas de l'arrêt optimal	13
I.2.1	Rappels sur l'arrêt optimal standard, et particularités de l'arrêt optimal champ-moyen	14
I.2.2	Problème de l'obstacle sur l'espace de Wasserstein	17
I.2.3	Caractérisation par les solutions de viscosité	20
I.3	Approximations d'équations aux dérivées partielles sur l'espace de Wasserstein	22
I.3.1	Cas du problème de l'obstacle	23
I.3.2	Cas d'une classe d'équations paraboliques	24
I.4	Vers une résolution numérique pour l'équation de l'obstacle en cascade	26
II	Introduction (English)	28
II.1	Generalities on mean field control	28
II.1.1	Motivations	28
II.1.2	Some elements of analysis on Wasserstein space	30
II.1.3	Dynamic programming principle	31
II.2	The case of optimal stopping	32
II.2.1	Reminder on standard optimal stopping, and particularities of mean field optimal stopping	33
II.2.2	Obstacle problem on Wasserstein space	36

CONTENTS

II.2.3	Characterization by viscosity solutions	38
II.3	Approximations of partial differential equations on Wasserstein space	40
II.3.1	Case of the obstacle problem	41
II.3.2	Case of a class of parabolic equations	43
II.4	Towards a numerical resolution for the cascade obstacle problem	44
III	Dynamic programming equation for the mean field optimal stopping problem	46
III.1	Introduction	46
III.2	Formulation of the mean field optimal stopping problem	50
III.3	Itô's formula for flows of laws of semimartingales	54
III.4	Obstacle problem on the Wasserstein space	57
III.4.1	The dynamic programming equation	57
III.4.2	The main results	58
III.4.3	Some discussions on optimal stopping policies	61
III.5	Examples	63
III.5.1	Connection with standard optimal stopping	63
III.5.2	Convex functions of the expectation	64
III.5.3	Construction of a smooth solution	64
III.6	Some extensions	67
III.6.1	Infinite horizon case	67
III.6.2	Mean field optimal stopping of a jump-diffusion	68
III.7	Proof of Proposition III.2.2	70
III.8	Proof of Theorem III.3.2	72
IV	Viscosity solutions for obstacle problems on Wasserstein space	74
IV.1	Introduction	74
IV.2	The obstacle problem on Wasserstein space	76
IV.2.1	Formulation	76
IV.2.2	Differential calculus	77
IV.2.3	The dynamic programming equation	78
IV.3	Viscosity solutions	80
IV.3.1	Definition and consistency	80

CONTENTS

IV.3.2	Some regularity results	81
IV.3.3	Viscosity property	83
IV.3.4	Stability	84
IV.3.5	Comparison	86
IV.3.6	Infinite horizon case	91
IV.4	Examples	91
IV.4.1	Connection with standard optimal stopping	91
IV.4.2	A generalization of the mean variance problem	92
IV.4.3	Expected shortfall	93
IV.4.4	Probability distortion	94
IV.5	Technical results	95
V	A finite-dimensional approximation for the obstacle problem on Wasserstein space	104
V.1	Introduction	104
V.2	Presentation of the problem	106
V.2.1	The mean field optimal stopping	106
V.2.2	Differential calculus on Wasserstein space	107
V.2.3	The multiple optimal stopping problem	108
V.3	Main results	108
V.4	Dynamic programming equations	109
V.4.1	Obstacle equation on Wasserstein space	110
V.4.2	The cascade obstacle equation	111
V.5	Convergence of the value function	112
V.6	Propagation of chaos for stopped McKean-Vlasov diffusions	119
V.7	Multiple optimal stopping: viscosity solutions with tangency in mean	124
V.7.1	Formulation of the problem	124
V.7.2	The single optimal stopping case	124
V.7.3	Reduction to a sequence of single optimal stopping problems	125
V.7.4	Viscosity solutions for the cascade obstacle equation	126
V.7.5	Construction of an optimal stopping strategy	127
V.8	Approximation of randomized stopping strategies by pure strategies	129
V.9	A useful estimate	131

VI A finite-dimensional approximation for viscosity solutions of partial differential equations on Wasserstein space	132
VI.1 Introduction	132
VI.2 Viscosity solutions of partial differential equations on Wasserstein space . . .	133
VI.2.1 Differentiability on Wasserstein space	133
VI.2.2 Partial differential equation on Wasserstein space	134
VI.2.3 Viscosity solutions	134
VI.3 Finite-dimensional approximation	137
VI.3.1 The approximating equation	137
VI.3.2 Viscosity solutions	137
VI.3.3 Main results	138
VI.4 Application to stochastic control	139
VI.4.1 Mean field control	139
VI.4.2 Zero-sum stochastic differential games	143
VI.4.3 The uncontrolled case	144
VI.5 Extension to path-dependent PDEs	144
VI.5.1 Pathwise derivatives	144
VI.5.2 Path-dependent equation on Wasserstein space	145
VI.5.3 Viscosity solutions	146
VI.5.4 Finite-dimensional approximation	146
VI.6 Proof of Theorems VI.3.4 and VI.5.6	148
VI.6.1 The Markovian setting	148
VI.6.2 The path-dependent setting	152
VI.7 A precompactness result	153
Acknowledgements	157
Bibliography	158

Résumé

Cette thèse comporte deux parties.

La première porte sur l'étude du problème d'arrêt optimal champ moyen, c'est-à-dire de l'arrêt optimal d'une diffusion du type McKean-Vlasov, lorsque le critère à optimiser est une fonction de la distribution du processus arrêté. Ce problème permet de modéliser la situation où un planificateur central, contrôlant une population infinie d'agents en interaction, doit attribuer à chaque agent un temps d'arrêt en vue d'optimiser un certain critère dépendant de la distribution du système.

Nous étudions ce problème via une approche type programmation dynamique, qui permet de caractériser sa fonction valeur par une équation aux dérivées partielles sur l'espace des mesures, que nous appelons *problème (ou équation) de l'obstacle sur l'espace de Wasserstein*, par analogie avec le problème de l'obstacle classique, rencontré notamment en arrêt optimal standard. Nous montrons en particulier que si cette équation dispose d'une solution classique, alors cette dernière est égale à la fonction valeur du problème d'arrêt optimal champ-moyen, et peut être utilisée pour caractériser les stratégies d'arrêt optimales.

Nous étendons ensuite cette étude au cas où la fonction valeur n'est pas nécessairement différentiable. Ainsi, nous introduisons une notion de solution de viscosité pour l'équation de l'obstacle sur l'espace de Wasserstein, pour laquelle nous montrons les propriétés de consistance avec les solutions classiques, de stabilité et d'unicité.

Dans la seconde partie de la thèse, nous nous intéressons au développement d'approximations pour certaines classes d'équations aux dérivées partielles sur l'espace des mesures de probabilité. Plus précisément, nous montrons que les solutions de viscosité de ces équations peuvent s'écrire comme limites de solutions de viscosité d'équations définies sur des espaces de dimension finie.

Nous nous intéressons d'abord au cas du problème de l'obstacle sur l'espace de Wasserstein, dont il s'avère que l'équation d'approximation correspond à l'équation de la programmation dynamique associée au problème d'arrêt optimal multiple, qui peut être vu comme une formulation du problème d'arrêt optimal champ-moyen en population finie.

Nous traitons enfin le cas d'une classe plus générale d'équations paraboliques sur l'espace de Wasserstein. Cette classe couvre notamment le cas des équations d'Hamilton-Jacobi-Bellman champ-moyen, ou encore celui des équations affleurant dans les problèmes de jeux

CONTENTS

différentiels champ-moyen. Nous montrons également que nos résultats peuvent être étendus aux équations dépendant de la trajectoire (c'est-à-dire, dont les variables de la solution ne sont plus le temps et une mesure sur \mathbb{R}^d , mais le temps et une mesure sur l'espace des trajectoires continues).

Abstract

This thesis consists in two parts.

The first one is concerned with the study of the mean field optimal stopping problem, that is the optimal stopping of a McKean-Vlasov diffusion, when the criterion to optimize is a function of the stopped process. This problem models the situation where a central planner controls a continuous infinity of interacting agents by assigning a stopping time to each of them, in order to optimize some criterion which depends on the distribution of the system.

We study this problem via a dynamic programming approach, which allows to characterize its value function by a partial differential equation on the space of probability measures, that we call *obstacle problem (or equation) on Wasserstein space* by analogy with the classical obstacle problem, which arises in particular in standard optimal stopping. We especially show that, if this equation has a classical solution, then it is equal to the value function of the mean field optimal stopping problem, and that it can be used to characterize optimal stopping policies.

We next extend our study to the case where the value function is not necessarily differentiable. Thus, we introduce a notion of viscosity solution for the obstacle problem on Wasserstein space, for which we prove the properties of consistency with classical solutions, stability and uniqueness.

In the second part of the thesis, we are interested in developing approximations for some classes of partial differential equations on the space of probability measures. More precisely, we show that viscosity solutions of these equations can be written as limits of viscosity solutions of equations defined on finite-dimensional spaces.

We first focus our study on the case of the obstacle problem on Wasserstein space, for which it turns out that the approximating equation corresponds to the dynamic programming equation associated with the multiple optimal stopping problem, which may be seen as a finite population formulation of the mean field optimal stopping problem.

We finally consider a larger class of parabolic equations on Wasserstein space. This class covers in particular the scope of mean field Hamilton-Jacobi-Bellman equations, or the case of equations arising in mean field differential games. We also prove that our results can be extended to the case of path-dependent equations (i.e., when the variables of the solution

are not the time and a measure on \mathbb{R}^d , but the time and a measure on the space of paths).

Notations

In the thesis, we shall systematically use the following notations:

- Let (E, \mathcal{A}) be a measurable space endowed with a metric d . We denote by $\mathcal{P}(E, \mathcal{A})$ the set of probability measures on (E, \mathcal{A}) , and by $\mathcal{P}_2(E, \mathcal{A})$ its subset of square integrable probability measures, equipped with the Wasserstein distance defined by

$$\mathcal{W}_2(\mu, \nu) := \inf_{\mathbb{Q} \in \Pi(\mu, \nu)} \left(\int_{E \times E} d^2(x, y) \mathbb{Q}(x, y) \right)^{1/2} \quad \text{for all } (\mu, \nu) \in \mathcal{P}_2(E, \mathcal{A})^2,$$

where $\Pi(\mu, \nu)$ is the set of couplings of μ and ν . When $\mathcal{A} = \mathcal{B}(E)$, the Borel σ -algebra of E , we simply write $\mathcal{P}(E)$ and $\mathcal{P}_2(E)$.

- We denote by $\text{supp}(\mu)$ the support of $\mu \in \mathcal{P}(E, \mathcal{A})$, defined as the smallest closed set $\mathcal{C} \subset E$ s.t. $\mu(\mathcal{C}^c) = 0$.
- Given a random variable Z and a probability \mathbb{P} , we denote by $\mathbb{P}_Z := \mathbb{P} \circ Z^{-1}$ the law of Z under \mathbb{P} .
- The space of the $d \times d$ real valued symmetric matrices is denoted by \mathbb{S}_d , and the subspace of non-negative symmetric matrices \mathbb{S}_d^+ .
- For vectors $x, y \in \mathbb{R}^d$ and matrices A, B , denote $x \cdot y := \sum_{i=1}^d x_i y_i$ and $A : B := \text{tr}(AB^\top)$.
- We shall also write "LSC" (resp. "USC") for "lower (resp. upper) semi-continuous".

Chapter I

Introduction (Français)

I.1 Généralités sur le contrôle optimal champ-moyen

I.1.1 Motivations

Bien qu'utilisée depuis longtemps en physique statistique, la notion de *champ-moyen* s'est durablement imposée en théorie du contrôle depuis qu'elle a été introduite, de manières indépendantes, par Lasry & Lions [53] et Caines, Huang & Malhammé [47] pour modéliser les problèmes de décision au sein de systèmes d'agents en interaction. La principale idée derrière de tels modèles est de postuler que, sous certaines hypothèses de symétrie et lorsque le nombre d'agents devient très grand, l'étude de ces systèmes peut se réduire à l'étude d'un agent représentatif qui agit en fonction l'état moyen du système, typiquement donné par la distribution des autres agents. Lorsque chaque agent cherche à optimiser son propre critère, de tels problèmes sont appelés *jeux à champ-moyen*; lorsqu'un planificateur central contrôle les agents en interaction pour optimiser un certain critère dépendant de l'état global du système, on parle de *contrôle optimal champ-moyen*. Le principal objet de cette thèse, l'arrêt optimal champ-moyen, appartient à cette seconde famille de problèmes.

Bien que moins traité que les problèmes de jeux, le contrôle optimal champ-moyen a bénéficié d'un intérêt croissant ces dernières années. D'abord étudiée par Bensoussan, Frehse & Yam [8] et Carmona, Delarue & Lachapelle [22] dans le cadre d'une étude comparative avec les jeux à champ-moyen, cette classe de problème a principalement été traitée via deux approches : par l'utilisation d'un principe du maximum (voir par exemple Buckdahn, Djehiche & Li [13], Andersson & Djehiche [1] ou Meyer-Brandis, Oksendal & Zhou [60]) et

par la programmation dynamique (voir par exemple Laurière & Pironneau [54], Pham & Wei [65], Bayraktar, Cosso & Pham [6], Wu & Zhang [77] ou Djete, Possamaï & Tan [34]). Des résumés de ces deux méthodes sont présentés dans Carmona & Delarue [21, Vol. I, Ch. 6]. L'approche par programmation dynamique aboutit principalement à la caractérisation de la fonction valeur du problème de contrôle par une équation d'Hamilton-Jacobi-Bellman sur l'espace des mesures de probabilité, parfois appelée en anglais *master equation*.

D'un point de vue probabiliste, l'intuition derrière les problèmes de contrôle optimal champ-moyen est la suivante : étant donné deux entiers $d, N \geq 1$, un ensemble de contrôles \mathcal{A} et $\mathbf{W} = (W^1, \dots, W^N)^\top$ un mouvement brownien standard de dimension $d \times N$, on considère le système de N particules en interaction dans \mathbb{R}^d :

$$dX_t^{k,N,\alpha} = b(t, X_t^{k,N,\alpha}, \mu^N(\mathbf{X}_t^{N,\alpha}), \alpha_t^k)dt + \sigma(t, X_t^{k,N,\alpha}, \mu^N(\mathbf{X}_t^{N,\alpha}), \alpha_t^k)dW_t^k, \quad (\text{I.1.1})$$

pour tout $k = 1, \dots, N$, avec $\mathbf{X}^{N,\alpha} := (X^{1,N,\alpha}, \dots, X^{N,N,\alpha})^\top$, $\alpha := (\alpha^1, \dots, \alpha^N)^\top \in \mathcal{A}^N$ le vecteur des contrôles et

$$\mu^N(\mathbf{X}_t^{N,\alpha}) := \frac{1}{N} \sum_{k=1}^N \delta_{X_t^{k,N,\alpha}}$$

la distribution empirique du système à l'instant t . Bien que chaque particule k soit contrôlée par le processus α^k , sa dynamique dépend également des autres contrôles $\{\alpha^j\}_{j \neq k}$, du fait de l'interaction via la distribution empirique du système. Adoptant le point de vue d'un planificateur central, on se donne alors le problème de contrôle

$$\sup_{\alpha \in \mathcal{A}^N} \mathbb{E} \left[\frac{1}{N} \sum_{k=1}^N \int_0^T f^k(r, \mathbf{X}_r^{N,\alpha}) dr + g(\mathbf{X}_T^{N,\alpha}) \right],$$

où $T > 0$ correspond à l'horizon de temps. Si f et g peuvent s'écrire comme des fonctionnelles définies sur l'espace des mesures de probabilité, c'est-à-dire peuvent se mettre sous la forme $f^k(t, \mathbf{x}) = f(t, x_k, \mu^N(\mathbf{x}))$ et $g(\mathbf{x}) = g(\mu^N(\mathbf{x}))$, on peut alors formuler l'approximation champ-moyen de ce problème. Plus précisément, quand $N \rightarrow \infty$, on considère que la dynamique d'une particule représentative du système est donnée par

$$dX_t^\alpha = b(t, X_t^\alpha, \mathcal{L}(X_t^\alpha), \alpha_t)dt + \sigma(t, X_t^\alpha, \mathcal{L}(X_t^\alpha), \alpha_t)dW_t, \quad (\text{I.1.2})$$

où $\mathcal{L}(X_t^\alpha)$ désigne la distribution de X_t^α , et W est un mouvement brownien standard. Le problème de contrôle champ-moyen s'écrit alors

$$\sup_{\alpha \in \mathcal{A}} \mathbb{E} \left[\int_0^T f(r, X_r^\alpha, \mathcal{L}(X_r^\alpha)) dr \right] + g(\mathcal{L}(X_T^\alpha)). \quad (\text{I.1.3})$$

Une telle approximation est essentiellement justifiée par les résultats du type *propagation du chaos* :

$$\frac{1}{N} \sum_{k=1}^N \delta_{\mathbf{X}^{k,N,\alpha}} \xrightarrow{N \rightarrow \infty} \mathcal{L}(X^\alpha) \text{ en distribution.}$$

Ces problèmes de convergence ont été abondamment étudiés. Nous renvoyons par exemple au célèbre article de Snitzman [72] ou à Oelschläger [63] pour le cas des diffusions non-contrôlées, et aux récents travaux de Lacker [52] et de Djete [33] pour le cas des diffusions contrôlées (pour le premier, dans le cadre des contrôles relaxés).

Outre l'intérêt d'un tel modèle pour étudier les systèmes de particules en interaction, il est intéressant d'observer que la classe des problèmes couverts par la formulation (I.1.3) est plus générale que celle traitée dans le cas du contrôle optimal standard, c'est-à-dire des problèmes de la forme

$$\sup_{\alpha \in \mathcal{A}} \mathbb{E} \left[\int_0^T f(r, X_r^\alpha) dr + g(X_T^\alpha) \right]. \quad (\text{I.1.4})$$

Mentionnons par exemple le problème d'optimisation moyenne-variance, donné par

$$\sup_{\alpha \in \mathcal{A}} \left\{ \mathbb{E}[X_T^\alpha] - \frac{1}{2} \text{Var}(X_T^\alpha) \right\}.$$

Comme la variance contient un terme quadratique en l'espérance, ce problème ne peut être écrit sous la forme (I.1.4), et n'est donc pas standard. Fischer & Livieri [42] en ont par ailleurs proposé une résolution lorsque X est un mouvement brownien géométrique, en résolvant explicitement le problème à N particules correspondant et en calculant sa limite lorsque $N \rightarrow \infty$.

I.1.2 Quelques éléments d'analyse sur l'espace de Wasserstein

Avant d'évoquer le principe de programmation dynamique pour le problème de contrôle champ-moyen et l'équation aux dérivées partielles associée, il est nécessaire d'introduire un certain nombre d'éléments d'analyse sur l'espace des mesures de probabilité. Pour un exposé plus détaillé sur ce sujet, nous renvoyons à Carmona & Delarue [21, Vol. I, Ch. 5]. Soit (E, d) un espace métrique, et $p \in [1, +\infty)$. On appelle *espace de Wasserstein d'ordre p* l'ensemble $\mathcal{P}_p(E)$ des mesures de probabilité qui possèdent un moment d'ordre p fini, et muni de la métrique

$$\mathcal{W}_p(\mu, \nu) := \left(\inf_{\gamma \in \Gamma(\mu, \nu)} \int_{E \times E} (d(x, y))^p \gamma(dx, dy) \right)^{1/p} \quad \text{pour tout } \mu, \nu \in \mathcal{P}_p(E),$$

où $\Gamma(\mu, \nu)$ désigne l'ensemble des couplages de μ et ν , c'est-à-dire l'ensemble des mesures de probabilité sur $E \times E$ ayant μ et ν comme lois marginales. Cette métrique est appelée *distance de Wasserstein d'ordre p* , ou encore *p -Wasserstein distance* en anglais. Si (E, d) est un espace polonais (c'est-à-dire complet et séparable), alors $(\mathcal{P}_p(E), \mathcal{W}_p)$ l'est également. Nous utiliserons souvent la propriété suivante, qui permet de caractériser la convergence au sens de \mathcal{W}_p (voir [21, Théorème 5.5]) : $\mathcal{W}_p(\mu_n, \mu) \xrightarrow{n \rightarrow \infty} 0$ si, et seulement si, la suite $\{\mu_n\}_{n \geq 1}$ converge faiblement vers μ et est p -uniformément intégrable, c'est-à-dire satisfait

$$\lim_{R \rightarrow \infty} \sup_{n \geq 1} \int_E (d(x_0, x))^p d\mu_n(x) = 0 \text{ pour un certain } x_0 \in E.$$

Il nous est également nécessaire d'introduire une notion de dérivée sur l'espace des mesures. On dit que $u : \mathcal{P}_p(E) \rightarrow \mathbb{R}$ admet une *dérivée fonctionnelle linéaire*, notée $\delta_m u$, si

$$u(\mu) - u(\nu) = \int_0^1 \int_E \delta_m u(\lambda\mu + (1-\lambda)\nu, x) d(\mu - \nu)(x) d\lambda \quad \text{pour tout } \mu, \nu \text{ dans } \mathcal{P}_p(E).$$

On remarque que cette dérivée est définie à une constante additive près, et qu'il s'agit d'une fonction de $\mathcal{P}_p(E) \times E$ dans \mathbb{R} . Par ailleurs, si $x \mapsto \delta_m u(\mu, x)$ est à son tour dérivable, les dérivées $\partial_x \delta_m u(\mu, \cdot)$ et $\partial_{xx}^2 \delta_m u(\mu, \cdot)$ coïncident avec les dérivées de Lions $\partial_\mu u(\mu, x)$ et $\partial_x \partial_\mu u(\mu, \cdot)$ (voir [58]), définies dans le cas $p = 2$ en réécrivant u comme une fonction définie sur l'espace de Hilbert des variables aléatoire de carré intégrable, et en utilisant la dérivée au sens de Fréchet.

Soit X est une semi-martingale continue à valeurs dans $E = \mathbb{R}^d$, $u : [0, T] \times \mathcal{P}_p(\mathbb{R}^d) \rightarrow \mathbb{R}$ admettant des dérivées $\partial_t u$, $\partial_x \delta_m u$ et $\partial_{xx}^2 \delta_m u$. Posons $\mu_t := \mathcal{L}(X_t)$ pour tout $t \in [0, T]$. Sous certaines conditions d'intégrabilité, nous avons la formule d'Itô,

$$\begin{aligned} u(t, \mu_t) &= u(0, \mu_0) + \int_0^t \partial_t u(s, \mu_s) ds \\ &\quad + \mathbb{E} \left[\int_0^t \partial_x \delta_m u(s, \mu_s, X_s) \cdot dX_s + \partial_{xx}^2 \delta_m u(s, \mu_s, X_s) : d\langle X \rangle_s \right], \end{aligned} \tag{I.1.5}$$

qui joue un rôle crucial dans l'établissement de l'équation de la programmation dynamique associée à un problème de contrôle champ-moyen.

I.1.3 Principe de la programmation dynamique

Comme dans le cas du contrôle standard, il est possible d'étudier les problèmes de contrôle champ-moyen en écrivant un principe de la programmation dynamique. Pour ce faire, il

est nécessaire de changer la variable d'état du problème. Dans le problème de contrôle standard (I.1.4), f et g sont définies sur un espace de dimension finie, typiquement \mathbb{R}^d , et la variable d'état adaptée est donc le processus X , à valeurs dans cet espace. En revanche, dans le problème de contrôle champ-moyen (I.1.3), elles sont définies sur un espace de dimension infinie, l'espace des mesures de probabilité. En conséquence, la variable d'état correspondante est la loi du processus X^α , ou plus précisément, dans le cas markovien, le flot de lois marginales $\{\mathcal{L}(X_t^\alpha)\}_{t \in [0, T]}$. Ainsi, après avoir réécrit (I.1.3) sous la forme dynamique

$$V(t, \mu) := \sup_{\alpha \in \mathcal{A}} \mathbb{E} \left[\int_t^T f(r, X_r^{t, \mu, \alpha}, \mathcal{L}(X_r^{t, \mu, \alpha})) dr \right] + g(\mathcal{L}(X_T^{t, \mu, \alpha})), \quad (t, \mu) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d),$$

où $X^{t, \mu, \alpha}$ correspond à la dynamique (I.1.2) conditionnée à $\mathcal{L}(X_t^\alpha) = \mu$, nous pouvons montrer que

$$V(t, \mu) = \sup_{\alpha \in \mathcal{A}} \mathbb{E} \left[\int_t^s f(r, X_r^{t, \mu, \alpha}, \mathcal{L}(X_r^{t, \mu, \alpha})) dr \right] + V(s, \mathcal{L}(X_s^{t, \mu, \alpha})) \text{ pour tout } s \in [t, T].$$

Si V est suffisamment régulière, nous pouvons alors associer cette égalité à la formule d'Itô (I.1.5) pour déduire que V est solution de

$$-\partial_t u(t, \mu) + H(t, \mu, \delta_m u(t, \mu, \cdot)) = 0, \quad u(T, \mu) = g(\mu), \quad (\text{I.1.6})$$

pour tout $(t, \mu) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$, avec

$$H(t, \mu, \varphi) := \int_E \sup_{a \in \mathcal{A}} \left(b(t, x, \mu, a) \cdot \partial_x \varphi(x) + \frac{1}{2} \sigma^2(t, x, \mu, a) : \partial_{xx}^2 \varphi(x) \right) \mu(dx)$$

pour toute $\varphi \in C^2(\mathbb{R}^d)$. Il est également possible d'établir (I.1.6) lorsque V n'est pas dérivable en recourant aux solutions de viscosité (voir par exemple Wu & Zhang [77] et Pham & Wei [65], qui utilisent deux notions de solutions de viscosité différentes).

Pour obtenir une caractérisation complète du problème (I.1.3) à l'aide de l'équation (I.1.6), il est nécessaire d'établir un résultat d'unicité, appelé *principe de comparaison* dans la littérature des solutions de viscosité. Un tel résultat est néanmoins difficile à prouver dans le cas général, et nous redirigeons le lecteur vers [77, Théorème 4.13] pour en voir une preuve dans un cas particulier.

I.2 Le cas de l'arrêt optimal

Les problèmes d'arrêt optimal correspondent à des problèmes de décision binaires : à chaque instant, la décision à prendre consiste à arrêter ou non un processus donné, en se basant

uniquement sur l'information disponible à cet instant et sans possibilité de revenir en arrière. Il est possible de l'illustrer par le problème du secrétaire : un employeur désire engager un secrétaire et se livre à une série d'entretiens avec les différents candidats. Après chaque entretien, il lui faut décider de recruter ou non le candidat.

Le principal but de cette thèse est d'établir une approche type programmation dynamique du problème d'arrêt optimal champ-moyen, qui constitue une sous-classe particulière des problèmes de contrôle optimal champ-moyen. Comme pour ces derniers, nous adoptons le point de vue d'un planificateur central contrôlant une distribution de particules en interaction. Dans ce cadre, il doit attribuer à chaque particule un temps d'arrêt en vue d'optimiser une certaine fonction de la distribution.

De la même manière que le contrôle champ-moyen se distingue des jeux à champ-moyen, le problème d'arrêt optimal champ-moyen est à distinguer des jeux d'arrêt à champ-moyen (parfois appelés *mean field games of timing* dans la littérature), qui ont été largement étudiés ces dernières années (voir, entre autres, Bertucci [9], Carmona, Delarue & Lacker [23], Nutz [62] et Bouveret, Dumitrescu & Tankov [11]).

I.2.1 Rappels sur l'arrêt optimal standard, et particularités de l'arrêt optimal champ-moyen

Rappelons d'abord quelques généralités à propos de l'arrêt optimal standard (nous renvoyons vers El Karoui [41] ou Shiryaev [70] pour des exposés détaillés). Soit X un processus de diffusion

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \tag{I.2.1}$$

\mathbb{F} la filtration engendrée par le mouvement Brownien W et, pour tout $t \in [0, T]$, soit $\mathcal{T}_{t,T}$ l'ensemble des \mathbb{F} -temps d'arrêt à valeurs dans $[t, T]$. On se donne le problème d'optimisation

$$\sup_{\tau \in \mathcal{T}_{0,T}} \mathbb{E} \left[\int_0^\tau f(r, X_r)dr + g(X_\tau) \right],$$

que l'on peut réécrire sous forme dynamique

$$V(t, x) := \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E} \left[\int_t^\tau f(r, X_r)dr + g(X_\tau) | X_t = x \right]. \tag{I.2.2}$$

Si V est continue, on peut montrer qu'elle satisfait le principe de la programmation dynamique

$$V(t, x) = \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E} \left[\int_t^{\tau \wedge \theta} f(r, X_r) dr + g(X_\tau) \mathbf{1}_{\theta \geq \tau} + V(\theta, X_\theta) \mathbf{1}_{\theta < \tau} \mid X_t = x \right],$$

qui permet de déduire, sous certaines hypothèses de croissance, que V est l'unique solution de viscosité du problème de l'obstacle

$$\min \left\{ -\partial_t u - b\partial_x u - \frac{1}{2}\sigma^2 \partial_{xx}^2 u - f, u - g \right\} = 0, \quad u(T, \cdot) = g. \quad (\text{I.2.3})$$

En utilisant cette équation dans le cas où $V \in C^{1,2}([0, T] \times \mathbb{R}^d)$, ou la théorie de l'enveloppe de Snell dans le cas général (voir par exemple Karatzas & Shreve [49, Appendice D]), on peut alors montrer que le temps d'arrêt

$$\tau^* := \inf \{s \geq t : V(s, X_s) = g(X_s)\}$$

est optimal pour le problème (I.2.2). En finance, ce type de problème apparaît naturellement dans le cadre des options américaines, qui permettent à leur acquéreur de les exercer à tout instant entre la date de souscription et la maturité. L'actif X sur lequel est construit l'option étant stochastique, l'acquéreur, qui ne peut prédire son évolution, doit résoudre un problème d'arrêt optimal pour déterminer la meilleure date d'exercice de l'option. De la même manière le vendeur, pour se couvrir, tarife l'option en calculant l'espérance de son *payoff* $g(X_\tau)$ sous la stratégie d'arrêt optimale.

Comme dans le cas du contrôle, et afin de mieux saisir l'intuition qui se cache derrière le modèle champ-moyen, nous présentons d'abord le problème d'arrêt optimal à N particules (appelé *arrêt optimal multiple*). Ce problème a été traité dans un cadre très général par Kobylanski, Quenez & Rouy-Mironescu [50]. On adopte une fois de plus le point de vue d'un planificateur central qui cherche à optimiser un certain critère en contrôlant toutes les particules du système. Le choix du contrôle pour une particule va influencer la dynamique des autres via le terme d'interaction. Les contrôles étant ici des temps d'arrêt, le seul moyen de rendre compte d'un tel impact est donc de rendre les temps d'arrêt intrinsèques à la dynamique du système, ce qui marque une première différence avec le cas de l'arrêt optimal standard où la dynamique (I.2.1) est non contrôlée. On choisit donc de formuler le problème ainsi : étant donné une suite de mouvements Browniens mutuellement indépendants $\{W^k\}_{k \geq 1}$, \mathbb{F}^N la filtration engendrée par $\{W^k\}_{1 \leq k \leq N}$, $\mathcal{T}_{0,T}^N$ l'ensemble des N -uplets

de \mathbb{F}^N -temps d'arrêt et $\boldsymbol{\tau} := (\tau_1, \dots, \tau_N) \in \mathcal{T}_{0,T}^N$, les particules suivent les dynamiques

$$dX_t^{k,N,\boldsymbol{\tau}} = b(t, X_t^{k,N,\boldsymbol{\tau}}, \mu^N(\mathbf{X}_t^{N,\boldsymbol{\tau}}))i_k \mathbf{1}_{t < \tau_k} dt + \sigma(t, X_t^{k,N,\boldsymbol{\tau}}, \mu^N(\mathbf{X}_t^{N,\boldsymbol{\tau}}))i_k \mathbf{1}_{t < \tau_k} dW_t^k \quad (\text{I.2.4})$$

pour tout $k = 1, \dots, N$, avec $\mathbf{X}^N = (X^{1,N,\boldsymbol{\tau}}, \dots, X^{N,N,\boldsymbol{\tau}})^\top$ et $\mathbf{i} := (i_1, \dots, i_N)^\top \in \{0, 1\}^N$. Cette équation est celle de la diffusion contrôlée (I.1.1) dans le cas où b et σ sont linéaires en α , pour

$$\boldsymbol{\alpha}_t = \mathbf{I}_t := (I_t^1, \dots, I_t^N)^\top := (i_1 \mathbf{1}_{t < \tau_1}, \dots, i_N \mathbf{1}_{t < \tau_N})^\top.$$

Le vecteur \mathbf{i} permet de signaler les particules qui sont arrêtées dès le temps initial, et s'avère crucial pour établir l'équation de la programmation dynamique caractérisant le problème d'arrêt optimal multiple, qui s'énonce ainsi:

$$\sup_{(\tau_1, \dots, \tau_N) \in \mathcal{T}_{0,T}^N} \mathbb{E} \left[\frac{1}{N} \sum_{k=1}^N \int_0^{\tau_k} f^{k,N}(r, \mathbf{X}_r^{N,\boldsymbol{\tau}}) dr + g^N(X_{\tau_1}^{1,N,\boldsymbol{\tau}}, \dots, X_{\tau_N}^{N,N,\boldsymbol{\tau}}) \right], \quad (\text{I.2.5})$$

Remarquons que, puisque $X_{\tau_k}^{k,N,\boldsymbol{\tau}} = X_T^{k,N,\boldsymbol{\tau}}$ pour tout k , et que $\boldsymbol{\tau}$ et \mathbf{I} sont en bijection, ce problème est équivalent à un problème de contrôle en \mathbf{I} . Néanmoins, étant donné la contrainte de monotonie sur \mathbf{I} (qui est décroissant), l'établissement d'un principe de la programmation dynamique n'est pas aussi immédiat que dans le cas du contrôle standard. Nous traitons le cas de l'arrêt optimal multiple dans le Chapitre V de cette thèse, qui correspond à [76].

Si $f^{k,N}(\cdot, \mathbf{x}) = f(\cdot, x_k, \mu^N(\mathbf{x}))$ pour tout $k \in \{1, \dots, N\}$ et $g^N(\mathbf{x}) = g(\mu^N(\mathbf{x}))$, alors le problème d'arrêt optimal multiple admet une formulation asymptotique, le problème d'*arrêt optimal champ-moyen*:

$$\sup_{\tau \in \mathcal{T}_{0,T}} \mathbb{E} \left[\int_0^\tau f(r, X_r, \mathcal{L}(X_r)) dr \right] + g(\mathcal{L}(X_\tau)), \quad (\text{I.2.6})$$

où X est une diffusion de type McKean-Vlasov arrêtée:

$$dX_t = b(t, X_t, \mathcal{L}(X_t))I_t dt + \sigma(t, X_t, \mathcal{L}(X_t))I_t dW_t, \quad \text{avec } I_t := I_{0-} \mathbf{1}_{t < \tau}, \quad (\text{I.2.7})$$

et I_{0-} est une variable à valeurs dans $\{0, 1\}$. Comme dans le cas de l'arrêt optimal multiple, il s'agit d'un problème de contrôle champ-moyen en I . Cependant, comme signalé ci-dessus, la monotonie inhérente à cette famille de contrôles rend moins immédiate l'écriture d'un principe de la programmation dynamique ; en particulier, il n'est pas suffisant de prendre $\mathcal{L}(X_t)$ comme variable d'état, contrairement au cas du contrôle optimal champ-moyen général.

Quelques exemples de tels problèmes d'arrêt optimal (critère non standard dépendant uniquement de la loi du processus arrêté) ont été abordés dans la littérature : nous pouvons ainsi évoquer les cas de l'arrêt optimal moyenne-variance par Pedersen & Peskir [64] et de la distortion de probabilité par Xu & Zhou [78], où le critère g est défini par

$$g(\mu) := \int_0^\infty \varphi(\mathbb{P}(U(\xi) \geq x)) dx,$$

où μ est une mesure de probabilité sur $(0, +\infty)$, ξ une variable aléatoire de loi μ sous une probabilité \mathbb{P} fixée, U une fonction d'utilité et φ une fonction continue de $[0, 1]$ dans lui-même appelée fonction de distortion. Ces deux exemples sont traités en horizon infini ($T = \infty$), lorsque X est un mouvement Brownien géométrique ayant un Dirac comme distribution de départ.

I.2.2 Problème de l'obstacle sur l'espace de Wasserstein

Le premier objectif de cette thèse est d'établir un principe de la programmation dynamique (PPD) pour caractériser la fonction valeur du problème. Cette tâche est retranscrite dans le Chapitre III, et correspond à [74].

Une idée naturelle est d'introduire une version dynamique du problème, sur le modèle du contrôle optimal champ-moyen (I.1.3). Soit $(t, \mu) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$, et

$$\begin{aligned} V(t, \mu) &:= \sup_{\tau \in \mathcal{T}_{t, T}} \mathbb{E} \left[\int_t^\tau f(r, X_r, \mathcal{L}(X_r)) dr \right] + g(\mathcal{L}(X_\tau)) \\ &= \sup_{I \in \mathbb{I}^0([t, T])} \mathbb{E} \left[\int_0^T f(r, X_r, \mathcal{L}(X_r)) I_r dr \right] + g(\mathcal{L}(X_T)), \end{aligned}$$

où X est une diffusion arrêtée (I.2.7) de processus de survie $I_s := \mathbf{1}_{s < \tau}$ telle que $\mathcal{L}(X_t) = \mu$, et où $\mathbb{I}^0([t, T])$ est l'ensemble des processus de survie (i.e., l'ensemble des processus de la forme $s \mapsto \mathbf{1}_{s < \theta}$, $\theta \in \mathcal{T}_{t, T}$). Il s'avère cependant que cette formulation est insuffisante pour obtenir un PPD. En effet, les particules arrêtées sur l'intervalle $[t, t + \delta)$, $\delta \in (0, T - t)$, le sont toujours sur l'intervalle $[t + \delta, T]$. Tout contrôle I admissible sur $[t + \delta, T]$ doit donc vérifier $I_{t+\delta} \leq I_{(t+\delta)-}^*$, où I^* est un contrôle optimal sur $[t, t + \delta)$. L'impossibilité d'obtenir un PPD avec cette formulation est donc due à la monotonie des éléments de $\mathbb{I}^0([t, T])$, qui sont par définition décroissants. C'est donc pour prendre ce fait en compte que l'on choisit pour variable d'état la loi jointe $\mathcal{L}((X_t, I_t))$, et non $\mathcal{L}(X_t)$. Intuitivement, cela signifie que l'information qui consiste à connaître les positions des particules dans \mathbb{R}^d (la distribution

$\mathcal{L}(X_t)$ n'est pas suffisante : il faut également connaître leur état, c'est-à-dire savoir si elles sont arrêtées ou non (information contenue dans la distribution $\mathcal{L}(X_t, I_t)$).

Nous adoptons alors la formulation suivante : $\Omega := C^0([0, T], \mathbb{R}^d) \times \mathbb{I}^0([-1, T])$, où $C^0([0, T], \mathbb{R}^d)$ désigne l'ensemble des trajectoires continues de $[0, T]$ dans \mathbb{R}^d , désigne notre espace canonique, et $Y := (X, I)$ le processus canonique sur Ω . Pour tout $(t, m) \in [0, T] \times \mathcal{P}_2(\mathbf{S})$, $\mathbf{S} := \mathbb{R}^d \times \{0, 1\}$, on désigne alors par $\mathcal{P}(t, m)$ l'ensemble des mesures de probabilité sur Ω telles que X ait la dynamique (I.2.7), \mathbb{P} -presque sûrement, et $\mathbb{P} \circ Y_t^{-1} = m$. Le problème d'arrêt optimal champ-moyen est alors défini par

$$V(t, m) := \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \mathbb{E} \left[\int_t^T f(r, X_r, \mathbb{P} \circ X_r^{-1}) dr \right] + g(\mathbb{P} \circ X_T^{-1}). \quad (\text{I.2.8})$$

Cette formulation, dite *faible* (le processus est fixé et l'on contrôle sa distribution), permet notamment de compactifier l'ensemble des contrôles en prenant également en compte les stratégies d'arrêt *randomisées*, qui consistent, à chaque instant, à pouvoir également arrêter X avec une certaine probabilité strictement comprise entre 0 et 1. Ainsi, il suffit que f et g soient semi-continues supérieurement pour que le problème (I.2.8) admette une solution. On s'assure par ailleurs qu'il s'agit effectivement d'un problème d'arrêt optimal en posant $\tau := \inf\{s \geq t : I_s = 0\}$ et en remarquant que, lorsque τ parcourt l'ensemble des temps d'arrêt (randomisés) à valeurs dans $[t, T]$, \mathbb{P} décrit $\mathcal{P}(t, m)$.

Une fois ce cadre posé, nous obtenons le PPD:

$$V(t, m) = \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \mathbb{E} \left[\int_t^s f(r, X_r, \mathbb{P}_{Y_r}) dr \right] + V(s, \mathbb{P}_{Y_{s-}}) \text{ pour tout } s \in [t, T]. \quad (\text{I.2.9})$$

Comme souvent en contrôle stochastique, on établit l'équation de la programmation dynamique associée à notre problème d'optimisation en appliquant une formule d'Itô au PPD. Cependant, le processus Y est une semimartingale càdlàg et donc potentiellement non continue, ce qui nous empêche d'utiliser (I.1.5) ; par ailleurs, son flot de lois marginales $\{\mathbb{P}_{Y_s}\}_{s \in [t, T]}$ est lui-même càdlàg, ce qui interdit d'utiliser certaines formules d'Itô pour les lois de diffusions à sauts trouvables dans la littérature (voir par exemple Burzoni, Ignazio, Reppen & Soner [15]). Nous sommes ainsi conduits à prouver une formule d'Itô pour le cas général d'une semimartingale Y càdlàg. Si $\mathbf{m} := \{m_t\}_{t \in [0, T]}$ désigne son flot de lois marginales, et $J_{\mathbb{T}}(\mathbf{m})$ l'ensemble des points de discontinuité de $t \mapsto m_t$ sur $\mathbb{T} \subset [0, T]$, alors

sous certaines conditions d'intégrabilité :

$$\begin{aligned}
 u(t, m_t) &= u(0, m_0) + \int_0^t \partial_t u(s, m_s) ds \\
 &+ \mathbb{E} \left[\int_0^t \partial_y \delta_m u(s, m_s, Y_s) \cdot dY_s^c + \partial_{yy}^2 \delta_m u(s, m_s, Y_s) : d\langle Y^c \rangle_s \right] \\
 &+ \sum_{s \in J_{(0,t]}(\mathbf{m})} (u(s, m_s) - u(s, m_{s-})) + \mathbb{E} \left[\sum_{s \in J_{(0,t]}^c(\mathbf{m})} (\delta_m u(s, m_s, Y_s) - \delta_m u(s, m_s, Y_{s-})) \right],
 \end{aligned} \tag{I.2.10}$$

où Y^c désigne la partie continue de Y . Toutefois, seul le cas particulier $Y = (X, I)$ nous intéresse dans le cadre de l'arrêt optimal champ-moyen.

L'application de cette formule au PPD prouve donc que V , si elle est suffisamment régulière, est solution de l'équation de la programmation dynamique :

$$\min_{m' \in C_u(t, m)} \{-(\mathbb{L}u + F)(t, m')\} = 0, \quad D_I u(t, m, \cdot) \geq 0, \quad u(T, \cdot) = g, \tag{I.2.11}$$

pour tout $(t, m) \in [0, T] \times \mathcal{P}_2(\mathbf{S})$, avec

$$\begin{aligned}
 \mathbb{L}u(t, m) &:= \partial_t u(t, m) \\
 &+ \int_{\mathbb{R}^d} (b(t, x, m) \cdot \partial_x \delta_m u(t, m, x, 1) + \frac{1}{2} \sigma^2(t, x, m) : \partial_{xx}^2 \delta_m u(t, m, x, 1)) m(dx, 1), \\
 F(t, m) &:= \int_{\mathbb{R}^d} f(t, x, m) m(dx, 1) \\
 C_u(t, m) &:= \{m' \preceq m : u(t, m') = u(t, m)\}, \\
 D_I u(t, m, x) &:= \delta_m u(t, m, x, 1) - \delta_m u(t, m, x, 0) \quad \text{pour tout } x \in \mathbb{R}^d.
 \end{aligned}$$

La notation \preceq désigne ici un ordre partiel sur $\mathcal{P}_2(\mathbf{S})$: nous avons $m' \preceq m$ s'il existe une fonction mesurable $p : \mathbb{R}^d \rightarrow [0, 1]$ telle que

$$m'(dx, 1) = p(x)m(dx, 1) \text{ et } m'(dx, 0) = (1 - p(x))m(dx, 1) + m(dx, 0).$$

Intuitivement, cela signifie que m' peut être obtenue en arrêtant chaque particule non déjà arrêtée de la distribution m se trouvant en position x avec probabilité $p(x)$. $C_u(t, m)$ désigne donc les stratégies d'arrêt à l'instant t qui préservent la fonction valeur. La condition $D_I u \geq 0$, qui implique que u est croissante pour l'ordre partiel \preceq , exprime la monotonie naturelle du problème : la fonction valeur est d'autant plus grande que l'ensemble des particules que l'on contrôle l'est.

Par analogie avec l'arrêt optimal standard, nous appelons cette équation *problème de l'obstacle sur l'espace de Wasserstein*. Nous montrons par ailleurs un théorème de vérification, c'est-à-dire que toute solution suffisamment régulière de cette équation est égale à la fonction valeur (I.2.8). Nous utilisons également ce problème de l'obstacle pour caractériser les stratégies d'arrêt optimales.

I.2.3 Caractérisation par les solutions de viscosité

Dans cette deuxième partie de notre travail, nous étendons la caractérisation du problème d'arrêt optimal champ-moyen (I.2.8) par le problème de l'obstacle (I.2.11) dans le cas où V n'est pas différentiable. Autrement dit, nous entendons définir la notion de solution faible de (I.2.11), au sens des solutions de viscosité. Introduite par Crandall & Lions [30] dans le cadre des équations d'Hamilton-Jacobi, cette notion permet de définir des solutions faibles pour des équations aux dérivées partielles du second ordre elliptiques, notamment non linéaires. Contrairement aux solutions au sens de Sobolev, les solutions de viscosité sont définies en tout point. Elles permettent d'obtenir des résultats d'existence et d'unicité sous des hypothèses de régularité minimales (typiquement, pour des fonctions localement bornées). Nous renvoyons le lecteur vers l'exposé de Crandall, Ishii & Lions [29] pour plus de détails sur les solutions de viscosité dans le cadre d'espaces de dimension finie.

Motivé entre autres par les problèmes de jeux et de contrôle champ-moyen, l'étude des solutions de viscosité des équations sur l'espace de Wasserstein a connu un vif intérêt ces dernières années et a été menée via des approches très diverses : par l'utilisation de sous-différentielles (Cardaliaguet & Quincampoix [18], Gangbo, Nguyen & Tudorascu [45]), en transposant l'équation sur l'espace de Hilbert des variables aléatoires (Pham & Wei [65]) ou en utilisant directement les dérivées sur l'espace des mesures et en construisant des voisinages de viscosité compacts dans l'espace de Wasserstein (Wu & Zhang [77], Burzoni, Ignazio, Reppen & Soner [15]).

Dans cette thèse, nous nous inspirons principalement de l'approche de [77], où les auteurs définissent une notion de solution de viscosité pour une classe générale d'équations paraboliques sur l'espace de Wasserstein. Cette notion est intrinsèque : contrairement à [65], qui utilise une méthode dite de *lifting*, elle ne recourt à aucun changement d'espace. Sa principale force réside dans le choix du voisinage de viscosité, dont la compacité est essentielle pour garantir les propriétés de stabilité et d'unicité des solutions. La boule

unité sur $\mathcal{P}_2(\mathbb{R}^d)$ n'étant pas compacte, les auteurs décident de considérer l'ensemble des flots de lois marginales de semimartingales à caractéristiques bornées partant d'un point $(t, \mu) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$. De simples arguments de tension de mesures montrent qu'un tel ensemble est compact (voir par exemple Meyer & Zheng [59]).

Dans le cadre de l'arrêt optimal champ-moyen, nous considérons ainsi l'ensemble des lois marginales des diffusions de type McKean-Vlasov arrêtées (I.2.7). Plus précisément, étant donné $(t, m) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$ et $\delta \in (0, T - t)$, nous définissons le voisinage de viscosité :

$$\mathcal{N}_\delta(t, m) := \{(s, \tilde{m}) : s \in [t, t + \delta], \tilde{m} \in \{\mathbb{P}_{Y_{s-}}, \mathbb{P}_{Y_s}\}, \mathbb{P} \in \mathcal{P}(t, m)\}.$$

Le flot $s \mapsto \mathbb{P}_{Y_s}$ étant càdlàg pour tout $\mathbb{P} \in \mathcal{P}(t, m)$, il est en effet nécessaire de prendre en compte tous les couples $\{\mathbb{P}_{Y_{s-}}, \mathbb{P}_{Y_s}\}$ pour assurer la compacité de $\mathcal{N}_\delta(t, m)$. Étant donné une fonction localement bornée $u : [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$, on définit alors les ensembles de fonctions test :

$$\begin{aligned} \overline{\mathcal{A}}u(t, m) &:= \left\{ \varphi \in C_2^{1,1}(\mathbf{Q}_t) : (\varphi - u_*)(t, m) = \max_{\mathcal{N}_\delta(t, m)} (\varphi - u_*) \text{ pour un } \delta > 0 \right\}, \\ \underline{\mathcal{A}}u(t, m) &:= \left\{ \varphi \in C_2^{1,1}(\mathbf{Q}_t) : (\varphi - u^*)(t, m) = \min_{\mathcal{N}_\delta(t, m)} (\varphi - u^*) \text{ pour un } \delta > 0 \right\}, \end{aligned}$$

où u_* et u^* désignent respectivement les enveloppes semi-continues inférieurement et supérieurement de u . En raison de la structure particulière de l'équation (I.2.11), la définition des solutions de viscosité de solutions est cependant moins classique qu'à l'accoutumée. En effet, pour une équation usuelle

$$\mathcal{D}u = 0,$$

où \mathcal{D} désigne un opérateur différentiel quelconque, les sur-solutions et sous-solutions de viscosité se définissent naturellement en tout point (t, m) par

$$\mathcal{D}\varphi(t, m) \geq 0 \quad \text{et} \quad \mathcal{D}\psi(t, m) \leq 0$$

pour tout $\varphi \in \overline{\mathcal{A}}u(t, m)$ et $\psi \in \underline{\mathcal{A}}u(t, m)$. La présence de l'inégalité $D_I u \geq 0$ dans (I.2.11) nous contraint à une certaine asymétrie entre les définitions de sur-solution et sous-solution. Ainsi, nous posons:

(i) u est une sur-solution de viscosité de (I.2.11) si, pour tout $(t, m) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$, $\varphi \in \overline{\mathcal{A}}u(t, m)$ et $m' \preceq m$,

$$u_*(t, m) \geq u_*(t, m') \quad \text{et} \quad -(\mathbb{L}\varphi + F)(t, m) \geq 0,$$

(ii) u est une sous-solution de viscosité de (I.2.11) si, pour tout $(t, m) \in [0, T) \times \mathcal{P}_2(\mathbb{R}^d)$ tel que $C_{u^*}(t, m) = \{m\}$ et $\varphi \in \underline{\mathcal{A}}u(t, m)$,

$$\min\{-(\mathbb{L}\varphi + F), (\mathbb{D}_I\varphi)_*\}(t, m) \leq 0,$$

où $\mathbb{D}_I\varphi(t, m) := \inf_{x \in \text{Supp}(m(\cdot, 1))} D_I u(t, m, x)$, $\text{Supp}(m(\cdot, 1))$ désignant le support de $m(\cdot, 1)$.

(iii) u est une solution de viscosité de (I.2.11) s'il s'agit à la fois d'une sur-solution et d'une sous-solution de viscosité.

Les solutions de viscosité ainsi définies sont consistantes avec les solutions classiques et satisfont la propriété de stabilité. Le résultat d'unicité, donné par un principe de comparaison, est comme souvent le résultat le plus délicat à prouver, d'autant plus lorsqu'il s'agit d'équations sur l'espace des mesures. C'est en renforçant les hypothèses de régularité sur les coefficients b et σ et en nous appuyant sur une suite régularisatrice pour les fonctions définies sur l'espace de Wasserstein (introduite par Mou & Zhang [61]) que nous parvenons à prouver ce résultat important.

I.3 Approximations d'équations aux dérivées partielles sur l'espace de Wasserstein

La deuxième partie de cette thèse est consacrée au développement d'approximations d'équations sur l'espace des mesures de probabilité. Plus précisément, notre but est d'approcher les solutions de viscosité de telles équations par les solutions de viscosité d'équations définies sur des espaces de dimension finie. Une telle démarche a déjà été effectuée par Gangbo, Mayora & Swiech [44] pour l'équation d'Hamilton-Jacobi-Bellman semi-linéaire (c'est-à-dire avec coefficient de diffusion non contrôlé), et dans le cadre d'une notion de solutions de viscosité différente de celle que nous adoptons. De manière plus lointaine, nous renvoyons également au travail de Cardaliaguet, Delarue, Lasry & Lions [17], qui montrent la convergence de la solution de l'équation caractérisant les jeux à N joueurs vers la solution de l'équation caractérisant l'équilibre de Nash des jeux à champ-moyen.

Dans l'esprit de Barles & Souganidis [2], qui prouvent la convergence d'un schéma numérique (dit *monotone*) pour les solutions de viscosité des équations d'Hamilton-Jacobi-Bellman, nous étudions la convergence de notre approximation de dimension finie vers la solution de l'équation champ-moyen via les semi-limites relaxées. Cette approche est complétée par l'utilisation de résultats type *propagation du chaos* (voir par exemple Snitzman

[72] ou Oelschläger [63]), qui permettent de faire le lien entre le problème à N particules et la formulation champ-moyen.

I.3.1 Cas du problème de l'obstacle

Du fait de sa structure particulière, nous traitons le cas du problème de l'obstacle sur l'espace de Wasserstein de manière isolée. Nous montrons principalement que l'équation d'approximation correspond à l'équation de la programmation dynamique caractérisant le problème d'arrêt optimal multiple (I.2.5).

Plus précisément, étant donné un espace de probabilité $(\Omega^0, \mathcal{F}_T^0, \mathbb{P}^0)$, on introduit le problème dynamique

$$V^N(t, \mathbf{y}) := \sup_{\tau \in \mathcal{T}_{t,T}^N} \mathbb{E}_{t,\mathbf{y}}^{\mathbb{P}^0} \left[\frac{1}{N} \sum_{k=1}^N \int_0^{\tau_k} f^{k,N}(r, \mathbf{X}_r^{N,\tau}) dr + g^N(X_{\tau_1}^{1,N,\tau}, \dots, X_{\tau_N}^{N,N,\tau}) \right] \quad (\text{I.3.1})$$

pour tout $\mathbf{y} \in \mathbf{S}^N$, où $\mathbf{Y}^{N,\tau} := (\mathbf{X}^{N,\tau}, \mathbf{I}^{N,\tau})$ suit la dynamique (I.2.4) et $\mathbb{E}_{t,\mathbf{y}}[\cdot] := \mathbb{E}[\cdot | Y_t^{N,\tau} = \mathbf{y}]$. Comme l'ont prouvé Kobylanski, Quenez & Rouy-Mironescu [50], ce problème peut être réduit à une suite de N problèmes d'arrêt optimal standards imbriqués, ce qui nous permet de montrer que (I.3.1) est caractérisé par l'équation

$$\begin{cases} \min \{ -\partial_t u(t, \mathbf{y}) - (\mathcal{L}u + \mathbf{f})(t, \mathbf{y}) \cdot \mathbf{i}, u(t, \mathbf{y}) - \max_{\mathbf{i}' < \mathbf{i}} u(t, \mathbf{x}, \mathbf{i}') \} = 0, \\ u|_{t=T} = u|_{\mathbf{i}=0} = g^N, \end{cases} \quad (\text{I.3.2})$$

où $\mathbf{y} = (\mathbf{x}, \mathbf{i})$, et $\mathbf{i}' < \mathbf{i}$ signifie que $i'_k \leq i_k$ pour tout $k = 1, \dots, N$ et $\mathbf{i}' \neq \mathbf{i}$, avec $\mathbf{i} := (i_1, \dots, i_N)^\top$, $\mathbf{i}' := (i'_1, \dots, i'_N)^\top$, $\mathbf{f} := (f^{1,N}, \dots, f^{N,N})^\top$. Nous appelons cette équation *problème de l'obstacle en cascade*. Nous montrons comment utiliser la solution de cette équation pour identifier une stratégie d'arrêt optimale pour (I.3.1).

Dans le cas où $f^{k,N}(\cdot, \mathbf{x}) = f(\cdot, x_k, \mu^N(\mathbf{x}))$ pour tout $k \in \{1, \dots, N\}$ et $g^N(\mathbf{x}) = g(\mu^N(\mathbf{x}))$, nous montrons que toute sur-solution ou sous-solution de viscosité de (I.3.2) converge, en un certain sens, vers une sur-solution ou sous-solution de viscosité de (I.2.11). Si les hypothèses garantissant l'unicité de la solution de (I.2.11) sont vérifiées, ce résultat implique que la fonction V^N converge vers V , la fonction valeur du problème d'arrêt optimal champ-moyen définie par (I.2.8), c'est-à-dire:

$$V^N(s, \mathbf{y}^N) \xrightarrow[\substack{N \rightarrow \infty, s \rightarrow t, \\ \mathcal{W}_2(m^N(\mathbf{y}^N), m) \rightarrow 0}]{} V(t, m) \quad \text{pour tout } (t, m) \in [0, T] \times \mathcal{P}_2(\mathbf{S}).$$

Le choix de la notion de solution de viscosité pour le problème de l'obstacle en cascade s'avère particulièrement important pour garantir ce résultat de convergence. En effet, il faut qu'elle s'accorde avec les solutions de viscosité pour l'équation de l'obstacle sur l'espace de Wasserstein dans le sens où, si φ est une fonction test pour ce dernier, alors sa projection sur \mathbf{S}^N définie par $\phi^N(t, \mathbf{y}) := \varphi(t, m^N(\mathbf{y}))$ doit être une fonction test pour le problème de dimension finie. Il s'avère alors que la bonne définition est celle donnée par le critère de « tangence en moyenne » : étant donné $u : [0, T] \times \mathbf{S}^N \rightarrow \mathbb{R}$ et $(t, \mathbf{y}) \in [0, T] \times \mathbf{S}^N$, on définit par exemple l'ensemble des fonctions test pour la sur-solution par

$$\bar{\mathcal{A}}u(t, \mathbf{y}) = \left\{ \phi \in C^{1,2}([t, T] \times \mathbf{S}) : (\phi - u)(t, \mathbf{y}) = \sup_{\theta \in \mathcal{T}_{t,T}} \mathbb{E}_{t, \mathbf{y}} [(\phi - u)(\theta \wedge H, \bar{Y}_{\theta \wedge H}^N)] \right\},$$

où \bar{Y} désigne la version non arrêtée de (I.2.4) et $H > t$ est un temps d'arrêt de localisation. Cette définition est une adaptation de celle introduite par Ekren [36] dans le cadre de l'arrêt optimal de diffusions dépendantes de la trajectoire au cas de l'arrêt optimal multiple.

Nous montrons également dans ce chapitre que la diffusion arrêtée à N particules (I.2.4) converge vers la diffusion McKean-Vlasov arrêtée (I.2.7) dans le sens suivant : pour tout $\tau \in \mathcal{T}_{t,T}^N$, quitte à se ramener à une sous-suite, la suite $\{\mathbb{P}^0 \circ (m^N(\mathbf{Y}^N, \tau))^{-1}\}_{N \geq 1}$ converge en loi vers une mesure supportée sur $\mathcal{P}(t, m)$. En particulier, si τ^N est une stratégie d'arrêt optimale pour (I.3.1), alors la limite est supportée sur l'ensemble des stratégies d'arrêt optimales pour (I.2.8). Réciproquement, tout élément de $\mathcal{P}(t, m)$ peut être approché par une suite de la forme $\mathbb{P}^0 \circ (m^N(\mathbf{Y}^N, \tau))^{-1}$; en particulier, tout élément optimal pour (I.2.8) peut s'écrire comme limite d'éléments ε_N -optimaux (avec $\varepsilon_N \searrow 0$) pour (I.3.1).

I.3.2 Cas d'une classe d'équations paraboliques

Dans ce chapitre, nous proposons une approximation pour une large classe d'équations paraboliques sur l'espace de Wasserstein, c'est-à-dire pour les équations de la forme:

$$-\partial_t u(t, \mu) - F(t, \mu, u(t, \mu), \partial_x \delta_m u(t, \mu, \cdot), \partial_{xx}^2 \delta_m u(t, \mu, \cdot)) = 0, \quad u(T, \cdot) = g, \quad (\text{I.3.3})$$

pour tout $(t, \mu) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$, avec $F : [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R} \times \mathbb{L}_2^0(\mathbb{R}^d, \mathbb{R}^d) \times \mathbb{L}_2^0(\mathbb{R}^d, \mathbb{S}_d) \rightarrow \mathbb{R}$, $\mathbb{L}_2^0(\mathbb{R}^d, \mathbb{R}^d)$ (resp. $\mathbb{L}_2^0(\mathbb{R}^d, \mathbb{R}^d)$) désignant l'ensemble des fonctions de \mathbb{R}^d dans \mathbb{R}^d (resp. \mathbb{S}_d) Borel-mesurables à croissance quadratique. La notion de solution de viscosité que nous adoptons pour cette classe d'équations est celle de Wu & Zhang [77], qui consiste à prendre

comme voisinage de viscosité l'ensemble \mathcal{P}_L des mesures sous lesquelles le processus canonique X est une semimartingale continue à caractéristiques (c'est-à-dire, coefficients de dérive et de diffusion) bornées par $L > 0$.

Nous construisons notre approximation ainsi : pour $N \geq 1$, et pour tout $\mathbf{x} = (x_1, \dots, x_N)$, $\mathbf{z} = (z_1, \dots, z_N) \in \mathbb{R}^{d \times N}$, $\boldsymbol{\gamma} = \text{Diag}(\gamma_1, \dots, \gamma_N) \in \mathbb{S}_{d \times N}^D$ et $y \in \mathbb{R}$, on définit l'opérateur

$$F^N(t, \mathbf{x}, y, \mathbf{z}, \boldsymbol{\gamma}) := F(t, \mu^N(\mathbf{x}), y, N\mathbf{z} \cdot \mathbf{1}_{\mathbf{x}}, N\boldsymbol{\gamma} \cdot \mathbf{1}_{\mathbf{x}}), \quad (\text{I.3.4})$$

avec $\mathbf{z} \cdot \mathbf{1}_{\mathbf{x}}(x) := \sum_{k=1}^N z_k \mathbf{1}_{x_k}(x)$ et $\boldsymbol{\gamma} \cdot \mathbf{1}_{\mathbf{x}}(x) := \sum_{k=1}^N \gamma_k \mathbf{1}_{x_k}(x)$ pour tout $x \in \mathbb{R}^d$. On introduit alors l'équation associée sur $[0, T] \times \mathbb{R}^{d \times N}$:

$$-\partial_t u(t, \mathbf{x}) - F^N(t, \mathbf{x}, u(t, \mathbf{x}), \partial_{\mathbf{x}} u(t, \mathbf{x}), \partial_{\mathbf{x}\mathbf{x}}^2 u(t, \mathbf{x})) = 0, \quad u(T, \cdot) = g^N. \quad (\text{I.3.5})$$

En employant une nouvelle fois les semi-limites relaxées, et en supposant que (I.3.3) dispose d'un principe de comparaison (dont la preuve reste un problème ouvert dans le cas général), nous parvenons à prouver que les solutions de viscosité de (I.3.5) converge vers la solution de viscosité de (I.3.3), dans le sens introduit dans le paragraphe précédent.

Comme dans le cas du problème de l'obstacle, le choix de la notion de solution de viscosité pour l'équation d'approximation est crucial et nous optons une nouvelle fois pour le critère de tangence en moyenne. Cependant, le caractère potentiellement non linéaire de (I.3.5) nous conduit à utiliser une espérance non linéaire, à l'image du travail d'Ekren, Keller, Touzi & Zhang [37] dans le cadre des équations aux dérivées partielles dépendant de la trajectoire (voir aussi Ren, Touzi & Zhang [67] pour un exposé général de cette théorie). Ainsi, étant donné $u : [0, T] \times \mathbb{R}^{d \times N} \rightarrow \mathbb{R}$ et $(t, \mathbf{x}) \in [0, T] \times \mathbb{R}^{d \times N}$, on définit par exemple l'ensemble des fonctions tests pour la sous-solution par

$$\underline{\mathcal{A}}u(t, \mathbf{x}) = \{ \phi \in C^{1,2}([0, T] \times \mathbb{R}^{d \times N}) : (\phi - u)(t, \mathbf{x}) = \inf_{\theta \in \mathcal{T}_{t,T}} \mathcal{E}_{t,\mathbf{x}}^N [(\phi - u)(\theta \wedge H, \mathbf{X}_{\theta \wedge H})] \},$$

où \mathbf{X} est le processus canonique sur $C^0([0, T], \mathbb{R}^{d \times N})$, $H > t$ un temps d'arrêt de localisation et $\mathcal{E}_{t,\mathbf{x}}^N$ est l'espérance non linéaire définie par

$$\mathcal{E}_{t,\mathbf{x}}^N[\cdot] := \inf_{\mathbb{P} \in \mathcal{P}_L^N} \mathbb{E}_{t,\mathbf{x}}^{\mathbb{P}}[\cdot],$$

avec \mathcal{P}_L^N l'ensemble de mesures semimartingales continues à caractéristiques bornées sur $C^0([0, T], \mathbb{R}^{d \times N})$.

Du fait de cette structure particulière, la preuve du théorème de convergence repose également sur un résultat-clé de précompacité, assimilable à une forme de *propagation du chaos* pour semimartingales continues à caractéristique bornées. Plus précisément, en invoquant des critères de tension de mesure et un problème martingale bien choisi, nous montrons que l'espace des mesures empiriques associées à des semimartingales continues à caractéristiques (uniformément) bornées est précompact, et que les valeurs d'adhérence appartiennent à \mathcal{P}_L . Ce résultat peut être vu comme une généralisation de Lacker [52, Theorem 5.1], qui traite le cas des diffusions soumises à des contrôles relaxés.

Étant donné que ces notions de solutions de viscosité, tant sur l'espace de Wasserstein qu'en dimension finie, ont été introduites par leurs auteurs dans le cadre d'équations dépendant de la trajectoire, nous parvenons également à prouver notre résultat de convergence dans le cadre non-markovien, c'est-à-dire quand la variable μ de (I.3.3) est une mesure sur $C^0([0, T], \mathbb{R}^d)$.

I.4 Vers une résolution numérique pour l'équation de l'obstacle en cascade

Le but de ce paragraphe est de discuter la possible résolution numérique du problème d'arrêt optimal multiple. On montre dans le Chapitre V que la fonction valeur de ce problème est caractérisée comme étant l'unique solution de viscosité de l'équation de la programmation dynamique (I.3.2). Notre idée est ainsi d'exploiter cette équation pour implémenter une résolution numérique du problème d'arrêt optimal multiple. Comme nous espérons pouvoir la résoudre pour des dimensions relative élevées, nous nous inspirons des méthodes de Sirignano & Spiliopoulos [71] et Raissi, Perdikaris & Karniadakis [66], qui approchent les solutions d'équations aux dérivées partielles par des réseaux de neurones, calibrés en utilisant les opérateurs des équations ainsi que les conditions aux bords pour définir la fonction de perte.

Soit $\hat{u}(\mathbf{w}; \cdot)$ le réseau de neurones approchant la solution u de (I.3.2), où $\mathbf{w} \in \mathbb{R}^K$ est le paramètre du réseau. Étant donné $(\mu_1, \mu_2, \mu_3) \in \mathcal{P}([0, T] \times \mathbf{S}^N) \times \mathcal{P}(\mathbf{S}^N) \times \mathcal{P}([0, T] \times \mathbb{R}^{d \times N})$,

nous définissons la fonction de perte théorique

$$L(\mathbf{w}) := \int_{[0,T] \times \mathbf{S}^N} |\mathcal{A}\hat{u}(\mathbf{w}; t, \mathbf{y})|^2 d\mu_1(t, \mathbf{y}) + \int_{\mathbf{S}^N} |\hat{u}(\mathbf{w}; T, \mathbf{y}) - \varphi(\mathbf{x})|^2 d\mu_2(\mathbf{y}) \quad (\text{I.4.1})$$

$$+ \int_{[0,T] \times (\mathbb{R}^d)^N} |\hat{u}(\mathbf{w}; t, \mathbf{x}, 0) - \varphi(\mathbf{x})|^2 d\mu_3(t, \mathbf{x}),$$

où

$$\mathcal{A}\hat{u}(\mathbf{w}; t, \mathbf{y}) := \min\{-\partial_t \hat{u}(\mathbf{w}; t, \mathbf{y}) - \mathcal{L}\hat{u}(\mathbf{w}; t, \mathbf{y}) \cdot \mathbf{i}, \hat{u}(\mathbf{w}; t, \mathbf{y}) - \max_{|\mathbf{i}'|=|\mathbf{i}|-1} \hat{u}(\mathbf{w}; t, \mathbf{x}, \mathbf{i}')\}.$$

Notre algorithme va consister à minimiser L sur l'ensemble des paramètres \mathbf{w} , via une descente de gradient stochastique. Plus précisément, il suivra les étapes suivante:

1. Simulation de points aléatoires $(t_n^1, \mathbf{y}_n^1) \in [0, T] \times \mathbf{S}^N$, $\mathbf{y}_n^2 \in \mathbf{S}^N$ and $(t_n^3, \mathbf{x}_n^3) \in [0, T] \times \mathbb{R}^{d \times N}$ selon les distributions μ_1, μ_2 and μ_3 , respectivement.
2. Calcul de l'erreur quadratique

$$\hat{L}(\mathbf{w}_n, \mathbf{q}) := (\mathcal{A}\hat{u}(\mathbf{w}_n; t_n, \mathbf{y}_n))^2 + (\hat{u}(\mathbf{w}_n; T, \mathbf{y}_n) - g^N(\mathbf{x}_n))^2 + (\hat{u}(\mathbf{w}_n; t_n, \mathbf{x}_n, 0) - g^N(\mathbf{x}_n))^2$$

3. Descente de gradient: $\mathbf{w}_{n+1} := \mathbf{w}_n - \alpha_n \partial_{\mathbf{w}} \hat{L}(\mathbf{w}_n, \mathbf{q})$, où $\alpha := \{\alpha_n\}_{n \geq 1}$ est le taux d'apprentissage,
4. Répétition jusqu'à satisfaction du critère d'arrêt.

Le critère d'arrêt peut prendre des formes diverses : il peut par exemple consister en un nombre fixé d'itérations, ou bien en une condition sur la norme du gradient de la fonction de perte.

Chapter II

Introduction (English)

II.1 Generalities on mean field control

II.1.1 Motivations

While it has been used for a long time in statistical physics, the notion of *mean field* has become inevitable in control theory since it has been independently introduced by Lasry & Lions [53] and Caines, Huang & Malhammé [47] to model decision problems in system of interacting agents. The main idea behind such models is to state that, under some symmetry assumptions and when the number of agents becomes very large, we may reduce the study of such systems to the study of one representative agent, who acts in function of the macroscopic state of the system, which is typically given by the distribution of the other agents. When any agent wants to optimize its own criterion, such problems are called *mean field games*; when a central planner controls all the interacting agents in order to optimize some criterion depending on the global system of agents, we face a *mean field control* problem. Mean field optimal stopping, which is the main subject of this thesis, belongs to this second class of problems.

Although not as extensively studied as games, mean field control has raised a growing interest over the past few years. First studied by Bensoussan, Frehse & Yam [8] and Carmona, Delarue & Lachapelle [22] in the context of a comparative study with mean field games, this class of problems has been mainly considered via two approaches: by the use of a maximum principle (see e.g. Buckdahn, Djehiche & Li [13], Andersson & Djehiche [1] or Meyer-Brandis, Oksendal & Zhou [60]) and by dynamic programming (see e.g. Laurière

& Pironneau [54], Pham & Wei [65], Bayraktar, Cosso & Pham [6], Wu & Zhang [77] or Djete, Possamai & Tan [34]). Summaries of these two methods are available in Carmona & Delarue [21, Vol. I, Ch. 6]. The dynamic programming approach mainly results in the characterization of the value function of the control problem by an Hamilton-Jacobi-Bellman equation (sometimes referred as *master equation*) on the space of probability measures.

From a probabilistic point of view, the intuition behind mean field control problems is the following: given two integers $d, N \geq 1$, a set of controls \mathcal{A} and $\mathbf{W} = (W^1, \dots, W^N)^\top$ a $d \times N$ -dimensional standard Brownian motion, we consider the interacting N -particles system in \mathbb{R}^d :

$$dX_t^{k,N,\alpha} = b(t, X_t^{k,N,\alpha}, \mu^N(\mathbf{X}_t^{N,\alpha}), \alpha_t^k)dt + \sigma(t, X_t^{k,N,\alpha}, \mu^N(\mathbf{X}_t^{N,\alpha}), \alpha_t^k)dW_t^k, \quad (\text{II.1.1})$$

for all $k = 1, \dots, N$, with $\mathbf{X}^{N,\alpha} := (X^{1,N,\alpha}, \dots, X^{N,N,\alpha})^\top$, $\alpha := (\alpha^1, \dots, \alpha^N)^\top \in \mathcal{A}^N$ the vector of controls and

$$\mu^N(\mathbf{X}_t^{N,\alpha}) := \frac{1}{N} \sum_{k=1}^N \delta_{X_t^{k,N,\alpha}}$$

the empirical distribution of the system at time t . Although each particle k is controlled by the process α^k , its dynamics also depends on the other controls $\{\alpha^j\}_{j \neq k}$ due to the interaction via the empirical distribution of the system. Taking a central planner point of view, we then consider the control problem

$$\sup_{\alpha \in \mathcal{A}^N} \mathbb{E} \left[\frac{1}{N} \sum_{k=1}^N \int_0^T f^k(r, \mathbf{X}_r^{N,\alpha}) dr + g(\mathbf{X}_T^{N,\alpha}) \right],$$

where $T > 0$ is the time horizon. If f and g can be written as functionals defined on the space of probability measures, i.e., can be written as $f^k(t, \mathbf{x}) = f(t, x_k, \mu^N(\mathbf{x}))$ and $g(\mathbf{x}) = g(\mu^N(\mathbf{x}))$, we may then formulate the mean field approximation of this problem. More precisely, when $N \rightarrow \infty$, we consider that the dynamics of a representative particle is given by

$$dX_t^\alpha = b(t, X_t^\alpha, \mathcal{L}(X_t^\alpha), \alpha_t)dt + \sigma(t, X_t^\alpha, \mathcal{L}(X_t^\alpha), \alpha_t)dW_t, \quad (\text{II.1.2})$$

where $\mathcal{L}(X_t^\alpha)$ denotes the distribution of X_T^α and W is a standard Brownian motion. The mean field control problem is then stated as

$$\sup_{\alpha \in \mathcal{A}} \mathbb{E} \left[\int_0^T f(r, X_r^\alpha, \mathcal{L}(X_r^\alpha)) dr \right] + g(\mathcal{L}(X_T^\alpha)). \quad (\text{II.1.3})$$

This approximation is mostly legitimated by *propagation of chaos* like results:

$$\frac{1}{N} \sum_{k=1}^N \delta_{\mathbf{X}^{k,N,\alpha}} \xrightarrow{N \rightarrow \infty} \mathcal{L}(X^\alpha) \text{ in distribution.}$$

These convergence problems have been extensively studied. We refer for example to Snitzman [72] or Oelschläger [63] for the case of uncontrolled diffusions, and to the recent works of Lacker [52] and Djete [33] for the case of controlled diffusions (with relaxed controls in the first one).

Besides the interest of such models to study systems of interacting particles, we observe that the scope of problems covered by the formulation (II.1.3) is more general than the one of standard optimal control, i.e. problems of the form

$$\sup_{\alpha \in \mathcal{A}} \mathbb{E} \left[\int_0^T f(r, X_r^\alpha) dr + g(X_T^\alpha) \right]. \quad (\text{II.1.4})$$

We may for instance mention the mean-variance optimization problem, given by

$$\sup_{\alpha \in \mathcal{A}} \left\{ \mathbb{E}[X_T^\alpha] - \frac{1}{2} \text{Var}(X_T^\alpha) \right\}.$$

As the variance contains the square of the mean, this problem cannot be written under the form (II.1.4), and therefore is not standard. Fischer & Livieri [42] have proposed a solution in the case where X is a geometric Brownian motion, by solving explicitly the corresponding N particles problem and computing its limit as N goes to infinity.

II.1.2 Some elements of analysis on Wasserstein space

Before stating the dynamic programming principle for the mean field control problem and the corresponding partial differential equation, it is necessary to introduce some elements of analysis on the space of probability measures. We refer to Carmona & Delarue [21, Vol. I, Ch. 5] for a more detailed presentation. Let (E, d) be a metric space, and $p \in [1, +\infty)$. We call *p-Wasserstein space* the set $\mathcal{P}_p(E)$ of probability measures of order p , endowed with the metric

$$\mathcal{W}_p(\mu, \nu) := \left(\inf_{\gamma \in \Gamma(\mu, \nu)} \int_{E \times E} (d(x, y))^p \gamma(dx, dy) \right)^{1/p} \quad \text{for all } \mu, \nu \in \mathcal{P}_p(E),$$

where $\Gamma(\mu, \nu)$ denotes the set of couplings of μ and ν , that is the set of probability measures on $E \times E$ with μ and ν as respective first and second marginal distributions. This metric is

called *p-Wasserstein distance*. If (E, d) is a Polish space (i.e. is complete and separable), then $(\mathcal{P}(E), \mathcal{W}_p)$ is a Polish space as well. We shall often use the following property, which allows to characterize the convergence in the sense of \mathcal{W}_p (see [21, Théorème 5.5]) : $\mathcal{W}_p(\mu_n, \mu) \xrightarrow[n \rightarrow \infty]{} 0$ if and only if the sequence $\{\mu_n\}_{n \geq 1}$ converges weakly to μ and is *p-uniformly integrable*, i.e. satisfies

$$\lim_{R \rightarrow \infty} \sup_{n \geq 1} \int_E (d(x_0, x))^p d\mu_n(x) = 0 \text{ for some } x_0 \in E.$$

It is also necessary to introduce a notion of derivative on the space of measures. We say that $u : \mathcal{P}_p(E) \rightarrow \mathbb{R}$ admits a *linear functional derivative* $\delta_m u$ if

$$u(\mu) - u(\nu) = \int_0^1 \int_E \delta_m u(\lambda\mu + (1-\lambda)\nu, x) d(\mu - \nu)(x) d\lambda \quad \text{for all } \mu, \nu \text{ in } \mathcal{P}_p(E).$$

We observe that this derivative is defined up to an additive constant, and that it is a function from $\mathcal{P}_p(E) \times E$ to \mathbb{R} . Furthermore, if $x \mapsto \delta_m u(\mu, x)$ is also differentiable, the derivatives $\partial_x \delta_m u(\mu, \cdot)$ and $\partial_{xx}^2 \delta_m u(\mu, \cdot)$ correspond to the Lions derivatives $\partial_\mu u(\mu, x)$ and $\partial_x \partial_\mu u(\mu, \cdot)$ (see [58]), defined in the case $p = 2$ by lifting u into a function defined on the Hilbert space of square integrable random variables, and by using the Fréchet derivative.

Let X be a continuous semimartingale taking its values in $E = \mathbb{R}^d$ and $u : [0, T] \times \mathcal{P}_p(\mathbb{R}^d) \rightarrow \mathbb{R}$ such that $\partial_t u$, $\partial_x \delta_m u$ and $\partial_{xx}^2 \delta_m u$ exist. Denote $\mu_t := \mathcal{L}(X_t)$ for all $t \in [0, T]$. Under some integrability conditions, we have the Itô's formula

$$\begin{aligned} u(t, \mu_t) &= u(0, \mu_0) + \int_0^t \partial_t u(s, \mu_s) ds \\ &\quad + \mathbb{E} \left[\int_0^t \partial_x \delta_m u(s, \mu_s, X_s) \cdot dX_s + \partial_{xx}^2 \delta_m u(s, \mu_s, X_s) : d\langle X \rangle_s \right], \end{aligned} \tag{II.1.5}$$

which plays a crucial role in the derivation of the dynamic programming equation corresponding to the mean field control problem.

II.1.3 Dynamic programming principle

As in the case of standard control, we may study mean field control problems by establishing a dynamic programming principle. This requires to change the state variable of the problem. In the standard setting (II.1.4), f and g are defined on a finite dimensional space (say \mathbb{R}^d), and therefore the \mathbb{R}^d -valued process X is the appropriate state variable. However, in the mean field control problem (II.1.3), they are defined on the space of probability measures,

which is infinite dimensional. Consequently, the appropriate state variable in this context is the distribution of the process X^α or, more precisely, in the Markovian case, the flow of marginal laws $\{\mathcal{L}(X_t^\alpha)\}_{t \in [0, T]}$. Thus, after rewriting (II.1.3) under the dynamical form

$$V(t, \mu) := \sup_{\alpha \in \mathcal{A}} \mathbb{E} \left[\int_t^T f(r, X_r^{t, \mu, \alpha}, \mathcal{L}(X_r^{t, \mu, \alpha})) dr \right] + g(\mathcal{L}(X_T^{t, \mu, \alpha})), \quad (t, \mu) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d),$$

where $X^{t, \mu, \alpha}$ corresponds to the dynamics (II.1.2) conditioned to $\mathcal{L}(X_t^\alpha) = \mu$, we can prove that

$$V(t, \mu) = \sup_{\alpha \in \mathcal{A}} \mathbb{E} \left[\int_t^s f(r, X_r^{t, \mu, \alpha}, \mathcal{L}(X_r^{t, \mu, \alpha})) dr \right] + V(s, \mathcal{L}(X_s^{t, \mu, \alpha})) \text{ for all } s \in [t, T].$$

If V is sufficiently smooth, we may then combine this equality with Itô's formula (II.1.5) to prove that V is a solution of

$$-\partial_t u(t, \mu) + H(t, \mu, \delta_m u(t, \mu, \cdot)) = 0, \quad u(T, \mu) = g(\mu), \quad (\text{II.1.6})$$

for all $(t, \mu) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$, with

$$H(t, \mu, \varphi) := \int_E \sup_{a \in A} \left(b(t, x, \mu, a) \cdot \partial_x \varphi(x) + \frac{1}{2} \sigma^2(t, x, \mu, a) : \partial_{xx}^2 \varphi(x) \right) \mu(dx)$$

for all $\varphi \in C^2(\mathbb{R}^d)$. It is also possible to derive (II.1.6) when V is not differentiable by using viscosity solutions (see for example Wu & Zhang [77]) and Pham & Wei [65], who use two different notions of viscosity solutions).

In order to fully characterize the problem (II.1.3) by the equation (II.1.6), it is also necessary to prove a uniqueness result, called *comparison principle* in the literature of viscosity solutions. However, this result is difficult to show in the general case, and we refer for example to [77, Théorème 4.13] for a proof in a particular case.

II.2 The case of optimal stopping

Optimal stopping problems correspond to binary decision problems: at each time, the decision to take consist in stopping or not a given process, based on the information available up to this time only and without possibility to come back. This can be illustrated by the secretary problem: an employer wants to hire a secretary and organize a sequence of interviews with the applicants. After each interview, he must decide whether he hires or not the candidate.

The main purpose of this thesis is to establish a dynamic programming approach for the mean field optimal stopping problem, which is a special subclass of mean field control problems. As for the latter, we take the point of view of a central planner, who controls a distribution of interacting particle. In this context, one has to choose for each particle a stopping time in order to optimize some functional of the distribution of the system.

The same way mean field control differs from mean field games, the mean field optimal stopping problem differs from mean field games of timing, which have been extensively studied recently (we refer, for example, to Bertucci [9], Carmona, Delarue & Lacker [23], Nutz [62] and Bouveret, Dumitrescu & Tankov [11]).

II.2.1 Reminder on standard optimal stopping, and particularities of mean field optimal stopping

We first recall some generalities on standard optimal stopping (we refer to El Karoui [41] or Shiryaev [70] for detailed surveys). Let X be a diffusion process

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \quad (\text{II.2.1})$$

\mathbb{F} the filtration generated by the Brownian motion W and, for all $t \in [0, T]$, let $\mathcal{T}_{t,T}$ be the set of $[t, T]$ -valued \mathbb{F} -stopping times. We consider the optimization problem

$$\sup_{\tau \in \mathcal{T}_{0,T}} \mathbb{E} \left[\int_0^\tau f(r, X_r)dr + g(X_\tau) \right],$$

which can be rewritten under the dynamical form

$$V(t, x) := \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E} \left[\int_t^\tau f(r, X_r)dr + g(X_\tau) | X_t = x \right]. \quad (\text{II.2.2})$$

If V is continuous, we may show that it satisfies the dynamic programming principle

$$V(t, x) = \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E} \left[\int_t^{\tau \wedge \theta} f(r, X_r)dr + g(X_\tau) \mathbf{1}_{\theta \geq \tau} + V(\theta, X_\theta) \mathbf{1}_{\theta < \tau} | X_t = x \right],$$

which allows to deduce, under some growth assumptions, that V is the unique viscosity solution of the obstacle problem

$$\min \left\{ -\partial_t u - b\partial_x u - \frac{1}{2}\sigma^2\partial_{xx}^2 u - f, u - g \right\} = 0, \quad u(T, \cdot) = g. \quad (\text{II.2.3})$$

By using the equation in the case where $V \in C^{1,2}([0, T] \times \mathbb{R}^d)$, or the Snell envelope theory in the general case (see e.g. Karatzas & Shreve [49, Appendice D]), we may then show that the stopping time

$$\tau^* := \inf\{s \geq t : V(s, X_s) = g(X_s)\}$$

is optimal for the problem (II.2.2). In finance, this type of problem arises naturally in the context of American options, which allow their buyer to exercise them at any time between the subscription and the maturity. Since the stock X on which the option is defined is stochastic, the buyer (who cannot predict its evolution) has to solve an optimal stopping problem to find the best time to exercise the option. Similarly, to hedge himself, the seller prices the optimal by computing the expectation of his *payoff* $g(X_\tau)$ under the optimal stopping strategy.

As in the case of stochastic control, and in order to provide a better understanding of the mean field model, we first present the N particles optimal stopping problem, called *multiple optimal stopping*. This problem has been studied in a very general context by Kobylanski, Quenez & Rouy-Mironescu [50]. We once again take the point of view of a central planner who wants to optimize a criterion by controlling all the particles of the system. The choice of the control for one particle impacts the dynamics of the other, due to the interaction term. In this setting, controls are stopping time, and the only way to highlight this impact is to make them intrinsic to the dynamics of the system, which is a first difference with the case of standard optimal stopping, where the dynamics (II.2.1) remains uncontrolled. We thus state the problem as follows: given a sequence of mutually independent Brownian motions $\{W^k\}_{k \geq 1}$, \mathbb{F}^N the filtration generated by $\{W^k\}_{1 \leq k \leq N}$, $\mathcal{T}_{0,T}^N$ the set of N -uplets of \mathbb{F}^N -stopping times and $\boldsymbol{\tau} := (\tau_1, \dots, \tau_N) \in \mathcal{T}_{0,T}^N$, the particles follow the dynamics

$$dX_t^{k,N,\boldsymbol{\tau}} = b(t, X_t^{k,N,\boldsymbol{\tau}}, \mu^N(\mathbf{X}_t^{N,\boldsymbol{\tau}}))i_k \mathbf{1}_{t < \tau_k} dt + \sigma(t, X_t^{k,N,\boldsymbol{\tau}}, \mu^N(\mathbf{X}_t^{N,\boldsymbol{\tau}}))i_k \mathbf{1}_{t < \tau_k} dW_t^k \quad (\text{II.2.4})$$

for all $k = 1, \dots, N$, with $\mathbf{X}^N = (X^{1,N,\boldsymbol{\tau}}, \dots, X^{N,N,\boldsymbol{\tau}})^\top$ and $\mathbf{i} := (i_1, \dots, i_N)^\top \in \{0, 1\}^N$. This equation corresponds to the controlled diffusion (II.1.1) in the case where b and σ are linear in α , for

$$\boldsymbol{\alpha}_t = \mathbf{I}_t := (I_t^1, \dots, I_t^N)^\top := (i_1 \mathbf{1}_{t < \tau_1}, \dots, i_N \mathbf{1}_{t < \tau_N})^\top.$$

The vector \mathbf{i} allows to signal the particles which are stopped from the initial time, and turns out to be crucial to derive the dynamic programming equation characterizing the multiple

optimal stopping problem, which writes as follows:

$$\sup_{(\tau_1, \dots, \tau_N) \in \mathcal{T}_{0,T}^N} \mathbb{E} \left[\frac{1}{N} \sum_{k=1}^N \int_0^{\tau_k} f^{k,N}(r, \mathbf{X}_r^{N,\tau}) dr + g^N(X_{\tau_1}^{1,N,\tau}, \dots, X_{\tau_N}^{N,N,\tau}) \right], \quad (\text{II.2.5})$$

Observe that, as $X_{\tau_k}^{k,N,\tau} = X_T^{k,N,\tau}$ for all k , and that τ and \mathbf{I} are in bijection, this problem is equivalent to a control problem in \mathbf{I} . Nonetheless, given the monotonicity constraint on \mathbf{I} (which is decreasing), the proof of a dynamic programming principle is not as straightforward as in the case of standard control. We study the case of multiple optimal stopping in the Chapter V of this thesis, which corresponds to [76].

If $f^{k,N}(\cdot, \mathbf{x}) = f(\cdot, x_k, \mu^N(\mathbf{x}))$ for all $k \in \{1, \dots, N\}$ and $g^N(\mathbf{x}) = g(\mu^N(\mathbf{x}))$, then the multiple optimal stopping problem has an asymptotical formulation, that is, the *mean field optimal stopping* problem:

$$\sup_{\tau \in \mathcal{T}_{0,T}} \mathbb{E} \left[\int_0^\tau f(r, X_r, \mathcal{L}(X_r)) dr \right] + g(\mathcal{L}(X_\tau)), \quad (\text{II.2.6})$$

where X is a stopped McKean-Vlasov diffusion:

$$dX_t = b(t, X_t, \mathcal{L}(X_t))I_t dt + \sigma(t, X_t, \mathcal{L}(X_t))I_t dW_t, \quad \text{with } I_t := I_{0-} \mathbf{1}_{t < \tau}, \quad (\text{II.2.7})$$

and I_{0-} is a $\{0, 1\}$ -random variable. As in the case of multiple optimal stopping, it is a mean field control problem in I . However, as said above, the monotonicity property inherent to this class of controls makes the proof of a dynamic programming principle less straightforward. In particular, it is not sufficient to take $\mathcal{L}(X_t)$ as state variable, unlike the case of general mean field control.

Some examples of such optimal stopping problems (with a nonstandard criterion depending only on the law of the stopped process) have been studied in the literature: we may thus mention the case of mean-variance optimal stopping by Pedersen & Peskir [64] and of the probability distortion by Xu & Zhou [78], where the criterion g is defined by

$$g(\mu) := \int_0^\infty \varphi(\mathbb{P}(U(\xi) \geq x)) dx,$$

where μ is a probability measure on $(0, +\infty)$, ξ is a random variable with distribution μ under some fixed probability measure \mathbb{P} , U a utility function and φ a continuous function from $[0, 1]$ into itself called distortion function. These two examples are studied in infinite horizon ($T = \infty$), when X is a geometric Brownian motion starting from a Dirac distribution.

II.2.2 Obstacle problem on Wasserstein space

The first objective of this thesis is to establish a dynamic programming principle (DPP) to characterize the value function of the problem. This task corresponds to Chapter III and [74].

A natural idea is to introduce a dynamical version of the problem, as for the mean field control problem (II.1.3). Let $(t, \mu) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$, and

$$\begin{aligned} V(t, \mu) &:= \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E} \left[\int_t^\tau f(r, X_r, \mathcal{L}(X_r)) dr \right] + g(\mathcal{L}(X_\tau)) \\ &= \sup_{I \in \mathbb{I}^0([t, T])} \mathbb{E} \left[\int_0^T f(r, X_r, \mathcal{L}(X_r)) I_r dr \right] + g(\mathcal{L}(X_T)), \end{aligned}$$

where X is a stopped diffusion (II.2.7) with survival process $I_s := \mathbf{1}_{s < \tau}$ such that $\mathcal{L}(X_t) = \mu$, and where $\mathbb{I}^0([t, T])$ is the set of survival processes (i.e., the sets of processes of the form $s \mapsto \mathbf{1}_{s < \theta}$, $\theta \in \mathcal{T}_{t,T}$). However, it turns out that this formulation is not sufficient to obtain a DPP. Indeed, the particles stopped on the interval $[t, t + \delta]$, $\delta \in (0, T - t)$, are still stopped on the interval $[t + \delta, T]$. Any control I which is admissible on $[t + \delta, T]$ must then satisfy $I_{t+\delta} \leq I_{(t+\delta)-}^*$, where I^* is an optimal control on $[t, t + \delta]$. The impossibility to prove a DPP with this formulation is then caused by the monotonicity of the elements of $\mathbb{I}^0([t, T])$, which are non-increasing by definition. Then, to take this into account, we choose $\mathcal{L}((X_t, I_t))$ for state variable instead of $\mathcal{L}(X_t)$. Intuitively, this means that the information which consists in knowing the positions of the particles in \mathbb{R}^d (the distribution $\mathcal{L}(X_t)$) is not sufficient: we also need to know their state, that is, whether they are stopped or not (such information is encapsulated in the joint distribution $\mathcal{L}((X_t, I_t))$).

We then use the following setting: $\Omega := C^0([0, T], \mathbb{R}^d) \times \mathbb{I}^0([-1, T])$, where $C^0([0, T], \mathbb{R}^d)$ denotes the set of continuous paths from $[0, T]$ to \mathbb{R}^d , is the canonical space, and $Y := (X, I)$ is the canonical process on Ω . For all $(t, m) \in [0, T] \times \mathcal{P}_2(\mathbf{S})$, $\mathbf{S} := \mathbb{R}^d \times \{0, 1\}$, we then denotes by $\mathcal{P}(t, m)$ the set of probability measures on Ω such that X follows the dynamics (II.2.7), \mathbb{P} -almost surely, and $\mathbb{P} \circ Y_t^{-1} = m$. The mean field optimal stopping problem is then defined by

$$V(t, m) := \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \mathbb{E} \left[\int_t^T f(r, X_r, \mathbb{P} \circ X_r^{-1}) dr \right] + g(\mathbb{P} \circ X_T^{-1}). \quad (\text{II.2.8})$$

This formulation, called *weak* (the process is fixed and we control its distribution), allows in particular to compactify the set of controls by also taking into account *randomized* stopping

strategies, which consist, at any time, in being able to stop X with some probability $p \in (0, 1)$. Thus, if f and g are upper semicontinuous, then the problem (II.2.8) admits a solution. We also observe that this problem is indeed an optimal stopping problem by defining $\tau := \inf\{s \geq t : I_s = 0\}$ and observing that, when τ goes through the set of (randomized) $[t, T]$ -valued stopping times, then \mathbb{P} goes through $\mathcal{P}(t, m)$.

Once this setting is defined, we are able to prove the DPP:

$$V(t, m) = \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \mathbb{E} \left[\int_t^s f(r, X_r, \mathbb{P}_{Y_r}) dr \right] + V(s, \mathbb{P}_{Y_{s-}}) \text{ for all } s \in [t, T]. \quad (\text{II.2.9})$$

As often in stochastic control, we derive the dynamic programming equation corresponding to our optimization problem by applying Itô's formula to the DPP. However, in our context, the process Y is a càdlàg semimartingale (thus possibly non continuous), which prevent us from using (II.1.5). Furthermore, its flow of marginal laws $\{\mathbb{P}_{Y_s}\}_{s \in [t, T]}$ is itself càdlàg, which also bans the use of some Itô's formula for laws of jump-diffusions available in the literature (see e.g. Burzoni, Ignazio, Reppen & Soner [15]). We then have to prove an Itô's formula for the general case of càdlàg semimartingale Y . If $\mathbf{m} := \{m_t\}_{t \in [0, T]}$ denotes its flow of marginal laws, and $J_{\mathbb{T}}(\mathbf{m})$ the set of discontinuity points of $t \mapsto m_t$ on $\mathbb{T} \subset [0, T]$, then, under some integrability conditions:

$$\begin{aligned} u(t, m_t) &= u(0, m_0) + \int_0^t \partial_t u(s, m_s) ds \\ &+ \mathbb{E} \left[\int_0^t \partial_y \delta_m u(s, m_s, Y_s) \cdot dY_s^c + \partial_{yy}^2 \delta_m u(s, m_s, Y_s) : d\langle Y^c \rangle_s \right] \\ &+ \sum_{s \in J_{(0, t]}(\mathbf{m})} (u(s, m_s) - u(s, m_{s-})) + \mathbb{E} \left[\sum_{s \in J_{(0, t]}^c(\mathbf{m})} (\delta_m u(s, m_s, Y_s) - \delta_m u(s, m_s, Y_{s-})) \right], \end{aligned} \quad (\text{II.2.10})$$

where Y^c denotes the continuous part of Y . Nonetheless, we are only interested in the particular case $Y = (X, I)$ in the context of mean field optimal stopping.

Applying this formula to the DPP (II.2.9) proves that, if V is sufficiently smooth, then it is a solution of the dynamic programming equation:

$$\min_{m' \in C_u(t, m)} \{- (\mathbb{L}u + F)(t, m')\} = 0, \quad D_I u(t, m, \cdot) \geq 0, \quad u(T, \cdot) = g, \quad (\text{II.2.11})$$

for all $(t, m) \in [0, T] \times \mathcal{P}_2(\mathbf{S})$, with

$$\begin{aligned} \mathbb{L}u(t, m) &:= \partial_t u(t, m) \\ &+ \int_{\mathbb{R}^d} (b(t, x, m) \cdot \partial_x \delta_m u(t, m, x, 1) + \frac{1}{2} \sigma^2(t, x, m) : \partial_{xx}^2 \delta_m u(t, m, x, 1)) m(dx, 1), \\ F(t, m) &:= \int_{\mathbb{R}^d} f(t, x, m) m(dx, 1) \\ C_u(t, m) &:= \{m' \preceq m : u(t, m') = u(t, m)\}, \\ D_I u(t, m, x) &:= \delta_m u(t, m, x, 1) - \delta_m u(t, m, x, 0) \quad \text{pour tout } x \in \mathbb{R}^d. \end{aligned}$$

The notation \preceq denotes a partial order on $\mathcal{P}_2(\mathbf{S})$: we have $m' \preceq m$ if there exists a measurable function $p : \mathbb{R}^d \rightarrow [0, 1]$ such that

$$m'(dx, 1) = p(x)m(dx, 1) \text{ and } m'(dx, 0) = (1 - p(x))m(dx, 1) + m(dx, 0).$$

Intuitively, this means that m' can be obtained by stopping each unstopped particle of the distribution m localized in position x with probability $p(x)$. $C_u(t, m)$ thus denotes the set of stopping strategies at time t which preserve the value function. The condition $D_I u \geq 0$, which implies that u is non-decreasing for the partial order \preceq , expresses the natural monotonicity of the problem: the value function is even greater that the set of particles that we control is.

By analogy with standard optimal stopping, we call this equation *obstacle problem on Wasserstein space*. In addition, we prove a verification theorem, i.e. that any sufficiently smooth solution of (II.2.11) is equal to the value function (II.2.8). We also use this equation to characterize optimal stopping strategies.

II.2.3 Characterization by viscosity solutions

In the second part of our work, we extend the characterization of the mean field optimal stopping problem (II.2.8) by the obstacle equation (II.2.11) to the case where V is not differentiable. In other words, we shall define a notion of weak solution for (II.2.11), in the sense of viscosity solutions. Introduced by Crandall & Lions [30] in the context of Hamilton-Jacobi equations, this notion allows to define weak solutions for second order elliptic partial differential equations, in particular nonlinear. Unlike Sobolev solutions, viscosity solutions are defined pointwise. They allow to obtain existence and uniqueness results under minimal regularity assumptions (typically, for locally bounded functions). We refer to the survey of

Crandall, Ishii & Lions [29] for more details about viscosity solutions on finite dimensional spaces.

Motivated in particular by mean field games and control, the study of viscosity solutions for equations on the Wasserstein space has raised a lot of interest over the past few years and has been realized via very diverse approaches: by the use of subdifferentials (Cardaliaguet & Quincampoix [18], Gangbo, Nguyen & Tudorascu [45]), in lifting the equation on the Hilbert space of square integrable random variables (Pham & Wei [65]) or by using directly the derivatives on the space of measures and by constructing compact viscosity neighborhoods in the Wasserstein space (Wu & Zhang [77], Burzoni, Ignazio, Reppen & Soner [15]).

In this thesis, we mainly follow the approach of [77], where the authors define a notion of viscosity solution for a general class of parabolic equations on Wasserstein space. This notion is intrinsic: unlike [65], which resorts to the *lifting* method, this does not require any change of space. Its main advantage lies in the choice of the viscosity neighborhood, whose compactness is crucial to ensure the stability and uniqueness properties. As the unit ball on $\mathcal{P}_2(\mathbb{R}^d)$ is not compact, the authors decide to consider the set of flows of marginal laws of continuous semimartingales with bounded characteristic, starting from a point $(t, \mu) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$. Then, simple tightness arguments prove that such a set is compact (see e.g. Meyer & Zheng [59]).

In the context of mean field optimal stopping, we thus consider the set of marginal laws of stopped McKean-Vlasov diffusions (II.2.7). More precisely, given $(t, m) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$ and $\delta \in (0, T - t)$, we define the viscosity neighborhood:

$$\mathcal{N}_\delta(t, m) := \{(s, \tilde{m}) : s \in [t, t + \delta], \tilde{m} \in \{\mathbb{P}_{Y_{s-}}, \mathbb{P}_{Y_s}\}, \mathbb{P} \in \mathcal{P}(t, m)\}.$$

Since the flow $s \mapsto \mathbb{P}_{Y_s}$ is càdlàg for all $\mathbb{P} \in \mathcal{P}(t, m)$, it is indeed necessary to take into account all pairs $\{\mathbb{P}_{Y_{s-}}, \mathbb{P}_{Y_s}\}$ to ensure the compactness of $\mathcal{N}_\delta(t, m)$. Given a locally bounded function $u : [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$, we then define the sets of test functions:

$$\begin{aligned} \overline{\mathcal{A}}u(t, m) &:= \left\{ \varphi \in C_2^{1,1}(\mathbf{Q}_t) : (\varphi - u_*)(t, m) = \max_{\mathcal{N}_\delta(t, m)} (\varphi - u_*) \text{ for some } \delta > 0 \right\}, \\ \underline{\mathcal{A}}u(t, m) &:= \left\{ \varphi \in C_2^{1,1}(\mathbf{Q}_t) : (\varphi - u^*)(t, m) = \min_{\mathcal{N}_\delta(t, m)} (\varphi - u^*) \text{ for some } \delta > 0 \right\}, \end{aligned}$$

where u_* and u^* respectively denote the lower and upper semicontinuous envelopes of u . Given the particular structure of the equation (II.2.11), the definition of viscosity solutions

is however less standard than usual. Indeed, for a usual equation

$$\mathcal{D}u = 0,$$

where \mathcal{D} denotes some differential operator, the viscosity supersolutions and subsolutions are naturally defined in any point (t, m) by

$$\mathcal{D}\varphi(t, m) \geq 0 \quad \text{and} \quad \mathcal{D}\psi(t, m) \leq 0$$

for all $\varphi \in \overline{\mathcal{A}u}(t, m)$ and $\psi \in \underline{\mathcal{A}u}(t, m)$. The presence of the inequality $D_I u \geq 0$ in (II.2.11) implies some asymmetry between our definitions of viscosity supersolutions and subsolutions. Thus, we define:

(i) u is a viscosity supersolution of (II.2.11) if, for all $(t, m) \in [0, T) \times \mathcal{P}_2(\mathbb{R}^d)$, $\varphi \in \overline{\mathcal{A}u}(t, m)$ and $m' \preceq m$,

$$u_*(t, m) \geq u_*(t, m') \quad \text{and} \quad -(\mathbb{L}\varphi + F)(t, m) \geq 0,$$

(ii) u is a viscosity subsolution of (II.2.11) if, for all $(t, m) \in [0, T) \times \mathcal{P}_2(\mathbb{R}^d)$ such that $C_{u^*}(t, m) = \{m\}$ and $\varphi \in \underline{\mathcal{A}u}(t, m)$,

$$\min\{-(\mathbb{L}\varphi + F), (\mathbb{D}_I \varphi)_*\}(t, m) \leq 0,$$

with $\mathbb{D}_I \varphi(t, m) := \inf_{x \in \text{Supp}(m(\cdot, 1))} D_I u(t, m, x)$, where $\text{Supp}(m(\cdot, 1))$ denotes the support of $m(\cdot, 1)$.

(iii) u is a viscosity solution of (II.2.11) if it is a viscosity supersolution and subsolution.

Our viscosity solutions are then consistent with classical solutions and satisfy the stability property. The uniqueness result, implied by a comparison principle, is as usual the most difficult to prove, even more that we consider equations on the space of probability measures. We manage to prove this important result by strengthening the regularity assumptions on the coefficients b and σ and by relying on a smooth mollifier for functions defined on Wasserstein space, introduced by Mou & Zhang [61].

II.3 Approximations of partial differential equations on Wasserstein space

The second part of this thesis is dedicated to the development of approximations of equations on the space of probability measures. More precisely, our goal is to approximate viscosity

solutions of such equations by viscosity solutions of equations defined on finite dimensional spaces. A related work has been realized by Gangbo, Mayora & Swiech [44] for the semi-linear Hamilton-Jacobi-Bellman equation (i.e. with uncontrolled diffusion coefficient), and with a notion of viscosity solutions different from ours. We may also refer to Cardaliaguet, Delarue, Lasry & Lions [17], who show the convergence of the solution of the equation characterizing N -players games to the solution of the equation characterizing the Nash equilibrium of mean field games.

In the spirit of Barles & Souganidis [2], who prove the convergence of a numerical scheme (called *monotone scheme*) for viscosity solutions of Hamilton-Jacobi-Bellman equations, we study the convergence of our finite-dimensional approximation to the solution of the mean field equation via the relaxed semi-limits. This approach is combined with the use of *propagation of chaos* like results (see e.g. Snitzman [72] ou Oelschläger [63]), which allow to connect the N particles setting to the mean field one.

II.3.1 Case of the obstacle problem

Given its particular structure, we study the case of the obstacle equation aside. We mainly show that the approximating equation corresponds to the dynamic programming equation characterizing the multiple optimal stopping problem (II.2.5).

More precisely, given a probability space $(\Omega^0, \mathcal{F}_T^0, \mathbb{P}^0)$, we introduce the dynamical formulation

$$V^N(t, \mathbf{y}) := \sup_{\tau \in \mathcal{T}_{t,T}^N} \mathbb{E}_{t,\mathbf{y}}^{\mathbb{P}^0} \left[\frac{1}{N} \sum_{k=1}^N \int_0^{\tau_k} f^{k,N}(r, \mathbf{X}_r^{N,\tau}) dr + g^N(X_{\tau_1}^{1,N,\tau}, \dots, X_{\tau_N}^{N,N,\tau}) \right] \quad (\text{II.3.1})$$

for all $\mathbf{y} \in \mathbf{S}^N$, where $\mathbf{Y}^{N,\tau} := (\mathbf{X}^{N,\tau}, \mathbf{I}^{N,\tau})$ follows the dynamics (II.2.4) and $\mathbb{E}_{t,\mathbf{y}}[\cdot] := \mathbb{E}[\cdot | Y_t^{N,\tau} = \mathbf{y}]$. As proved by Kobylanski, Quenez & Rouy-Mironescu [50], this problem can be reduced to a sequence of N interlocked standard optimal stopping problems, which allows us to prove that (II.3.1) is characterized by the equation

$$\begin{cases} \min \left\{ -\partial_t u(t, \mathbf{y}) - (\mathcal{L}u + \mathbf{f})(t, \mathbf{y}) \cdot \mathbf{i}, u(t, \mathbf{y}) - \max_{\mathbf{i}' < \mathbf{i}} u(t, \mathbf{x}, \mathbf{i}') \right\} = 0, \\ u|_{t=T} = u|_{\mathbf{i}=0} = g^N, \end{cases} \quad (\text{II.3.2})$$

where $\mathbf{y} = (\mathbf{x}, \mathbf{i})$, and $\mathbf{i}' < \mathbf{i}$ means that $i'_k \leq i_k$ for all $k = 1, \dots, N$ and $\mathbf{i}' \neq \mathbf{i}$, with $\mathbf{i} := (i_1, \dots, i_N)^\top$, $\mathbf{i}' := (i'_1, \dots, i'_N)^\top$, $\mathbf{f} := (f^{1,N}, \dots, f^{N,N})^\top$. We call this equation *cascade*

obstacle problem. We show how to use the solution of this equation to identify an optimal stopping strategy for (II.3.1).

In the case where $f^{k,N}(\cdot, \mathbf{x}) = f(\cdot, x_k, \mu^N(\mathbf{x}))$ for all $k \in \{1, \dots, N\}$ and $g^N(\mathbf{x}) = g(\mu^N(\mathbf{x}))$, we show that any viscosity supersolution or subsolution of (II.3.2) converges (in some sense) to the viscosity supersolution or subsolution of (II.2.11). If the assumptions ensuring the uniqueness of the solution of (II.2.11) are satisfied, this result implies that the function V^N converges to V , the value function of the mean field optimal stopping problem defined by (II.2.8), that is:

$$V^N(s, \mathbf{y}^N) \xrightarrow[\substack{N \rightarrow \infty, s \rightarrow t, \\ w_2(m^N(\mathbf{y}^N), m) \rightarrow 0}]{} V(t, m) \quad \text{for all } (t, m) \in [0, T] \times \mathcal{P}_2(\mathbf{S}).$$

The choice of the notion of viscosity solution for the cascade obstacle problem turns out to be particularly important to ensure this convergence result. Indeed, it has to be consistent with our notion of viscosity solution for the obstacle problem on Wasserstein space in the sense where, if φ is a test function for the latter, then its projection on \mathbf{S}^N defined by $\phi^N(t, \mathbf{y}) := \varphi(t, m^N(\mathbf{y}))$ must be a test function for the finite dimensional problem. It then turns out that the good definition is the one given by the "tangency through the mean" criterion : given $u : [0, T] \times \mathbf{S}^N \rightarrow \mathbb{R}$ and $(t, \mathbf{y}) \in [0, T] \times \mathbf{S}^N$, we define for example the set of test functions for the supersolution by

$$\bar{\mathcal{A}}u(t, \mathbf{y}) = \left\{ \phi \in C^{1,2}([t, T] \times \mathbf{S}) : (\phi - u)(t, \mathbf{y}) = \sup_{\theta \in \mathcal{T}_{t,T}} \mathbb{E}_{t, \mathbf{y}} [(\phi - u)(\theta \wedge H, \bar{Y}_{\theta \wedge H}^N)] \right\},$$

where \bar{Y} denotes the unstopped version of (II.2.4) and $H > t$ is a localization stopping time. This definition is an adaptation to the case of the multiple optimal stopping of the definition introduced by Ekren [36] in the context of path-dependent optimal stopping.

We also show in this chapter that the N particles stopped diffusion (II.2.4) converges to the stopped McKean-Vlasov diffusion (II.2.7) in the following sense: for all $\tau \in \mathcal{T}_{t,T}^N$, after possibly passing to a subsequence, the sequence $\{\mathbb{P}^0 \circ (m^N(\mathbf{Y}^N, \tau))^{-1}\}_{N \geq 1}$ converges in distribution to a measure supported on $\mathcal{P}(t, m)$. In particular, if τ^N is an optimal stopping strategy for (II.3.1), then the limit is supported on the set of optimal stopping strategies for (II.2.8). Conversely, any element of $\mathcal{P}(t, m)$ can be approximated by a sequence of the form $\mathbb{P}^0 \circ (m^N(\mathbf{Y}^N, \tau))^{-1}$. In particular, any element that is optimal for (II.2.8) can be written as the limit of ε_N -optimal controls (with $\varepsilon_N \searrow 0$) for (II.3.1).

II.3.2 Case of a class of parabolic equations

In this chapter, we propose an approximation for a large class of parabolic equations on Wasserstein space, i.e. for the equations of the form

$$-\partial_t u(t, \mu) - F(t, \mu, u(t, \mu), \partial_x \delta_m u(t, \mu, \cdot), \partial_{xx}^2 \delta_m u(t, \mu, \cdot)) = 0, \quad u(T, \cdot) = g, \quad (\text{II.3.3})$$

for all $(t, \mu) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d)$, with $F : [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R} \times \mathbb{L}_2^0(\mathbb{R}^d, \mathbb{R}^d) \times \mathbb{L}_2^0(\mathbb{R}^d, \mathbb{S}_d) \rightarrow \mathbb{R}$, where $\mathbb{L}_2^0(\mathbb{R}^d, \mathbb{R}^d)$ (resp. $\mathbb{L}_2^0(\mathbb{R}^d, \mathbb{S}_d)$) denotes the set of functions from \mathbb{R}^d to \mathbb{R}^d (resp. \mathbb{S}_d) Borel-measurables with quadratic growth. The notion of viscosity solutions that we use for this class of equations is the one of Wu & Zhang [77], which consists in taking for viscosity neighborhood the set \mathcal{P}_L of measures under which the canonical process X is a continuous semimartingale with characteristics (i.e., drift and diffusion coefficients) bounded by $L > 0$.

We build our approximation as follows: for $N \geq 1$, and for all $\mathbf{x} = (x_1, \dots, x_N)$, $\mathbf{z} = (z_1, \dots, z_N) \in \mathbb{R}^{d \times N}$, $\gamma = \text{Diag}(\gamma_1, \dots, \gamma_N) \in \mathbb{S}_{d \times N}^D$ and $y \in \mathbb{R}$, we define the operator

$$F^N(t, \mathbf{x}, y, \mathbf{z}, \gamma) := F(t, \mu^N(\mathbf{x}), y, N\mathbf{z} \cdot \mathbf{1}_{\mathbf{x}}, N\gamma \cdot \mathbf{1}_{\mathbf{x}}), \quad (\text{II.3.4})$$

with $\mathbf{z} \cdot \mathbf{1}_{\mathbf{x}}(x) := \sum_{k=1}^N z_k \mathbf{1}_{x_k}(x)$ and $\gamma \cdot \mathbf{1}_{\mathbf{x}}(x) := \sum_{k=1}^N \gamma_k \mathbf{1}_{x_k}(x)$ for all $x \in \mathbb{R}^d$. We then introduce the corresponding equation on $[0, T] \times \mathbb{R}^{d \times N}$:

$$-\partial_t u(t, \mathbf{x}) - F^N(t, \mathbf{x}, u(t, \mathbf{x}), \partial_{\mathbf{x}} u(t, \mathbf{x}), \partial_{\mathbf{xx}}^2 u(t, \mathbf{x})) = 0, \quad u(T, \cdot) = g^N. \quad (\text{II.3.5})$$

By using once again relaxed semi-limits, and assuming that we have a comparison principle for (II.3.3) (whose proof remains an open problem in the general case), we manage to prove that the viscosity solutions of (II.3.5) converge to the viscosity solution of (II.3.3), in the sense introduced in the previous paragraph.

As in the case of the obstacle equation, the choice of the notion of viscosity solution for the approximating equation is crucial and we resort once again to the tangency through the mean criterion. However, the possible nonlinearity of (II.3.5) lead us to use a nonlinear expectation, similarly to the work of Ekren, Keller, Touzi & Zhang [37] in the context of path-dependent partial differential equations (see also Ren, Touzi & Zhang [67] for a general survey of this theory). Thus, given $u : [0, T] \times \mathbb{R}^{d \times N} \rightarrow \mathbb{R}$ and $(t, \mathbf{x}) \in [0, T] \times \mathbb{R}^{d \times N}$, we define for example the set of test functions for the subsolution by

$$\underline{\mathcal{A}}u(t, \mathbf{x}) = \left\{ \phi \in C^{1,2}([0, T] \times \mathbb{R}^{d \times N}) : (\phi - u)(t, \mathbf{x}) = \inf_{\theta \in \mathcal{T}_{t,T}} \underline{\mathcal{E}}_{t,\mathbf{x}}^N [(\phi - u)(\theta \wedge H, \mathbf{X}_{\theta \wedge H})] \right\},$$

where \mathbf{X} is the canonical process on $C^0([0, T], \mathbb{R}^{d \times N})$, $H > t$ is a localization stopping time and $\mathcal{E}_{t, \mathbf{x}}^N$ is the nonlinear expectation defined by

$$\mathcal{E}_{t, \mathbf{x}}^N[\cdot] := \inf_{\mathbb{P} \in \mathcal{P}_L^N} \mathbb{E}_{t, \mathbf{x}}^{\mathbb{P}}[\cdot],$$

with \mathcal{P}_L^N the set measures such that \mathbf{X} is a $\mathbb{R}^{d \times N}$ -valued continuous semimartingale with characteristics bounded by L .

Due to this particular structure, the proof of the convergence theorem also relies on a key precompactness result, similar to a form of *propagation of chaos* for continuous semimartingales with uniformly bounded characteristics. More precisely, by using tightness arguments and a well-chosen martingale problem, we show that the space of empirical measures associated with such martingales is precompact, and that its accumulation points belong to \mathcal{P}_L . This result may be seen as generalization of Lacker [52, Theorem 5.1], who studies the case of diffusions conducted by relaxed controls.

Given that these notions of viscosity solutions, whether it is on Wasserstein space or in finite dimension, have been introduced by their authors in the context of path-dependent equations, we also manage to prove our convergence result in the non-Markovian case, that is, when the variable μ in `eqrefIntro-parPDE-en` is a measure on the set of continuous paths $C^0([0, T], \mathbb{R}^d)$.

II.4 Towards a numerical resolution for the cascade obstacle problem

The purpose of this paragraph is to provide an idea about a possible numerical resolution of the multiple optimal stopping problem. We prove in Chapter V that the value function of the multiple optimal stopping problem is characterized as the unique viscosity solution of the dynamic programming equation (II.3.2). Our idea is to exploit this equation to implement a numerical approximation of the multiple optimal stopping problem. As we intend to handle relatively high dimensions, we choose to follow the approach of Sirignano & Spiliopoulos [71] and Raissi, Perdikaris & Karniadakis [66], i.e., we may approximate the value function with a neural network, whose calibration is made by using the PDE operator and the boundary conditions as loss function.

Denote $\hat{u}(\mathbf{w}; \cdot)$ the neural network approximating the solution u of (II.3.2), where $\mathbf{w} \in \mathbb{R}^K$ is the parameter of the network. Given $(\mu_1, \mu_2, \mu_3) \in \mathcal{P}([0, T] \times \mathbf{S}^N) \times \mathcal{P}(\mathbf{S}^N) \times \mathcal{P}([0, T] \times$

$\mathbb{R}^{d \times N}$), we define the theoretical loss function

$$\begin{aligned}
 L(\mathbf{w}) &:= \int_{[0,T] \times \mathbf{S}^N} |\mathcal{A}\hat{u}(\mathbf{w}; t, \mathbf{y})|^2 d\mu_1(t, \mathbf{y}) + \int_{\mathbf{S}^N} |\hat{u}(\mathbf{w}; T, \mathbf{y}) - \varphi(\mathbf{x})|^2 d\mu_2(\mathbf{y}) \quad (\text{II.4.1}) \\
 &\quad + \int_{[0,T] \times (\mathbb{R}^d)^N} |\hat{u}(\mathbf{w}; t, \mathbf{x}, 0) - \varphi(\mathbf{x})|^2 d\mu_3(t, \mathbf{x}),
 \end{aligned}$$

where

$$\mathcal{A}\hat{u}(\mathbf{w}; t, \mathbf{y}) := \min\{-\partial_t \hat{u}(\mathbf{w}; t, \mathbf{y}) - \mathcal{L}\hat{u}(\mathbf{w}; t, \mathbf{y}) \cdot \mathbf{i}, \hat{u}(\mathbf{w}; t, \mathbf{y}) - \max_{|\mathbf{i}'|=|\mathbf{i}|-1} \hat{u}(\mathbf{w}; t, \mathbf{x}, \mathbf{i}')\}.$$

Our algorithm shall consist in minimizing L over \mathbf{w} via a stochastic gradient descent. More precisely, it should follow the following steps:

1. Simulation of a random $(t_n^1, \mathbf{y}_n^1) \in [0, T] \times \mathbf{S}^N$, $\mathbf{y}_n^2 \in \mathbf{S}^N$ and $(t_n^3, \mathbf{x}_n^3) \in [0, T] \times \mathbb{R}^{d \times N}$ according to μ_1, μ_2 and μ_3 , respectively.
2. Computation of the squared error

$$\hat{L}(\mathbf{w}_n, \mathbf{q}) := (\mathcal{A}\hat{u}(\mathbf{w}_n; t_n, \mathbf{y}_n))^2 + (\hat{u}(\mathbf{w}_n; T, \mathbf{y}_n) - g^N(\mathbf{x}_n))^2 + (\hat{u}(\mathbf{w}_n; t_n, \mathbf{x}_n, 0) - g^N(\mathbf{x}_n))^2$$

3. Gradient descent: $\mathbf{w}_{n+1} := \mathbf{w}_n - \alpha_n \partial_{\mathbf{w}} \hat{L}(\mathbf{w}_n, \mathbf{q})$, where $\alpha := \{\alpha_n\}_{n \geq 1}$ is the learning rate,
4. Repeat until the stopping criterion is satisfied.

Note that the stopping criterion may take diverse forms (a fixed number of iterations, or a condition on the norm of the gradient of the loss function).

Chapter III

Dynamic programming equation for the mean field optimal stopping problem

III.1 Introduction

In this chapter we study a McKean-Vlasov type of optimal stopping problem, where the state dynamics and/or the reward function depend on the law of the stopped process. To be precise, given X_0 and an independent Brownian motion W , consider

$$X_t = X_0 + \int_0^{t \wedge \tau} b(s, X_s, \mathcal{L}_{X_s}) ds + \int_0^{t \wedge \tau} \sigma(s, X_s, \mathcal{L}_{X_s}) dW_s, \quad (\text{III.1.1})$$

where τ is a stopping time and \mathcal{L}_{X_s} denotes the law of X_s . We emphasize the impact of τ on \mathcal{L}_{X_s} , in particular, \mathcal{L}_{X_s} is neither equal to $\mathcal{L}_{X_{\tau \wedge s}^0}$ nor to $\mathcal{L}_{X_s^0 | s=\tau}$, where X^0 denotes the unstopped process:

$$X_t^0 = X_0 + \int_0^t b(s, X_s^0, \mathcal{L}_{X_s^0}) ds + \int_0^t \sigma(s, X_s^0, \mathcal{L}_{X_s^0}) dW_s. \quad (\text{III.1.2})$$

Our optimization problem is, for some functionals f and g defined on a space of probability laws,

$$V_0 := \sup_{\tau} \mathbb{E} \left[\int_0^{\tau} f(s, X_s, \mathcal{L}_{X_s}) ds \right] + g(\mathcal{L}_{X_{\tau}}). \quad (\text{III.1.3})$$

When b, σ and f do not depend on \mathcal{L}_{X_s} and $g(\mathcal{L}_{X_\tau}) = \mathbb{E}[\varphi(X_\tau)]$ for some function $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$, the above problem reduces to a standard optimal stopping problem, see e.g. Shiryaev [70]. The mean field optimal stopping problem (III.1.3) can be viewed as the limit of a multiple stopping problem over a large system interacting through the empirical measure:

$$\begin{aligned} X_t^i &= x_i + \int_0^{t \wedge \tau_i} b(s, X_s^i, \bar{\mu}_s) ds + \sigma(s, X_s^i, \bar{\mu}_s) dW_s^i, \quad \bar{\mu}_s := \frac{1}{N} \sum_{i=1}^N \delta_{X_s^i}; \\ V_0^N &:= \sup_{(\tau_1, \dots, \tau_N)} \mathbb{E} \left[\frac{1}{N} \sum_{i=1}^N \int_0^{\tau_i} f(s, X_s^i, \bar{\mu}_s) ds + g \left(\frac{1}{N} \sum_{i=1}^N \delta_{X_{\tau_i}^i} \right) \right], \end{aligned} \quad (\text{III.1.4})$$

where δ_x denotes the Dirac-measure, (W^1, \dots, W^N) are $N \times d$ -dimensional Brownian motions. We refer to Kobylanski, Quenez & Rouy-Mironescu [50] for general multiple stopping problems, and we shall investigate the convergence issue in Chapter V.

There has been a strong attention on mean field games of optimal stopping in the literature, see, e.g., Bertucci [9], Bouveret, Dumitrescu & Tankov [11], Carmona, Delarue & Lacker [23], and Nutz [62]. Given $\{\mu_t\}_{t \geq 0}$, consider the optimal stopping problem:

$$V_0^{\mu \cdot} := \sup_{\tau} \mathbb{E} \left[\int_0^{\tau} f(s, X_s^{\mu \cdot}, \mu_s) ds + g(\tau, X_{\tau}^{\mu \cdot}, \mu_{\tau}) \right], \quad (\text{III.1.5})$$

where X^{μ} is unstopped and solves a standard SDE (not McKean-Vlasov type as in (III.1.2)):

$$X_t^{\mu \cdot} = X_0 + \int_0^t b(s, X_s^{\mu \cdot}, \mu_s) ds + \int_0^t \sigma(s, X_s^{\mu \cdot}, \mu_s) dW_s.$$

Assume the above problem has an optimal stopping time $\tau^*(\mu \cdot)$, then the mean field game problem is to find a fixed point $\{\mu_t\}_{t \geq 0}$, namely the mean field equilibrium: $\mathcal{L}_{X_{\tau^*(\mu \cdot)}^{\mu \cdot}} = \mu_t$, $t \geq 0$. We remark that in the last mean field game, for given $\{\mu_t\}_{t \geq 0}$, the dynamics of X^{μ} does not depend on the stopping time τ and the optimal stopping problem (III.1.5) is a standard one as in [70], so it has a completely different structure than our optimal stopping problem. We would also like to mention Li [56], Briand, Elie & Hu [12], and Djehiche, Elie & Hamadene [31] for closely related works on mean field type reflected BSDEs, and Belomestny & Schoenmakers [7] for a numerical method for mean field type optimal stopping problems. However, in all these works again the dynamics of the state process does not depend on the stopping time τ . To our best knowledge, our work is the first in the literature to study the optimal stopping problem where the dynamics depends on the law of the stopped process, or say in (III.1.4) the interaction is through the stopped particles.

Besides the obvious connection with large interacting particle systems, the general form (III.1.3) is convenient for many other applications. For example, by considering the unstopped state process X^0 in (III.1.2), the optimal stopping of mean variance problem $\sup_{\tau} \{ \mathbb{E}[X_{\tau}^0] - \frac{1}{2} \text{Var}(X_{\tau}^0) \}$ corresponds to $g(\mu) = \int_{\mathbb{R}} (x - \frac{1}{2}x^2)\mu(dx) + (\int_{\mathbb{R}} x\mu(dx))^2$ for a square integrable measure μ . Another example is the optimal stopping problem under probability distortion, used in behavioral economics, which corresponds to $g(\mu) = \int_0^{\infty} \varphi(\mu([U^{-1}(y), \infty))dy$, for some utility function $U : \mathbb{R} \rightarrow [0, \infty)$, and some distortion function $\varphi : [0, 1] \rightarrow [0, 1]$. When X^0 is a Geometric Brownian motion and the time horizon is infinite, Pedersen & Peskir [64] proved the existence of optimal stopping time for the mean variance problem, and Xu & Zhou [78] obtained the optimal stopping time for the probability distortion problem for some special shapes of the functions φ and U (convex, concave, or reverse S-shaped). We remark that these problems are typically considered as time inconsistent problems, as we will explain in the next paragraph, and the existing literature considers only the static problem, namely the existence of optimal stopping time for the problem over a fixed time interval $([0, \infty)$ or $[0, T]$). We shall study the problem (III.1.3) systematically, and more importantly, dynamically. We remark that, even when we consider only the unstopped state process X^0 , our dynamic approach for the optimal stopping problem (III.1.3) seems new.

It is well known that standard optimal stopping problems can be solved by the dynamic programming approach, see e.g. El Karoui [41] and Shiryaev [70]. The situation here is more subtle because of the involvement of the law. In order to have Dynamic Programming Principle (DPP, for short), it is crucial to choose the right variable, which stands for the information one needs to make the dynamic system "Markovian". Indeed, if we define $V(t, x)$ as the dynamic value function for problem (III.1.3) on $[t, T]$ with initial condition $X_t = x$, which in the case (III.1.4) means we observe only the state x_i of one particular player i , the DPP would fail. Consequently the problem is often viewed as time inconsistent in the standard sense. Moreover, even if we define $V(t, \mu)$ as the dynamic value function for problem (III.1.3) on $[t, T]$ with initial condition $\mathcal{L}_{X_t} = \mu$, the DPP would still fail.

Our first observation is that a successful DPP requires the introduction of the survival process $I_t := \mathbf{1}_{\{\tau > t\}}$. To be precise, we will have the desired DPP if we write the dynamic value function as $V(t, \mathcal{L}_{(X_t, I_t)})$, that is, to maintain the time consistency, we need to know not only the current states of all particles, but also which particles are still surviving. Moreover, we formulate a weak relaxed version of (III.1.1) by allowing for randomized stopping

times induced by the set $\mathcal{P}(t, m)$ of all joint distributions \mathbb{P} of the stopped process and the corresponding stopping time, started at time t from the initial distribution m . Such a weak formulation is particularly convenient here for two reasons:

- the set of controls has been shifted from the stopping times into $\mathcal{P}(t, m)$, that we will prove to be compact, implying the existence of an optimal \mathbb{P}^* to the mean field optimal stopping problem as long as f and g are upper-semicontinuous;
- shifting the state variable from the process X into the flow of joint marginal distributions, denoted as $\{\mathbb{P}_{(X_t, I_t)}\}$ in order to emphasize its dependence on \mathbb{P} , enables us to establish a DPP and to derive a dynamic programming equation on the space of measures to characterize the value function V .

More precisely, given that the laws are deterministic, our following DPP is very easy to establish:

$$V(t, m) = \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \int_t^s \mathbb{E}[f(r, X_r, \mathbb{P}_{(X_r, I_r)}) I_r] dr + V(s, \mathbb{P}_{(X_s, I_s)}),$$

Such dynamic programming approach has also been used successfully in the mean field control literature, where the state variable is \mathcal{L}_{X_t} , see, e.g., Carmona & Delarue [21, Vol. 1, ch. 6], Pham & Wei [65], Wu & Zhang [77], and Djete, Possamai & Tan [34].

The corresponding dynamic programming equation is as usual derived by means of Itô's formula. Itô's formula for functions on Wasserstein space of probability measures has been established for continuous diffusions by Buckdahn, Li, Peng & Rainer [14] and Chassagneux, Crisan & Delarue [27], and for jump diffusions by Li [57] and Burzoni, Ignazio, Reppen & Soner [15]. However, [57, 15] require the law of the state process to be continuous under the Wasserstein distance, while in our case it is quite possible that $t \mapsto \mathbb{P}_{I_t}$ is discontinuous. We thus first extend Itô's formula so that both the state process and its law can have jumps. Our proof follows the standard derivation, based on the linear functional derivative. We introduce an appropriate time discretization and reduce our derivation to the standard Itô's formula for càdlàg semimartingales. We also refer to the independent work of Guo, Pham & Wei [46], who prove similar results by using density arguments.

Together with the DPP, our Itô's formula immediately leads to the desired dynamic programming equation, an obstacle problem on the Wasserstein space. We shall characterize the value function, provided its sufficient regularity, as the unique classical solution of the obstacle problem, and we will use the value function to characterize the structure of the optimal stopping time. The regularity of the value function, of course, remains a challenging

problem in general, and we will therefore investigate the viscosity solution approach for the obstacle problem in Chapter IV.

The chapter is structured as follows. In Section III.2, we set the mean field optimal stopping problem in weak formulation, and establish the dynamic programming principle. In Section III.3 we prove the Itô's formula for possibly discontinuous flows of measures of semimartingales, that in particular allows us to differentiate smooth functions along the flow $\{\mathbb{P}_{(X_t, I_t)}\}_{t \in [0, T]}$. In Section III.4 we derive the dynamic programming equation for the value function and establish its classical solution theory. Section III.5 is dedicated to some examples illustrating the connection with the standard optimal stopping theory, and shedding more light on a class of criteria including the mean-variance one. We also provide an explicit example which exhibits both features of pure stopping strategies and randomized ones. In Section III.6 we provide two extensions. Subsection III.6.1 extends our results to the infinite horizon setting, and Subsection III.6.2 provides a quick discussion of the extension to the case where the process X is a jump-diffusion. Finally, Appendices III.7 and III.8 report some technical proofs.

III.2 Formulation of the mean field optimal stopping problem

Let $T < \infty$ be fixed, and $\Omega := C^0([-1, T], \mathbb{R}^d) \times \mathbb{I}^0([-1, T])$ the canonical space, where:

- $C^0([-1, T], \mathbb{R}^d)$ is the set of continuous paths from $[-1, T]$ to \mathbb{R}^d , constant on $[-1, 0]$;
- $\mathbb{I}^0([-1, T])$ is the set of non-increasing and càdlàg maps from $[-1, T]$ to $\{0, 1\}$, constant on $[-1, 0]$, and ending with value 0 at T .

We equip Ω with the Skorokhod distance, under which it is a Polish space. The choice of the extension to -1 is arbitrary, the extension of time to the left of the origin is only needed to allow for an immediate stop at time $t = 0$.

We denote $Y := (X, I)$ the canonical process, with state space $\mathbf{S} := \mathbb{R}^d \times \{0, 1\}$, its canonical filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in [-1, T]}$, and the corresponding jump time of the survival process I :

$$\tau := \inf\{t \geq 0 : I_t = 0\}, \text{ so that } I_t := I_{0-} \mathbf{1}_{t < \tau} \text{ for all } t \in [-1, T]. \quad (\text{III.2.1})$$

By the càdlàg property of I , τ is an \mathbb{F} -stopping time. Denote further

$$\mathbf{Q}_t := [t, T] \times \mathcal{P}_2(\mathbf{S}), \quad \text{and} \quad \bar{\mathbf{Q}}_t := [t, T] \times \mathcal{P}_2(\mathbf{S}), \quad t \in [0, T].$$

CHAPTER III. DYNAMIC PROGRAMMING EQUATION FOR THE MEAN FIELD OPTIMAL STOPPING PROBLEM

Let $(b, \sigma, f) : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbf{S}) \rightarrow \mathbb{R}^d \times \mathbb{S}_d^+ \times \mathbb{R}$ and $g : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$, where \mathbb{S}_d^+ denotes the set of $d \times d$ non-negative symmetric matrices. Throughout the chapter, the following assumption will always be in force, where $\mathcal{P}_2(\mathbf{S})$ is equipped with the \mathcal{W}_2 -distance.

Assumption III.2.1 (i) b, σ are continuous in t , and uniformly Lipschitz continuous in (x, m) .

(ii) f is Borel measurable and has quadratic growth in $x \in \mathbb{R}^d$, and the following function F is continuous on $[0, T] \times \mathcal{P}_2(\mathbf{S})$:

$$F(t, m) := \int_{\mathbb{R}^d} f(t, x, m) m(dx, 1). \quad (\text{III.2.2})$$

(iii) g is upper-semicontinuous and locally bounded; and extended to $\mathcal{P}_2(\mathbf{S})$ by $g(m) := g(m(\cdot, \{0, 1\}))$.

Define the stopped McKean-Vlasov dynamics on $[0, T]$:

$$X_s = X_0 + \int_0^s b(r, X_r, \mathbb{P}_{Y_r}) I_r dr + \int_0^s \sigma(r, X_r, \mathbb{P}_{Y_r}) I_r dW_r^\mathbb{P} \text{ and } I_s = I_{0-} \mathbf{1}_{s < \tau}, \quad (\text{III.2.3})$$

where a solution \mathbb{P} of the last SDE is defined by the requirement that the following processes M and N are \mathbb{P} -martingales on $[0, T]$:

$$M := X \cdot - \int_0^\cdot b(r, X_r, \mathbb{P}_{Y_r}) I_r dr \text{ and } N := M^2 - \int_0^\cdot \sigma^2(r, X_r, \mathbb{P}_{Y_r}) I_r dr. \quad (\text{III.2.4})$$

Note that $X \cdot = X_{\cdot \wedge \tau}$, and in particular $X_T = X_\tau$, \mathbb{P} -a.s.

We then focus on the mean field optimal stopping problem: given $\mu \in \mathcal{P}_2(\mathbb{R}^d)$,

$$V_0 := \sup_{\mathbb{P}} \mathbb{E}^\mathbb{P} \left[\int_0^\tau f(r, X_r, \mathbb{P}_{Y_r}) dr \right] + g(\mathbb{P}_{X_\tau}) = \sup_{\mathbb{P}} \int_0^T F(r, \mathbb{P}_{Y_r}) dr + g(\mathbb{P}_{Y_T}), \quad (\text{III.2.5})$$

where the supremum is taken over all solutions \mathbb{P} of the McKean-Vlasov SDE satisfying the constraint $\mathbb{P}_{X_0} = \mu$ and $\mathbb{P}(I_{0-} = 1) = 1$. We recall that this problem is motivated by the N -multiple optimal stopping problem (III.1.4), whose convergence is studied in Chapter V.

In order to solve this problem, we use the dynamic programming approach, made possible by an appropriate dynamic version of the problem. This requires to take as a state the joint distribution m_t of the variables $Y_t = (X_t, I_t)$, which leads to the dynamic value function

$$V(t, m) := \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \int_t^T F(r, \mathbb{P}_{Y_r}) dr + g(\mathbb{P}_{Y_T}), \quad (t, m) \in \overline{\mathbf{Q}}_0, \quad (\text{III.2.6})$$

where $\mathcal{P}(t, m)$ is the set of probability measures \mathbb{P} on (Ω, \mathcal{F}_T) such that

- $\mathbb{P}_{Y_{t-}} = m$ and $s \in [-1, t) \rightarrow Y_s$ is constant, \mathbb{P} -a.s.
- The processes M, N of (III.2.4) are \mathbb{P} -martingales on $[t, T]$, so that, for some \mathbb{P} -Brownian motion $W^\mathbb{P}$,

$$X_s = X_t + \int_t^s b(r, X_r, \mathbb{P}_{Y_r}) I_r dr + \sigma(r, X_r, \mathbb{P}_{Y_r}) I_r dW_r^\mathbb{P}, \quad I_s = I_{t-} \mathbf{1}_{s < \tau}, \quad \mathbb{P} - \text{a.s.} \quad (\text{III.2.7})$$

Proposition III.2.2 *For any $(t, m) \in \mathbf{Q}_0$, the set $\mathcal{P}(t, m)$ is compact under the Wasserstein distance \mathcal{W}_2 . Consequently, existence holds for the mean field optimal stopping problem (III.2.6).*

We relegate this proof to Appendix III.7. Our main result of this section is the following dynamic programming principle (DPP for short).

Theorem III.2.3 *For any $(t, m) \in \mathbf{Q}_0$ and $s \in [t, T]$, we have the DPP:*

$$\begin{aligned} V(t, m) &= \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \int_t^s F(r, \mathbb{P}_{Y_r}) dr + V(s, \mathbb{P}_{Y_{s-}}) \\ &= \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \int_t^s F(r, \mathbb{P}_{Y_r}) dr + V(s, \mathbb{P}_{Y_s}). \end{aligned} \quad (\text{III.2.8})$$

Proof Denote, for any probability measure \mathbb{P} on (Ω, \mathcal{F}_T) ,

$$J(t, \mathbb{P}) := \int_t^T F(r, \mathbb{P}_{Y_r}) dr + g(\mathbb{P}_{Y_T}).$$

We start with proving the first equality of (III.2.8). Let $\tilde{V}(t, m)$ denote the middle term of (III.2.8). Fix an arbitrary $\mathbb{P} \in \mathcal{P}(t, m)$, and denote $\tilde{m} := \mathbb{P}_{Y_{s-}}$.

First, for any time partition $\pi : -1 = t_0 < \dots < t_m < s \leq t_{m+1} < \dots < t_{m+n} = T$, introduce the finite measure: for any $A_i \in \mathcal{B}(\mathbf{S})$,

$$\nu_\pi(A_0 \times \dots \times A_{m+n}) := \mathbb{P}\left(Y_{s-} \in \cap_{i=0}^m A_i, Y_{t_{m+j}} \in A_{m+j}, j = 1, \dots, n\right).$$

It is clear that $\{\nu_\pi\}_\pi$ satisfies the consistency condition, and thus it follows from the Kolmogorov extension theorem that there exists a probability measure $\tilde{\mathbb{P}}$ on (Ω, \mathcal{F}_T) such that $\{\nu_\pi\}_\pi$ is the finite distribution of the process Y under $\tilde{\mathbb{P}}$. It is straightforward to verify $\tilde{\mathbb{P}} \in \mathcal{P}(s, \tilde{m})$, and $\tilde{\mathbb{P}}_{Y_r} = \mathbb{P}_{Y_r}$ for all $r \in [s, T]$. Thus,

$$J(t, \mathbb{P}) = \int_t^s F(r, \mathbb{P}_{Y_r}) dr + J(s, \mathbb{P}) = \int_t^s F(r, \mathbb{P}_{Y_r}) dr + J(s, \tilde{\mathbb{P}}) \leq \int_t^s F(r, \mathbb{P}_{Y_r}) dr + V(s, \mathbb{P}_{Y_{s-}}).$$

Since $\mathbb{P} \in \mathcal{P}(t, m)$ is arbitrary, we obtain $V(t, m) \leq \tilde{V}(t, m)$.

On the other hand, given \tilde{m} , by Proposition III.2.2 there exists $\tilde{\mathbb{P}} \in \mathcal{P}(s, \tilde{m})$ such that $J(s, \tilde{\mathbb{P}}) = V(s, \tilde{m})$. For the above time partition π , we introduce another finite measure: for any $A_i \in \mathcal{B}(\mathbf{S})$,

$$\nu_\pi(A_0 \times \cdots \times A_{m+n}) := \int_{\mathbf{S}} \mathbb{E}^{\mathbb{P}} \left[\prod_{i=0}^m \mathbf{1}_{A_i}(Y_{t_i}) \middle| Y_{s-} = y \right] \times \mathbb{E}^{\tilde{\mathbb{P}}} \left[\prod_{j=1}^n \mathbf{1}_{A_{m+j}}(Y_{t_{m+j}}) \middle| Y_{s-} = y \right] \tilde{m}(dy).$$

Applying the Kolmogorov extension theorem again there exists a probability measure $\hat{\mathbb{P}}$ on (Ω, \mathcal{F}_T) such that $\{\nu_\pi\}_\pi$ is the finite distribution of the process Y under $\hat{\mathbb{P}}$. It is clear that $\hat{\mathbb{P}} = \mathbb{P}$ on \mathcal{F}_{s-} , and $\{Y_{s-}, Y_r, s \leq r \leq T\}$ has the same distribution under $\hat{\mathbb{P}}$ and $\tilde{\mathbb{P}}$, and $\{Y_r, r < s\}$ and $\{Y_r, r \geq s\}$ are conditionally independent under $\hat{\mathbb{P}}$, conditional on Y_{s-} . We shall emphasize that this conditional independence is valid only conditional on Y_{s-} , the process Y is in general not Markov under $\hat{\mathbb{P}}$. It is obvious that the processes M, N in (III.2.4) remain to be $\hat{\mathbb{P}}$ -martingales on $[t, s]$. Moreover, for any $s \leq s_1 < s_2 \leq T$, any $0 = t_0 < \cdots < t_m < s \leq t_{m+1} < \cdots < t_{m+n} = s_1$, and any bounded measurable function $\varphi_1 : \mathbb{R}^{(m+1)d} \rightarrow \mathbb{R}$, $\varphi_2 : \mathbb{R}^{nd} \rightarrow \mathbb{R}$,

$$\begin{aligned} & \mathbb{E}^{\hat{\mathbb{P}}} \left[[M_{s_2} - M_{s_1}] \varphi_1(Y_{t_0}, \dots, Y_{t_m}) \varphi_2(Y_{t_{m+1}}, \dots, Y_{t_{m+n}}) \right] \\ &= \mathbb{E}^{\hat{\mathbb{P}}} \left[\mathbb{E}^{\tilde{\mathbb{P}}} \left[[M_{s_2} - M_{s_1}] \varphi_2(Y_{t_{m+1}}, \dots, Y_{t_{m+n}}) \middle| Y_{s-} \right] \times \mathbb{E}^{\mathbb{P}} \left[\varphi_1(Y_{t_0}, \dots, Y_{t_m}) \middle| Y_{s-} \right] \right] \\ &= \mathbb{E}^{\hat{\mathbb{P}}} \left[0 \times \mathbb{E}^{\mathbb{P}} \left[\varphi_1(Y_{t_0}, \dots, Y_{t_m}) \middle| Y_{s-} \right] \right] = 0. \end{aligned}$$

Then M is a $\hat{\mathbb{P}}$ -martingale on $[s, T]$ as well, and hence a $\hat{\mathbb{P}}$ -martingale on $[t, T]$. Similarly we can show that N is a $\hat{\mathbb{P}}$ -martingale on $[t, T]$, then $\hat{\mathbb{P}} \in \mathcal{P}(t, m)$. Therefore,

$$\begin{aligned} \int_t^s F(r, \mathbb{P}_{Y_r}) dr + V(s, \mathbb{P}_{Y_{s-}}) &= \int_t^s F(r, \mathbb{P}_{Y_r}) dr + J(s, \tilde{\mathbb{P}}) \\ &= \int_t^s F(r, \hat{\mathbb{P}}_{Y_r}) dr + J(s, \hat{\mathbb{P}}) = J(t, \hat{\mathbb{P}}) \leq V(t, m). \end{aligned}$$

Since $\mathbb{P} \in \mathcal{P}(t, m)$ is arbitrary, we obtain $\tilde{V}(t, m) \leq V(t, m)$, and hence the first equality of (III.2.8).

It remains to prove the second equality of (III.2.8). First, since $I_s \leq I_{s-}$, it is obvious that $\mathcal{P}(s, \mathbb{P}_{Y_s}) \subset \mathcal{P}(s, \mathbb{P}_{Y_{s-}})$, and thus $V(s, \mathbb{P}_{Y_s}) \leq V(s, \mathbb{P}_{Y_{s-}})$ for all $\mathbb{P} \in \mathcal{P}(t, m)$. On the other hand, for any $\mathbb{P} \in \mathcal{P}(t, m)$, set $\tilde{\mathbb{P}} \in \mathcal{P}(t, m)$ be such that $\tilde{\mathbb{P}} = \mathbb{P}$ on \mathcal{F}_{s-} and $I_r = I_{s-}$

for all $r \geq s$, $\tilde{\mathbb{P}}$ -a.s. Then $\tilde{\mathbb{P}}_{Y_s} = \mathbb{P}_{Y_{s-}}$, and thus

$$\begin{aligned} \int_t^s F(r, \mathbb{P}_{Y_r}) dr + V(s, \mathbb{P}_{Y_{s-}}) &= \int_t^s F(r, \tilde{\mathbb{P}}_{Y_r}) dr + V(s, \tilde{\mathbb{P}}_{Y_s}) \\ &\leq \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \int_t^s F(r, \mathbb{P}_{Y_r}) dr + V(s, \mathbb{P}_{Y_s}). \end{aligned}$$

This completes the proof immediately. \blacksquare

In order to derive the dynamic programming equation, we follow the usual procedure, which requires Itô's formula along the flow of measures $\{\mathbb{P}_{Y_s}\}_{t \leq s \leq T}$, as we shall develop in the next section.

III.3 Itô's formula for flows of laws of semimartingales

In contrast with the available literature reviewed in the introduction, our Itô's formula allows for possible jumps for both the semimartingale and its flow of marginal laws $\mathbf{m} = \{m_s\}$. The mapping $s \mapsto m_s$ is also càdlàg and we shall denote

$$J_{\mathbb{T}}(\mathbf{m}) := \{s \in \mathbb{T} : m_s \neq m_{s-}\}, \quad J_{\mathbb{T}}^c(\mathbf{m}) := \{s \in \mathbb{T} : m_s = m_{s-}\}, \quad \text{for all } \mathbb{T} \subset [0, T] \quad (\text{III.3.1})$$

We first introduce the notion of *linear functional derivative*, in the same spirit as Carmona & Delarue [21, Vol 1, Definition 5.43] and Cardialaguet, Delarue, Lasry & Lions [17]:

Definition III.3.1 (i) $u : \mathcal{P}_2(\mathbb{R}^{d'}) \rightarrow \mathbb{R}$ has a linear functional derivative if there exists

$$\delta_m u : \mathcal{P}_2(\mathbb{R}^{d'}) \times \mathbb{R}^{d'} \rightarrow \mathbb{R}$$

such that $\delta_m u$ is continuous for the product topology and

- the mapping $y \mapsto \delta_m u(m, y)$ has quadratic growth in y , locally uniformly in m . That is, for any compact set $\Xi \subset \mathcal{P}_2(\mathbb{R}^{d'})$, $\sup_{m \in \Xi} |\delta_m u(m, y)| \leq C_{\Xi}[1 + |y|^2]$.

- for all $m, m' \in \mathcal{P}_2(\mathbb{R}^{d'})$,

$$u(m') - u(m) = \int_0^1 \int_{\mathbb{R}^{d'}} \delta_m u(\lambda m' + (1 - \lambda)m, y)(m' - m)(dy) d\lambda. \quad (\text{III.3.2})$$

(ii) $C_2^{1,1}([0, T] \times \mathcal{P}_2(\mathbb{R}^{d'}))$ denotes the set of functions $u : [0, T] \times \mathcal{P}_2(\mathbb{R}^{d'}) \rightarrow \mathbb{R}$ such that

- $\partial_t u, \delta_m u, \partial_y \delta_m u, \partial_{yy}^2 \delta_m u$ exist and are continuous in all variables;
- $\partial_{yy}^2 \delta_m u$ is bounded in y , locally uniformly in (t, m) .

Here the subscript $_2$ in $C_2^{1,1}$ is to refer the growth conditions so as to ensure appropriate square integrability in the analysis below.

By abusing the notation, in the following statement, we let Y denote a general càdlàg \mathbb{R}^d -valued semimartingale on $[0, T]$. We denote Y^c the continuous part of Y ; $Y_t^c = Y_0 + M_t^c + A_t^c$ the Doob-Meyer decomposition, where M^c is the martingale part and A^c is the finite variation part; $\|A^c\|_t$ the total variation process of A^c and $\langle M^c \rangle_t$ the quadratic variation process of M^c .

Theorem III.3.2 (Itô's formula) *Let $u \in C_2^{1,1}([0, T] \times \mathcal{P}_2(\mathbb{R}^d))$, and assume*

$$\mathbb{E} \left[\|A^c\|_T^2 + \langle M^c \rangle_T + \left(\sum_{0 < s \leq T} |Y_s - Y_{s-}| \right)^2 \right] < \infty. \quad (\text{III.3.3})$$

Then, denoting $\mathbf{m} = \{m_s\}_{0 \leq s \leq T}$ the marginal laws of Y_s ,

$$\begin{aligned} u(T, m_T) &= u(0, m_0) + \int_0^T \partial_t u(s, m_s) ds \\ &+ \mathbb{E} \left[\int_0^T \partial_y \delta_m u(s, m_s, Y_s) \cdot dA_s^c + \frac{1}{2} \int_0^T \partial_{yy}^2 \delta_m u(s, m_s, Y_s) : d\langle M^c \rangle_s \right] \\ &+ \sum_{s \in J_{(0, T]}(\mathbf{m})} [u(s, m_s) - u(s, m_{s-})] + \mathbb{E} \left[\sum_{s \in J_{(0, T]}^c(\mathbf{m})} (\delta_m u(s, m_s, Y_s) - \delta_m u(s, m_s, Y_{s-})) \right]. \end{aligned} \quad (\text{III.3.4})$$

The proof of this result is relegated to Appendix III.8. Note that (III.3.4) exhibits two different sums: one refers to the jumps of Y , while the other to the jumps of the marginals \mathbf{m} . The Poisson process provides a simple example of pure jump process with continuous marginals (i.e., $J_{(0, T]}(\mathbf{m}) = \emptyset$).

We now specialize the discussion to the case $Y := (X, I)$. Note that $\mathcal{P}_2(\mathbf{S}) \subset \mathcal{P}_2(\mathbb{R}^{d+1})$, we may restrict Definition III.3.1 to $\mathcal{P}_2(\mathbf{S})$ only.

Definition III.3.3 *Let $C_2^{1,1}(\overline{\mathbf{Q}}_0)$ denote the set of functions $u : \overline{\mathbf{Q}}_0 \rightarrow \mathbb{R}$ such that $\partial_t u$, $\delta_m u$, $\partial_x \delta_m u$, $\partial_{xx}^2 \delta_m u$ exist and are continuous in all variables, and $\partial_{xx}^2 \delta_m u$ is bounded in x , locally uniformly in (t, m) , where the functional linear derivative takes the form $\delta_m u : (t, m, x, i) \in \overline{\mathbf{Q}}_0 \times \mathbb{R}^d \times \{0, 1\} \rightarrow \mathbb{R}$ satisfying, for any $t \in [0, T]$ and $m, m' \in \mathcal{P}_2(\mathbf{S})$,*

$$u(t, m') - u(t, m) = \int_0^1 \int_{\mathbf{S}} \delta_m u(t, \lambda m' + (1 - \lambda)m, x, i) (m' - m)(dx, di) d\lambda.$$

In this case, of course there is no need to consider the derivative of $\delta_m u$ with respect to the i -variable. Instead, we denote

$$\delta_m u_i(t, m, x) := \delta_m u(t, m, x, i) \text{ for } i \in \{0, 1\}, \text{ and } D_I u := \delta_m u_1 - \delta_m u_0. \quad (\text{III.3.5})$$

Example III.3.4 Let us define, for a given probability measure \mathbb{P} , $u(m) := \varphi(m[\psi])$, with ψ smooth and $m[\psi] := \sum_{i=0,1} \int_{\mathbb{R}^d} \psi(x, i) m(dx, i)$. Then we compute

$$\delta_m u(m, x, i) = \varphi'(m[\psi])\psi(x, i) \text{ and } D_I u(m, x) = \varphi'(m[\psi])[\psi(x, 1) - \psi(x, 0)].$$

Recall the infinitesimal generator of X , we define

$$\begin{aligned} \mathbb{L}u(t, m) &:= \partial_t u(t, m) + \int_{\mathbb{R}^d} \mathcal{L}_x \delta_m u_1(t, m, x) m(dx, 1), \quad \text{where} \\ \mathcal{L}_x \delta_m u_1(t, m, x) &:= b(t, x, m) \cdot \partial_x \delta_m u_1(t, m, x) + \frac{1}{2} \sigma^2(t, x, m) : \partial_{xx}^2 \delta_m u_1(t, m, x). \end{aligned} \quad (\text{III.3.6})$$

We now state the Itô formula for $\mathbf{m} := \{m_s := \mathbb{P}_{Y_s}\}_{s \in [-1, T]}$. Note that in Theorem III.3.2, we consider the jumps on $(0, T]$. However, in light of DPP (III.2.8), it is more convenient to consider the jumps on $[0, T)$, namely we include the jump at the initial point instead of the ending point. Such an adjustment is straightforward.

Corollary III.3.5 Let $m \in \mathcal{P}_2(\mathbf{S})$, $\mathbb{P} \in \mathcal{P}(0, m)$, and $u \in C_2^{1,1}(\overline{\mathbf{Q}}_0)$. Then,

$$\begin{aligned} u(T, m_{T-}) &= u(0, m) + \int_0^T \mathbb{L}u(s, m_s) ds \\ &+ \sum_{s \in J_{[0, T)}(\mathbf{m})} [u(s, m_s) - u(s, m_{s-})] + \mathbb{E}^{\mathbb{P}} \left[\int_{J_{[0, T)}^c(\mathbf{m})} D_I u(s, m_s, X_s) dI_s \right]. \end{aligned} \quad (\text{III.3.7})$$

Proof We can easily see that $Y_s - Y_{s-} = (0, I_s - I_{s-})$ and

$$Y_s^c = (X_s, I_{0-}), \quad dM_s^c = (\sigma(s, X_s, m_s) dW_s^{\mathbb{P}}, 0), \quad dA_s^c = (b(s, X_s, m_s) ds, 0).$$

Then (III.3.3) obviously holds true. Now following Theorem III.3.2, but by considering the jump at 0 instead of at T , we have

$$\begin{aligned} u(T, m_{T-}) - u(0, m) &= \int_0^T \partial_t u(s, m_s) ds + \mathbb{E}^{\mathbb{P}} \left[\int_0^T \partial_y \delta_m u_1(s, m_s, X_s) \cdot b(s, X_s, m_s) ds \right. \\ &+ \left. \frac{1}{2} \int_0^T \partial_{xx}^2 \delta_m u_1(s, m_s, X_s) : d\langle X \rangle_s \right] + \sum_{s \in J_{[0, T)}(\mathbf{m})} [u(s, m_s) - u(s, m_{s-})] \\ &+ \mathbb{E}^{\mathbb{P}} \left[\sum_{s \in J_{[0, T)}^c(\mathbf{m})} [\delta_m u(s, m_s, Y_s) - \delta_m u(s, m_s, Y_{s-})] \right] \\ &= \int_0^T \mathbb{L}u(s, m_s) ds + \sum_{s \in J_{[0, T)}(\mathbf{m})} [u(s, m_s) - u(s, m_{s-})] + \mathbb{E}^{\mathbb{P}} \left[\int_{J_{[0, T)}^c(\mathbf{m})} D_I u(s, m_s, X_s) dI_s \right], \end{aligned}$$

where the last equality thanks to the fact that $I_s \neq I_{s-}$ if and only if $I_s = 0, I_{s-} = 1$. \blacksquare

We remark that, in this case $J_{[0, T]}(\mathbf{m}) = \{s \in [0, T] : \mathbb{P}(\tau = s) > 0\}$. That is, $J_{[0, T]}(\mathbf{m})$ is the collection of all atoms of τ under \mathbb{P} .

III.4 Obstacle problem on the Wasserstein space

III.4.1 The dynamic programming equation

We first introduce a partial order \preceq on $\mathcal{P}_2(\mathbf{S})$: we say that $m' \preceq m$ if

$$m'(dx, 1) = p(x)m(dx, 1), \text{ and } m'(dx, 0) = [1 - p(x)]m(dx, 1) + m(dx, 0), \quad (\text{III.4.1})$$

for some measurable $p : \mathbb{R}^d \rightarrow [0, 1]$, i.e. $m'(dx, 1)$ is obtained from m by randomly stopping a proportion $1 - p(x)$ of the surviving particles. In our context, $m_{t-} = \mathbb{P}_{Y_{t-}}$ and $m_t = \mathbb{P}_{Y_t}$, with $\mathbb{P} \in \mathcal{P}(t, m)$, so that $m_t \preceq m_{t-}$ with conditional transition probability

$$p(x) = p(t, x) := \mathbb{P}(I_t = 1 \mid X_t = x, I_{t-} = 1). \quad (\text{III.4.2})$$

Remark III.4.1 The set $\{m' : m' \preceq m\}$ is compact, as it is in continuous bijection with $\{\hat{m} \in \mathcal{P}_2(\mathbf{S} \times \{0, 1\}) : \hat{m} \circ (\mathbf{x}, \mathbf{i})^{-1} = m\}$, with $(\mathbf{x}, \mathbf{i}, \mathbf{i}')$ the projection coordinates on $\mathbf{S} \times \{0, 1\}$.

Our main objective is to show that the dynamic programming equation corresponding to our mean field optimal stopping problem, as deduced from the DPP (III.2.8), is

$$\begin{aligned} \min_{m' \in C_u(t, m)} [-(\mathbb{L}u + F)(t, m')] = 0, \quad D_I u(t, m, \cdot) \geq 0, \quad u(T, \cdot) = g, \text{ for all } (t, m) \in \mathbf{Q}_0, \\ \text{where } C_u(t, m) := \left\{ m' \preceq m : u(t, m') = u(t, m) \right\}. \end{aligned} \quad (\text{III.4.3})$$

By analogy with standard optimal stopping, we call (III.4.3) *obstacle problem* on the Wasserstein space. The different components of this equation have the following interpretation. ■

Remark III.4.2 (i) As will be proved in Lemma III.4.3, the inequality $D_I u(t, m, \cdot) \geq 0$ expresses the natural monotonicity of the optimal stopping problem, i.e. u is increasing for \preceq . In other words, the larger the set of surviving particles is, the larger the value function is.

(ii) $C_u(t, m)$ is the collection of *admissible stopping strategies at time t* , i.e. those that preserve the value function for smaller sets of surviving particles.

(iii) The equation $\min_{m' \in C_u(t, m)} -(\mathbb{L}u + F)(t, m') = 0$ characterizes the sets of particles that are optimal to keep diffusing (in the same spirit as the classical HJB equation, where the min characterizes the optimal controls). Note that $C_u(t, m)$ is compact, as a closed subset of the compact set $\{m' \preceq m\}$, see Remark III.4.1. Therefore the min is attained by the continuity of $(\mathbb{L}u + F)(t, \cdot)$. Finally, as $m \in C_u(t, m)$, we have $-(\mathbb{L}u + F)(t, m) \geq 0$.

(iv) The boundary condition $u(T, \cdot) = g$ is due to (III.2.6) directly. Moreover, the boundary condition implies that $u(t, m) = g(m)$, for all $t \in [0, T]$ and $m \in \partial\mathcal{P}_2(\mathbf{S}) := \{m \in \mathcal{P}_2(\mathbf{S}) : m(\mathbb{R}^d, 1) = 0\}$, i.e. all particles are stopped. Indeed, in this case $\{m' : m' \preceq m\} = \{m\}$ and thus $C_u(t, m) = \{m\}$. Recall (III.2.2) and (III.3.6), then (III.4.3) implies to $-\partial_t u(t, m) = -(\mathbb{L}u + F)(t, m) = 0$. This clearly implies that $u(t, m) = u(T, m) = g(m)$ for all $t \in [0, T]$. ■

Lemma III.4.3 *Let $u : \mathcal{P}_2(\mathbf{S}) \rightarrow \mathbb{R}$ admit a linear functional derivative. Then u is nondecreasing for \preceq if and only if $D_I u(m, \cdot) \geq 0$ for all $m \in \mathcal{P}_2(\mathbf{S})$.*

Proof First, assume $D_I u(m, \cdot) \geq 0$ for all $m \in \mathcal{P}_2(\mathbb{R}^d)$. Then, for $m' \preceq m$ with corresponding transition probability p , we have

$$u(m) - u(m') = \int_0^1 \int_{\mathbb{R}^d} D_I u(\lambda m + (1 - \lambda)m', x)[1 - p(x)]m(dx, 1)d\lambda \geq 0.$$

Conversely, assume that u is nondecreasing for \preceq , i.e. $u(m') \preceq u(m)$ for all $m' \preceq m$. Introduce $\mathcal{N} := \{x : D_I u(m, x) < 0\}$; $p_\varepsilon(x) := 1 - \varepsilon \mathbf{1}_{\mathcal{N}}(x)$, $\varepsilon \in (0, 1)$; and the corresponding measure m'_ε defined by (III.4.1). Then $(m - m'_\varepsilon)(dx, di) = (2i - 1)\varepsilon \mathbf{1}_{\mathcal{N}}(x)m(dx, 1)$, and thus

$$\begin{aligned} 0 &\leq \frac{1}{\varepsilon}[u(m) - u(m')] = \frac{1}{\varepsilon} \int_0^1 \int_{\mathbf{S}} \delta_m u(\lambda m + [1 - \lambda]m'_\varepsilon, x, i)(m - m'_\varepsilon)(dx, di)d\lambda \\ &= \int_0^1 \int_{\mathcal{N}} D_I u(\lambda m + [1 - \lambda]m'_\varepsilon, x)m(dx, 1)d\lambda. \end{aligned} \quad (\text{III.4.4})$$

Note that $\{\lambda m + [1 - \lambda]m'_\varepsilon : \lambda \in [0, 1], \varepsilon \in [0, 1]\} \subset \mathcal{P}_2(\mathbf{S})$ is compact, then $D_I u(\lambda m + [1 - \lambda]m'_\varepsilon, x)$ has quadratic growth in x , uniformly in λ, ε . Moreover, sending $\varepsilon \rightarrow 0$, since $m'_\varepsilon \rightarrow m$ and $D_I u$ is continuous in m , applying the dominated convergence theorem we obtain from (III.4.4) that $\int_{\mathcal{N}} D_I u(m, x)m(dx, 1) \geq 0$, which is possible only if $m(\mathcal{N}, 1) = 0$. That is, $D_I u(m, x) \geq 0$ for $m(\cdot, 1)$ -a.e. x . Since $D_I u$ is continuous in (m, x) and the set $\{m \in \mathcal{P}_2(\mathbf{S}) : \text{supp}(m(\cdot, 1)) = \mathbb{R}^d\}$ is dense in $\mathcal{P}_2(\mathbf{S})$, then one can easily show that $D_I u(m, x) \geq 0$ for all $(m, x) \in \mathcal{P}_2(\mathbf{S}) \times \mathbb{R}^d$. ■

III.4.2 The main results

Theorem III.4.4 *If V defined in (III.2.6) is in $C_2^{1,1}(\overline{\mathbf{Q}}_0)$, then it is a solution of (III.4.3).*

Moreover, for any $(t, m) \in [0, T) \times \mathcal{P}_2(\mathbf{S})$, $\mathbb{P}^* \in \mathcal{P}(t, m)$ is optimal for $V(t, m)$ if and only if, denoting $\mathbf{m}^* = \{m_s^* := \mathbb{P}_{Y_s}^*\}_{0 \leq s \leq T}$,

$$\begin{aligned} -(\mathbb{L}V + F)(s, m_s^*) &= 0, \quad V(s, m_s^*) = V(s, m_{s-}^*), \quad \text{for all } s \in [t, T], \\ D_I V(\tau, m_\tau^*, X_\tau) \mathbf{1}_{J_{[t, T]}^c}(\mathbf{m}^*)(\tau) &= 0, \quad \mathbb{P}^*\text{-a.s.}, \quad \text{where } m_\tau^* := m_s^*|_{s=\tau}. \end{aligned} \quad (\text{III.4.5})$$

Proof *Step 1:* We first prove that

$$-(\mathbb{L}V + F)(t, m) \geq 0, \quad D_I V(t, m, x) \geq 0, \quad \text{for all } (t, m, x) \in \mathbf{Q}_0 \times \mathbb{R}^d. \quad (\text{III.4.6})$$

Fix (t, m) . For any $m' \preceq m$, we may choose $\mathbb{P} \in \mathcal{P}(t, m)$ such that $\mathbb{P}_{Y_t} = m'$, and $I_s = I_t$, $s \in [t, T]$, \mathbb{P} -a.s. For $\delta \in (0, T - t)$, by DPP (III.2.8) we have

$$V(t, m) \geq \int_t^{t+\delta} F(s, \mathbb{P}_{Y_s}) ds + V(t + \delta, \mathbb{P}_{(X_{t+\delta}, I_t)}).$$

Send $\delta \rightarrow 0$, note that $\mathbb{P}_{(X_{t+\delta}, I_t)} \rightarrow \mathbb{P}_{(X_t, I_t)} = m'$. Then by the continuity of V we have $V(t, m) \geq V(t, m')$. Since $m' \preceq m$ is arbitrary, by Lemma III.4.3 we see that $D_I V \geq 0$.

To prove that $-(\mathbb{L}V + F)(t, m) \geq 0$, we consider $\mathbb{P} \in \mathcal{P}(t, m)$ such that $I_s = I_{t-}$, $s \in [t, T]$, \mathbb{P} -a.s. Apply Itô's formula (III.3.7) on $[t, t + \delta]$ under \mathbb{P} , we see that all the terms involving the jumps are equal to 0. Then by DPP (III.2.8) we have, denoting $m_s := \mathbb{P}_{Y_s}$,

$$0 \geq V(t + \delta, m_{(t+\delta)-}) - V(t, m) + \int_t^{t+\delta} F(s, m_s) ds = \int_t^{t+\delta} (\mathbb{L}u + F)(s, m_s) ds.$$

Note that $m_s \rightarrow m$ as $s \downarrow t$. Then by the continuity of $\mathbb{L}u + F$ one can easily see that $-(\mathbb{L}V + F)(t, m) \geq 0$.

Step 2: In this step we prove the equivalence of the optimality condition (III.4.5). First, if $\mathbb{P}^* \in \mathcal{P}(t, m)$ satisfies (III.4.5), applying Itô's formula (III.3.7) on $[t, T]$ we obtain immediately

$$V(t, m) = V(T, \mathbb{P}_{Y_T}^*) - \int_t^T \mathbb{L}V(s, \mathbb{P}_{Y_s}^*) ds = g(\mathbb{P}_{Y_T}^*) + \int_t^T F(s, \mathbb{P}_{Y_s}^*) ds.$$

As f has quadratic growth in x , locally uniformly in (t, m) , we may switch the integral and the expectation in the expression of F , and thus \mathbb{P}^* is optimal.

On the other hand, for any optimal $\mathbb{P}^* \in \mathcal{P}(t, m)$ such that $V(t, m) = \int_t^T F(s, \mathbb{P}_{Y_s}^*) ds + g(\mathbb{P}_{Y_T}^*)$. Denoting $m_s^* := \mathbb{P}_{Y_s}^*$, $s \geq t$, with $m = m_{t-}^*$, then by DPP (III.2.8) and Itô's formula (III.3.7) we have

$$0 = \int_t^T (\mathbb{L}V + F)(s, m_s^*) ds + \sum_{s \in J_{[t, T]}(\mathbf{m}^*)} [V(s, m_s^*) - V(s, m_{s-}^*)] + \mathbb{E} \left[\int_{J_{[t, T]}^c(\mathbf{m}^*)} D_I V(s, m_s^*, X_s) dI_s \right].$$

By Step 1 we have $V(s, m_s^*) \leq V(s, m_{s-}^*)$. Together with (III.4.6), we see that all the three terms in the right side above are nonpositive, then all of them should be 0:

$$\begin{aligned} (\mathbb{L}V + F)(s, m_s^*) = 0, \text{ a.e. } s \in [t, T]; \quad V(s, m_s^*) = V(s, m_{s-}^*), \text{ for all } s \in J_{[t, T]}^c(\mathbf{m}^*); \\ \int_{J_{[t, T]}^c(\mathbf{m}^*)} D_I V(s, m_s^*, X_s) dI_s = 0, \quad \mathbb{P}^* - \text{a.s.} \end{aligned} \quad (\text{III.4.7})$$

Since $\mathbb{L}V + F$ is continuous and, for $s \in J_{[t, T]}^c(\mathbf{m}^*)$, by definition $m_s^* = m_{s-}^*$ and hence $V(s, m_s^*) = V(s, m_{s-}^*)$, then the first line of (III.4.7) implies that the first line of (III.4.5) holds for all $s \in [t, T]$. Moreover, since τ is the only jump point of I , the second line of (III.4.7) is clearly equivalent to the second line of (III.4.5).

Step 3: Finally we complete the verification of (III.4.3). First by (III.2.6) $V(T, m) = g(m)$. Then, by Step 1, it remains to verify $\min_{m' \in C_V(t, m)} [-(\mathbb{L}V + F)(t, m')] = 0$. Note that $m_{t-}^* = m$ and set $s = t$ in the first line of (III.4.5), we have $m_t^* \in C_V(t, m)$. Thus

$$0 \leq \min_{m' \in C_V(t, m)} [-(\mathbb{L}V + F)(t, m')] \leq -(\mathbb{L}V + F)(t, m_t^*) = 0,$$

and therefore the equality holds. ■

Theorem III.4.5 (Verification) *Let $u \in C_2^{1,1}(\overline{\mathbf{Q}}_0)$ be a solution of (III.4.3). Then $u = V$.*

Proof We prove the theorem by using the obstacle equation (III.4.3) to construct an ε -optimal control for (III.2.6). We fix $m \in \mathcal{P}_2(\mathbf{S})$ and assume for simplicity that $t = 0$.

Step 1: We first prove that $u \geq V$. For an arbitrary $\mathbb{P} \in \mathcal{P}(0, m)$, we apply Itô's formula (III.3.7) and obtain: again denoting $\mathbf{m} = \{m_s := \mathbb{P}_{Y_s}\}$,

$$\begin{aligned} u(T, m_{T-}) = u(0, m) + \int_0^T \mathbb{L}u(s, m_s) ds \\ + \sum_{s \in J_{[0, T]}(\mathbf{m})} [u(s, m_s) - u(s, m_{s-})] + \mathbb{E}^{\mathbb{P}} \left[\int_{J_{[0, T]}^c(\mathbf{m})} D_I u(s, m_s, X_s) dI_s \right]. \end{aligned} \quad (\text{III.4.8})$$

By (III.4.3) and Lemma III.4.3 we have $u(s, m_s) \leq u(s, m_{s-})$. Then, (III.4.3) and (III.4.8) imply that

$$u(0, m) \geq u(T, m_{T-}) - \int_t^T \mathbb{L}u(s, m_s) ds \geq g(m_T) + \int_t^T F(s, m_s) ds.$$

Since $\mathbb{P} \in \mathcal{P}(0, m)$ is arbitrary, we obtain $u(0, m) \geq V(0, m)$.

Step 2: We now show that $u \leq V$. Let $n \geq 1$, $t_j := \frac{j}{n}T$, $j = 0, \dots, n$. We define $\mathbb{P}^n \in \mathcal{P}(0, m)$ and $m_s^n := \mathbb{P}_{Y_s}^n$ recursively such that $m_{0-}^n = m$, and for $j = 0, \dots, n-1$, thanks to Remark III.4.1,

$$m_{t_j}^n \in C_u(t_j, m_{t_{j-}}^n) \text{ s.t. } -(\mathbb{L}u + F)(t_j, m_{t_j}^n) = 0 \quad \text{and} \quad m_s^n \circ \mathbf{i}^{-1} = m_{t_j}^n \circ \mathbf{i}^{-1}, \quad s \in [t_j, t_{j+1}).$$

By the arguments of Proposition III.2.2 applied to $\mathcal{P}(0, m)$, and by (III.2.7), one can easily show that $\mathcal{W}_2(m_s^n, m_{t_j}^n) \leq \frac{C_m}{\sqrt{n}}$, $s \in [t_j, t_{j+1})$, for some constant $C_m > 0$ which may depend on m but is uniform on n . Moreover, by Proposition III.2.2 and the compactness of $[0, T]$, we see that the set $\{\mathbb{P}_{Y_s}, \mathbb{P}_{Y_{s-}} : s \in [0, T], \mathbb{P} \in \mathcal{P}(0, m)\}$ is compact. As $u \in C_2^{1,1}(\overline{\mathbf{Q}}_0)$, $\mathbb{L}u + F$ is continuous and then uniformly continuous on this set. Then, there exists a modulus of continuity function ρ such that

$$-(\mathbb{L}u + F)(s, m_s^n) = -(\mathbb{L}u + F)(s, m_s^n) + (\mathbb{L}u + F)(t_j, m_{t_j}^n) \leq \rho\left(\frac{T}{n} + \frac{C_m}{\sqrt{n}}\right), \quad s \in [t_j, t_{j+1}).$$

By Itô's formula (III.3.7), and noting that \mathbb{P}^n is constructed such that there is no contribution of the jump terms, we have

$$\begin{aligned} u(0, m) &= u(T, m_{T-}^n) - \int_0^T \mathbb{L}u(s, m_s^n) ds \\ &\leq g(m_T^n) + \int_t^T F(s, m_s^n) ds + T\rho\left(\frac{T}{n} + \frac{C_m}{\sqrt{n}}\right) \leq V(0, m) + T\rho\left(\frac{T}{n} + \frac{C_m}{\sqrt{n}}\right). \end{aligned}$$

Send $n \rightarrow \infty$, we obtain $u(0, m) \leq V(0, m)$. ■

III.4.3 Some discussions on optimal stopping policies

Proposition III.2.2 (i) guarantees that the mean field optimal stopping problem has an optimal randomized stopping strategy, i.e. a probability measure \mathbb{P}^* on Ω s.t. $g(\mathbb{P}_{Y_T}^*) + \int_t^T F(s, \mathbb{P}_{Y_s}^*) ds = V(t, m)$. A pure stopping strategy corresponds to the case where the conditional transition probability in (III.4.2) $p_s(\cdot) \in \{0, 1\}$ for all $s \in [t, T]$. In this case, the optimal stopping time is in closed-loop, i.e. τ is a stopping time w.r.t to the \mathbb{P}^* -augmented filtration of X , and the obstacle equation (III.4.3) reduces to:

$$\min_{A \in \mathcal{B}_u(t, m)} [-(\mathbb{L}u + F)(t, m^A)] = 0, \quad u(t, m) = \max_{A \in \mathcal{B}(\mathbb{R}^d)} u(t, m^A), \quad u|_{t=T} = g, \quad (t, m) \in \overline{\mathbf{Q}}_0 \quad (\text{III.4.9})$$

where $m^A := m \circ (\mathbf{x}, \mathbf{i}\mathbf{1}_A(\mathbf{x}))^{-1}$, and $\mathcal{B}_u(t, m) := \{A \in \mathcal{B}(\mathbb{R}^d) : u(t, m^A) = u(t, m)\}$.

We now discuss heuristically how to use the value function V to construct an optimal stopping time, provided $V \in C_2^{1,1}(\overline{\mathbf{Q}}_0)$. In light of (III.4.5) and recalling that $D_I V \geq 0$, introduce

$$K(t, m) := \{x \in \mathbb{R}^d : D_I V(t, m, x) = 0\}. \quad (\text{III.4.10})$$

Fix $(0, m_{0-})$. We set $m_{0-}^* := m_{0-}$ and construct \mathbf{m}^* for $V(0, m_{0-})$ in several steps.

Step 1. First, by (III.4.3) and Remark (III.4.1), there exists $m_0^* \in C_V(0, m_{0-}^*)$ such that $m_0^* \preceq m_{0-}^*$ and $(\mathbb{L}V + F)(0, m_0^*) = 0$. In particular, if $m_{0-}^* \circ \mathbf{x}^{-1}$ is continuous on $\{I_{0-} = 1\}$, there exists $A \in \mathcal{B}(\mathbb{R}^d)$ such that $I_0 = I_{0-} \mathbf{1}_{A^c}(X_0)$, and thus the optimal stopping time is a pure strategy at 0.

Step 2. Let \mathbb{P}^* be a weak solution to the following McKean-Vlasov SDE:

$$\mathbb{P}_{Y_0}^* = m_0^*, \quad X \text{ satisfies (III.2.3) and } I_t = I_0 \mathbf{1}_{\{X_s \in K(s, \mathbb{P}_{Y_s}^*), 0 < s \leq t\}}, \quad \mathbb{P}^*\text{-a.s.} \quad (\text{III.4.11})$$

Assume $m_t^* := \mathbb{P}_{Y_t}^*$ is continuous up to certain $t_1 > 0$. Then the optimal stopping time between $[0, t_1)$ is a pure strategy: $\tau = \inf\{t \geq 0 : X_t \notin K(t, m_t^*)\}$. Note that, since V is the value function, we should have $(\mathbb{L}V + F)(t, m_t^*) = 0$ for $t \in [0, t_1)$. We shall remark though, the McKean-Vlasov SDE (III.4.11) is path dependent and has discontinuous coefficients, so in general it is hard to solve. Moreover, the case that $t_1 = 0$ is even more difficult to solve.

Step 3. We have obtained $m_{t_1-}^*$ from Step 2. As in Step 1, we may find $m_{t_1}^* \in C_V(t_1, m_{t_1-}^*)$ at t_1 such that $m_{t_1}^* \preceq m_{t_1-}^*$ and $(\mathbb{L}V + F)(t_1, m_{t_1}^*) = 0$. Then following Step 2 again we can hopefully extend m^* to certain $t_2 > t_1$. Repeat the procedure, we may construct m^* on $[0, T]$.

We emphasize again that this procedure is just to illustrate the idea, in particular, it could be helpful for constructing approximate optimal stopping times, as we saw in Theorem III.4.5 Step 2. In general it is hard to realize this procedure, in fact, even the existence of classical solution is a very challenging task. Nevertheless, in Subsection III.5.3 below we will present an example where V is smooth and we can construct the τ explicitly. We also remark again that, the optimal stopping time in the continuous region constructed in Step 2 above is always a pure strategy, while in the jump region in Step 1 the optimal stopping time could be indeed mixed, but will also be a pure strategy when the distribution of the survival particles at that time is continuous.

III.5 Examples

III.5.1 Connection with standard optimal stopping

In this subsection we consider the case that b and σ do not depend on the $\mathcal{P}_2(\mathbf{S})$ -valued variable. For a measurable function φ , we define the optimal stopping problem

$$V(t, m) := \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \mathbb{E}^{\mathbb{P}}[\varphi(X_T)], \quad (t, m) \in \overline{\mathbf{Q}}_0, \quad (\text{III.5.1})$$

We also introduce $v(t, x) := V(t, \delta_{(x, 1)})$ which is related to the standard obstacle problem

$$\min\{-\partial_t v + \mathcal{L}v, v - \varphi\} = 0, \quad v(T, \cdot) = \varphi, \quad \text{where } \mathcal{L}v := b \cdot \partial_x v + \frac{1}{2}\sigma^2 : \partial_{xx}^2 v. \quad (\text{III.5.2})$$

To be consistent with Definition III.3.3, let $C_2^{1,1}([0, T] \times \mathbb{R}^d)$ denote the set of $v \in C^{1,2}([0, T] \times \mathbb{R}^d)$ such that $\partial_{xx}^2 v$ is bounded. This condition can be relaxed in this case though.

Proposition III.5.1 *Assume $v \in C_2^{1,1}([0, T] \times \mathbb{R}^d)$. Then:*

- (i) $V(t, m) = \int_{\mathbf{S}} (v(t, x)i + \varphi(x)(1 - i))m(dx, di)$, and V is a classical solution of the corresponding obstacle equation on the Wasserstein space;
- (ii) The probability measure \mathbb{P}^* s.t. $\tau = \inf\{s \geq t : v(s, X_s) = \varphi(X_s)\}$ on $\{I_{t-} = 1\}$, \mathbb{P}^* -a.s., is optimal for the problem $V(t, m)$. In particular, we see that τ is a pure stopping strategy under \mathbb{P}^* .

Proof Denote by u the right-hand side of the expression in (i). Then $u \in C_2^{1,1}(\overline{\mathbf{Q}}_0)$ with

$$\begin{aligned} \partial_t u(t, m) &= \int_{\mathbb{R}^d} \partial_t v(t, x)m(dx, 1), \quad \delta_m u(t, m, x, i) = v(t, x)i + \varphi(x)(1 - i), \\ \partial_x \delta_m u_1(t, m, x) &= \partial_x v(t, x), \quad \partial_{xx}^2 \delta_m u_1(t, m, x) = \partial_{xx}^2 v(t, x) \quad \text{for all } (x, i) \in \mathbf{S}. \end{aligned}$$

We then show that u is a solution of the equation (III.4.3). First, by (III.5.2),

$$D_I u = v - \varphi \geq 0 \quad \text{and} \quad -\mathbb{L}u(t, m) = \int_{\mathbb{R}^d} -\mathcal{L}v(t, x)m(dx, 1) \geq 0.$$

Defining $A_t := \{x : v(t, x) - \varphi(x) > 0\}$ and $m^{A_t} := m \circ (\mathbf{x}, \mathbf{i} \mathbf{1}_{A_t}(\mathbf{x}))^{-1}$, we have

$$u(t, m) - u(t, m^{A_t}) = \int_{A_t^c} [v(t, x) - \varphi(x)]m(dx, 1) = 0,$$

and therefore $m^{A_t} \in C_u(t, m)$. As $-\mathcal{L}v(t, x) = 0$, $x \in A_t$, we have $-\mathbb{L}u(t, m^{A_t}) = 0$. Thus, u is a solution of (III.4.3), and we deduce that $u = V$ by Theorem III.4.5.

To see that (ii) holds, notice that the flow $m_s^* := \mathbb{P}_{Y_s}^*$ is s.t. $m_s^* = (m_{s-}^*)^{A_s}$ for all $s \in [t, T]$. Then \mathbb{P}^* clearly satisfies (III.4.5), and thus is optimal for $V(t, m)$. \blacksquare

III.5.2 Convex functions of the expectation

Let $d = 1$, $\psi, h, \varphi : \mathbb{R} \rightarrow \mathbb{R}$, with φ convex. We consider the optimal stopping problem:

$$V(t, m) := \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \left[\mathbb{E}^{\mathbb{P}}[\psi(X_T)] + \varphi(\mathbb{E}^{\mathbb{P}}[h(X_T)]) \right]. \quad (\text{III.5.3})$$

This is an extension of the mean-variance optimal stopping problem. Introducing the convex dual $\varphi^*(\alpha) := \sup_{\beta \in \mathbb{R}} \{\alpha\beta - \varphi(\beta)\}$, we may write

$$V(t, m) = \sup_{\alpha \in \mathbb{R}} \left[-\varphi^*(\alpha) + V_{\alpha}(t, m) \right], \quad \text{with } V_{\alpha}(t, m) := \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \mathbb{E}^{\mathbb{P}}[\psi_{\alpha}(X_T)], \quad \psi_{\alpha} := \psi + \alpha h.$$

Assuming $u_{\alpha}(t, x) := V_{\alpha}(t, \delta_{(x,1)}) \in C_2^{1,1}([0, T] \times \mathbb{R})$, it follows from Proposition III.5.1 that

$$\begin{aligned} V_{\alpha}(t, m) &= \int_{\mathbf{S}} [u_{\alpha}(t, x)i + f_{\alpha}(x)[1 - i]] m(dx, di), \\ -\mathbb{L}V_{\alpha}(t, m^{A_t}) &= 0, \quad \text{where } A_t := \{x : u_{\alpha}(t, x) > \varphi(t, x)\}; \\ D_I V_{\alpha}(t, m, x) &= u_{\alpha}(t, x) - \varphi(t, x) = 0, \quad \text{when } u_{\alpha}(t, x) = \varphi(t, x). \end{aligned} \quad (\text{III.5.4})$$

Since α is one dimensional, it is not hard to find $\alpha^*(t, m)$ s.t. $V(t, m) = V_{\alpha^*(t, m)}(t, m) - \varphi^*(\alpha^*(t, m))$.

Moreover, fix (t, m) and let $\mathbb{P}^* \in \mathcal{P}(t, m)$ be the optimal measure for the problem $V_{\alpha^*(t, m)}(t, m)$, as constructed in the previous subsection. Then it is obvious that \mathbb{P}^* is optimal for $V(t, m)$ as well, and by Proposition (III.5.1) (ii), τ is an optimal stopping strategy under \mathbb{P}^* .

III.5.3 Construction of a smooth solution

In this subsection we construct an example where the obstacle problem indeed has a classical solution. First, set $b = 0$, $\sigma = 1$, and thus, for any $\mathbb{P} \in \mathcal{P}(t, m)$,

$$X_s = X_t + W_{\tau \wedge s}^{\mathbb{P}} - W_t^{\mathbb{P}} \quad \mathbb{P}\text{-a.s. on } \{I_{t-} = 1\}. \quad (\text{III.5.5})$$

Next, let $a \in C^1([0, T])$ and $\varphi \in C_2^{1,1}([0, T] \times \mathbb{R}_+)$ be positive functions such that

$$\partial_x \varphi(t, x) = 0 \text{ for } x \geq a_t, \quad \text{and} \quad \partial_x \varphi(t, x) > 0 \text{ for } x < a_t. \quad (\text{III.5.6})$$

One such example can be $\varphi(t, x) := e^{-[(a_t - x)^+]^3}$. Moreover, we introduce another positive function $\psi \in C^2(\mathbb{R})$ with bounded derivatives, and set

$$u_0(t, m) := [T - t]\varphi(t, v_0(m)), \quad \text{where} \quad v_0(m) := \int_{\mathbb{R}} \psi(x)m(dx, 1). \quad (\text{III.5.7})$$

Proposition III.5.2 *Under the above setting, $u_0 \in C_2^{1,1}([0, T] \times \mathcal{P}_2(\mathbf{S}))$, and u_0 is the classical solution to the obstacle problem (III.4.3) with*

$$F(t, m) := -\mathbb{L}u_0(t, m) - [v_0(m) - a_t]^+, \quad g := 0. \quad (\text{III.5.8})$$

We remark that this F may not take the specific form of (III.2.2), which is mainly motivated from applications but not really required for our theory. Since this example is just for illustration purpose of the theory, we content ourselves by allowing for this more general F . We emphasize again that in general it is hard to have classical solution for our obstacle problem, and therefore we shall investigate viscosity solutions in Chapter IV.

Proof First, by Definition III.3.1 one may easily verify: $\delta_m v_0(m, x, 1) = \psi(x)$, $\delta_m v_0(m, x, 0) = 0$, $\partial_t u_0(t, m) = [T-t]\partial_t \varphi(t, v_0(m)) - \varphi(t, v_0(m))$, $\delta_m u_0(t, m, x, 1) = [T-t]\partial_x \varphi(t, v_0(m))\psi(x)$, $\delta_m u_0(t, m, x, 0) = 0$. Then it is clear that $u_0 \in C_2^{1,1}(\overline{\mathbf{Q}}_0)$.

We now show that u_0 satisfies (III.4.3). Clearly, $u_0(T, \cdot) = 0 = g$, and

$$\begin{cases} D_I u_0(t, m, x) = [T-t]\partial_x \varphi(t, v_0(m))\psi(x) \geq 0, \\ -\mathbb{L}u_0(t, m) - F(t, m) = [v_0(m) - a_t]^+ \geq 0. \end{cases} \quad (\text{III.5.9})$$

In particular, $-(\mathbb{L}u_0 + F)(t, m) = 0$ when $v_0(m) \leq a_t$. Finally, when $v_0(m) > a_t$, combining (III.5.6) and (III.5.7), we have

$$C_{u_0}(t, m) = \{m' \preceq m : v_0(m') \in [a_t, v_0(m)]\}, \quad t < T. \quad (\text{III.5.10})$$

Set $m'_* \preceq m$ by (III.4.1) with $p(x) \equiv \frac{a_t}{v_0(m)}$. Then $m'_* \in C_{u_0}(t, m)$ with $v_0(m'_*) = a_t$. Therefore,

$$\min_{m' \in C_{u_0}(t, m)} -(\mathbb{L}u_0 + F)(t, m') \leq -(\mathbb{L}u_0 + F)(t, m'_*) = [v_0(m'_*) - a_t]^+ = 0.$$

This, together with (III.5.9), completes the proof. \blacksquare

In the rest of this subsection, we construct an optimal $\mathbb{P}^* \in \mathcal{P}(0, m_{0-})$ for the problem $V_0 := V(0, m_{0-}) = u_0(0, m_{0-})$. For simplicity we assume

$$T = 2, \quad \psi(x) := e^{-\frac{x^2}{2}}, \quad \text{and} \quad X_0 = 0, \quad I_{0-} = 1, \quad m_{0-} \text{ a.s.} \quad (\text{III.5.11})$$

We next specify the function a , which relies on two functions κ_i on $[0, T] \times \mathbb{R}$:

$$\begin{aligned} \kappa_0(t, x) &:= \mathbb{E}[\psi(x + W_t)], & \kappa_1(t, x) &:= \mathbb{E}[\psi(x + W_t)\mathbf{1}_{\{W_t^* < 1\}}], \\ a_t &:= \frac{1}{2}[\kappa_0(t, 0) + t^2(1-t)^2]\mathbf{1}_{[0,1]}(t) + \frac{1}{2}\mathbb{E}[\kappa_1(t-1, W_1)]\mathbf{1}_{(1,2]}(t), \end{aligned} \quad (\text{III.5.12})$$

where $W_t^* := \sup_{0 \leq s \leq t} W_s$. Recall Karatzas & Shreve [48, Chapter 2, Proposition 8.1] for the joint density of (W_t, W_t^*) , by direct calculations we have $0 \leq \partial_t \kappa_0(t, x) - \partial_t \kappa_1(t, x) \rightarrow 0$ as $t \rightarrow 0$. Then $\partial_t \kappa_1(0, x) = \partial_t \kappa_0(0, x)$, which implies that $a'_{1+} = h'_1 = a'_{1-}$, that is, $a \in C^1([0, T])$.

Proposition III.5.3 *Under the above setting, an optimal \mathbb{P}^* has the following structure:*

- (i) *At time 0, there is a massive stop with $\mathbb{P}^*(I_0 = 1) = \frac{1}{2}$.*
- (ii) *There is no stop during the time interval $(0, 1]$: $I_t = I_0$, $0 \leq t \leq 1$, \mathbb{P}^* -a.s.*
- (iii) *Particles stop continuously during the time interval $(1, 2)$:*

$$\tau = \inf\{t > 1 : X_t - X_1 \geq 1\} \wedge 2, \quad \mathbb{P}^*\text{-a.s. on } \{I_0 = 1\}. \quad (\text{III.5.13})$$

- (iv) *All the remaining particles stop at time 2.*

Proof (i) Note that in this case

$$v_0(m_{0-}) = \mathbb{E}^{m_{0-}}[\psi(X_0)I_{0-}] = \psi(0) = 1 > \frac{1}{2} = \frac{1}{2}\kappa_0(0, 0) = a_0. \quad (\text{III.5.14})$$

Then $-(\mathbb{L}u_0 + F)(0, m_{0-}) = v_0(m_{0-}) - a_0 > 0$, we have to stop some particles immediately. We may choose m_0^* such that $m_0^*(I_0 = 1) = \frac{1}{2}$, and then $t = 0$ is a jump point of m^* , and

$$v_0(m_0^*) = \mathbb{E}^{m_0^*}[\psi(X_0)I_0] = \frac{1}{2}, \quad \text{and thus} \quad -(\mathbb{L}u_0 + F)(0, m_0^*) = v_0(m_0^*) - a_0 = 0.$$

Moreover, since $v_0(m_0^*) \in [a_0, v_0(m_{0-})]$, by (III.5.10) we have $m_0^* \in C_{u_0}(0, m_{0-})$. This implies that $u_0(0, m_0^*) = u_0(0, m_{0-}^*)$, and then it follows from (III.4.5) that \mathbb{P}^* is optimal at $t = 0$.

(ii) For the \mathbb{P}^* specified in the proposition, we have $I_t = I_0$ and hence $X_t = W_t^{\mathbb{P}^*}$ on $\{I_0 = 1\}$, $0 \leq t \leq 1$, \mathbb{P}^* -a.s. By (III.5.12) we see that $v_0(m_t^*) = \frac{1}{2}\kappa_0(t, 0) \leq a_t$, which implies that $-(\mathbb{L}u_0 + F)(0, m_t^*) = 0$, $0 \leq t \leq 1$. Since no particle stops during this period, then by (III.4.5) again \mathbb{P}^* is optimal on $[0, 1]$.

(iii) We first note that, if we continue to keep all particles on $\{I_0 = 1\}$ alive after $t = 1$, then we will have $v_0(m_t) = \frac{1}{2}\kappa_0(t, 0) > a_t$ (since $\kappa_0 > \kappa_1$) and thus $-(\mathbb{L}u_0 + F)(0, m_t) > 0$, which is not optimal. So after $t = 1$, we start to stop particles, and our structure allows us to stop the particles continuously in the sense m_t^* is continuous in t . Indeed, by (III.5.5) and (III.5.13),

$$\tau = \inf\{t > 1 : W_t^{\mathbb{P}^*} - W_1^{\mathbb{P}^*} \geq 1\} \wedge 2, \quad \mathbb{P}^*\text{-a.s. on } \{I_0 = 1\}.$$

Then, for $t \in (1, 2)$,

$$\begin{aligned} v_0(m_t^*) &= \mathbb{E}^{\mathbb{P}^*} \left[\psi(X_t) I_t \right] = \mathbb{E}^{\mathbb{P}^*} \left[\psi(X_t) I_0 \mathbf{1}_{\{\tau > t\}} \right] \\ &= \mathbb{E}^{\mathbb{P}^*} \left[\psi(W_1^{\mathbb{P}^*} + W_t^{\mathbb{P}^*} - W_1^{\mathbb{P}^*}) \mathbf{1}_{\{I_0=1\}} \mathbf{1}_{\{\sup_{1 \leq s \leq t} [W_s^{\mathbb{P}^*} - W_1^{\mathbb{P}^*}] < 1\}} \right] \\ &= \frac{1}{2} \mathbb{E}^{\mathbb{P}^*} \left[\kappa_1(t-1, W_1^{\mathbb{P}^*}) \right] = a_t. \end{aligned}$$

Therefore, $-(\mathbb{L}u_0 + F)(0, m_t^*) = 0$, $1 < t < 2$.

Next, for any $1 < t < 2$, clearly m_t^* is continuous, and thus $u_0(t, m_t^*) = u_0(t, m_{t-}^*)$. Moreover, since $v_0(m_t^*) = a_t$, by (III.5.9) and (III.5.6) we have

$$D_I u_0(t, m_t^*, X_t) = [T - t] \partial_x \varphi(t, v_0(m_t^*)) \psi(X_t) = [T - t] \partial_x \varphi(t, a_t) \psi(X_t) = 0.$$

Then by (III.4.5) again we see that \mathbb{P}^* is optimal on $[1, 2)$.

(iv) This is required by our formulation of the problem. ■

Remark III.5.4 (i) For the \mathbb{P}^* in Proposition (III.5.3), m_t^* has two jumps, one at $t = 0$ and the other at $t = 2$. In particular, the stopping at $t = 0$ is randomized. Indeed, since $X_0 \equiv 0$ under m_{0-} , there is no $A \in \mathcal{B}(\mathbb{R})$ such that $\mathbb{E}^{m_{0-}}[\psi(X_0) \mathbf{1}_A(X_0)] = a_0$.

(ii) If X_0 has continuous distribution under m_{0-} , say with density $\rho_0(x)$, then it is possible to have pure stopping strategy. Indeed, let x_0 be a median of X_0 . Set T , ψ , I_{0-} , κ_0 , κ_1 as in (III.5.11) and (III.5.12), and modify the a in (III.5.12) as follows:

$$a_t := \left[\int_{-\infty}^{x_0} \kappa_0(t, x) \rho_0(x) dx + t^2(1-t)^2 \right] \mathbf{1}_{[0,1]}(t) + \int_{-\infty}^{x_0} \mathbb{E}[\kappa_1(t-1, x + W_1)] \rho_0(x) dx \mathbf{1}_{(1,2]}(t).$$

By the same arguments as in Proposition III.5.3, the following pure stopping strategy is optimal:

- At time 0, there is a massive stop for the particles $X_0 > x_0$: $I_0 = \mathbf{1}_{\{X_0 \leq x_0\}}$, \mathbb{P}^* -a.s.
- There is no stop during the time interval $(0, 1]$: $I_t = I_0$, $0 \leq t \leq 1$, \mathbb{P}^* -a.s.
- Particles stop continuously during the time interval $(1, 2)$ following (III.5.13).
- All the remaining particles stop at time 2. ■

III.6 Some extensions

III.6.1 Infinite horizon case

This subsection is dedicated to the case $T = +\infty$. For any $(t, m) \in \overline{\mathbf{Q}}_0$, let $\mathcal{P}(t, m)$ denote the set of \mathbb{P} such that $\mathbb{P}_{Y_{t-}} = m$ and (III.2.7) holds on $[t, \infty)$. We shall always assume

Assumption III.6.1 (i) *Assumption III.2.1 holds true on $[0, \infty)$;*

(ii) $\int_0^\infty \sup_{m \in \mathcal{P}_2(\mathbf{S})} |F(t, m)| dt < \infty$;

(iii) *For any (t, m) and $\mathbb{P} \in \mathcal{P}(t, m)$, $X_\infty := \lim_{t \rightarrow \infty} X_t$ exists, \mathbb{P} -a.s.*

We remark that one sufficient condition of (ii) above is that $|f(t, x, m)| \leq Ce^{-\lambda t}$ for some constants $C, \lambda > 0$, and a special case of (iii) is:

$$d = 1, \quad b = b_0 x, \quad \sigma = \sigma_0 x, \quad b_0 - \frac{1}{2} \sigma_0^2 < 0. \quad (\text{III.6.1})$$

That is, the unstopped process X^0 in (III.1.2) is a Geometric Brownian motion and $X_\infty^0 = 0$.

We also define $I_\infty := 0$. This allows the case $\tau = +\infty$, and guarantees that $\mathcal{P}(t, m)$ is compact. The infinite horizon optimal stopping problem then simply writes:

$$V(t, m) := \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \int_t^\infty F(s, \mathbb{P}_{Y_s}) ds + g(\mathbb{P}_{X_\infty}) \quad \text{for all } (t, m) \in \overline{\mathbf{Q}}_0. \quad (\text{III.6.2})$$

The corresponding obstacle equation on Wasserstein space is

$$\min_{m' \in C_u(t, m)} -[\mathbb{L}u + F](t, m') = 0, \quad D_I u(t, m, \cdot) \geq 0, \quad (t, m) \in \overline{\mathbf{Q}}_0, \quad (\text{III.6.3})$$

with boundary condition $u(\infty, \cdot) = g$.

Now by considering the problem on $[0, \infty]$, we see that all the definitions as well as all the results in the previous sections on the finite horizon remain true in the infinite horizon.

Remark III.6.2 In the infinite horizon, one may naturally consider the time homogeneous case, that is, b, σ, f do not depend on t . Then $V = V(m)$ is also time homogeneous, and thus (III.6.3) becomes an elliptic problem: recalling (III.3.6),

$$\min_{m' \in C_u(m)} - \left[\int_{\mathbb{R}^d} \mathcal{L}_x \delta_m u_1(m', x) m'(dx, 1) + F(m') \right] = 0, \quad D_I u(m, \cdot) \geq 0, \quad m \in \mathcal{P}_2(\mathbf{S}) \quad (\text{III.6.4})$$

with boundary condition $u = g$ on $\partial \mathcal{P}_2(\mathbf{S}) := \{m \in \mathcal{P}_2(\mathbf{S}) : m(\mathbb{R}^d, 1) = 0\}$. We leave the details to interested readers. ■

III.6.2 Mean field optimal stopping of a jump-diffusion

This last subsection is dedicated to an informal discussion about the case where (X, I) is a stopped jump-diffusion, i.e., $I_s = I_t - \mathbf{1}_{s < \tau}$ and

$$X_s = X_t + \int_t^s b(r, X_r, m_r) I_r dr + \int_t^s \sigma(r, X_r, m_r) I_r dW_r + \int_t^s \gamma(r, X_{r-}, m_{r-}) I_{r-} d\eta_r \quad (\text{III.6.5})$$

where η is a pure jump process with intensity $\lambda_s := \lambda_s(s, X_s, m_s)$ and whose jump size is defined by a distribution ν , and $\gamma_s := \gamma(s, X_{s-}, m_{s-})$ satisfies the usual conditions. We refer to Burzoni, Ignazio, Reppen & Soner [15], who characterized the mean field optimal control of a jump-diffusion by a dynamic programming equation (in the viscosity sense). The result of this section may be seen as a complement to the context of mean field optimal stopping. We consider the optimal stopping problem (III.2.6), where $\mathcal{P}(t, m)$ is the set of probability measures such that the canonical process (X, I) satisfies (III.6.5). Then, the value function still satisfies the DPP (III.2.8). In order to formally derive the corresponding dynamic programming equation, we need to find the differential operator associated with the dynamics (III.6.5), which follows from Itô's formula (III.3.4) in the present jump-diffusion case. Let $u \in C_2^{1,1}(\overline{\mathbf{Q}}_0)$. Observing that the discontinuities in the flow $\mathbf{m} = \{m_s\}$ are only due to I (as η has an intensity, hence no atoms), by shifting the jump at s to t as in (III.3.7), we have

$$u(s, m_{s-}) = u(t, m_{t-}) + \int_t^s \mathbb{L}u(r, m_r) dr + \sum_{r \in J_{[t,s]}^c(\mathbf{m})} [u(r, m_r) - u(r, m_{r-})] + \mathcal{J}_D,$$

where $\mathcal{J}_D := \mathbb{E}^{\mathbb{P}} \left[\sum_{r \in J_{[t,s]}^c(\mathbf{m})} (\delta_m u(r, m_r, X_r, I_r) - \delta_m u(t, m_r, X_{r-}, I_{r-})) \right]$.

We next compute \mathcal{J}_D . Denote $\varphi_r(\cdot) := \varphi(r, m_r, \cdot)$ for any function φ and $\Delta\eta_r := \eta_r - \eta_{r-}$. Note again that $J_{[t,s]}(\mathbf{m})$ is countable and thus η does not jump at $J_{[t,s]}(\mathbf{m})$, a.s. Then

$$\begin{aligned} \mathcal{J}_D &= \mathbb{E}^{\mathbb{P}} \left[\sum_{r \in J_{[t,s]}^c(\mathbf{m})} [\delta_m u_r(X_{r-} + \gamma_r(X_{r-})I_{r-} \Delta\eta_r, I_r) - \delta_m u_r(X_{r-}, I_{r-})] \right] \\ &= \mathbb{E}^{\mathbb{P}} \left[\sum_{r \in J_{[t,s]}^c(\mathbf{m})} [\delta_m u_r(X_{r-} + \gamma_r(X_{r-})I_{r-} \Delta\eta_r, I_r) - \delta_m u_r(X_{r-}, I_r)] \right] \\ &\quad + \mathbb{E}^{\mathbb{P}} \left[\sum_{r \in J_{[t,s]}^c(\mathbf{m})} [\delta_m u_r(X_{r-}, I_r) - \delta_m u_r(X_{r-}, I_{r-})] \right] \\ &= \mathbb{E}^{\mathbb{P}} \left[\int_t^s \int_{\mathbb{R}^d} [\delta_m u_r(X_r + y\gamma_r(X_r)I_r, I_r) - \delta_m u_r(X_r, I_r)] \nu(dy) \gamma_r(X_r) \lambda_r(X_r) I_r dr \right] \\ &\quad + \mathbb{E}^{\mathbb{P}} \left[\int_{J_{[t,s]}^c(\mathbf{m})} D_I u_r(X_r) dI_r \right], \end{aligned}$$

which implies that the differential operator corresponding to the dynamics (III.6.5) is

$$\begin{aligned} \mathbb{L}^{JD} u(t, m) &:= \mathbb{L}u(t, m) \\ &\quad + \int_{(\mathbb{R}^d)^2} [\delta_m u_1(t, m, x + y\gamma(t, m, x)) - \delta_m u_1(t, m, x)] \gamma \lambda(t, m, x) \nu(dy) m(dx, 1). \end{aligned}$$

Then, the dynamic programming equation corresponding to our problem is

$$\min_{m' \in C_u(t,m)} - (\mathbb{L}^{JD} u + F)(t, m') = 0, \quad D_I u(t, m, \cdot) \geq 0, \quad u(T, \cdot) = g, \quad (t, m) \in [0, T] \times \mathcal{P}_2(\mathbf{S}).$$

All the results of the previous sections can be adapted under appropriate assumptions.

III.7 Proof of Proposition III.2.2

We assume for simplicity $t = 0$ and fix $m \in \mathcal{P}_2(\mathbf{S})$. Let C_m denote a generic constant which may depend on T and m but independent of \mathbb{P} . We proceed in three steps.

Step 1. We first prove the following uniform integrability: denoting $X_T^* := \sup_{0 \leq s \leq T} |X_s|$,

$$\sup_{\mathbb{P} \in \mathcal{P}(0,m)} \mathbb{E}^{\mathbb{P}} [|X_T^*|^2] \leq C_m, \quad \lim_{R \rightarrow \infty} \sup_{\mathbb{P} \in \mathcal{P}(0,m)} \mathbb{E}^{\mathbb{P}} [|X_T^*|^2 \mathbf{1}_{\{X_T^* \geq R\}}] = 0. \quad (\text{III.7.1})$$

Indeed, for any $\mathbb{P} \in \mathcal{P}(0, m)$, first by standard arguments we derive from (III.2.7) that

$$\mathbb{E}^{\mathbb{P}} [|X_T^*|^2] \leq C \mathbb{E}^m [1 + |X_0|^2] \leq C_m.$$

In particular, this implies that the set $\{\mathbb{P}_{Y_s} : \mathbb{P} \in \mathcal{P}(0, m), 0 \leq s \leq T\}$ is bounded under \mathcal{W}_2 . Then, for any $p > 2$, by (III.2.7) again we have $\mathbb{E}^{\mathbb{P}} [|X_T^*|^p | \mathcal{F}_0] \leq C_{m,p} [1 + |X_0|^p]$, \mathbb{P} -a.s., where $C_{m,p}$ may depend on p as well, but is still independent of \mathbb{P} . Now for any $R > 0$,

$$\begin{aligned} \mathbb{E}^{\mathbb{P}} [|X_T^*|^2 \mathbf{1}_{\{X_T^* \geq R\}}] &\leq \mathbb{E}^{\mathbb{P}} [|X_T^*|^2 \mathbf{1}_{\{1+|X_0| \geq \sqrt{R}\}}] + \mathbb{E}^{\mathbb{P}} [|X_T^*|^2 \mathbf{1}_{\{\frac{X_T^*}{1+|X_0|} \geq \sqrt{R}\}}] \\ &\leq \mathbb{E}^{\mathbb{P}} [|X_T^*|^2 \mathbf{1}_{\{1+|X_0| \geq \sqrt{R}\}}] + \frac{1}{\sqrt{R}} \mathbb{E}^{\mathbb{P}} \left[\frac{|X_T^*|^3}{1+|X_0|} \right] \\ &= \mathbb{E}^{\mathbb{P}} \left[\mathbb{E}^{\mathbb{P}} [|X_T^*|^2 | \mathcal{F}_0] \mathbf{1}_{\{1+|X_0| \geq \sqrt{R}\}} \right] + \frac{1}{\sqrt{R}} \mathbb{E}^{\mathbb{P}} \left[\frac{\mathbb{E}^{\mathbb{P}} [|X_T^*|^3 | \mathcal{F}_0]}{1+|X_0|} \right] \\ &\leq C_{m,2} \mathbb{E}^m [[1 + |X_0|^2] \mathbf{1}_{\{1+|X_0| \geq \sqrt{R}\}}] + \frac{C_{m,3}}{\sqrt{R}} \mathbb{E}^m [1 + |X_0|^2]. \end{aligned}$$

Notice that the right side above does not depend on \mathbb{P} , then it clearly implies (III.7.1).

Step 2. We next show that $\mathcal{P}(0, m)$ is closed under the weak convergence. Let $\{\mathbb{P}^n\}_{n \geq 1} \subset \mathcal{P}(0, m)$ converge weakly to some \mathbb{P}^∞ . Since $\mathbb{P}_{(X_0, I_{0-})}^n = m$ for all n , we have $\mathbb{P}_{Y_{0-}}^\infty = m$. Then it suffices to show that the processes M, N in (III.2.4) are \mathbb{P}^∞ -martingales on $[0, T]$. We shall report only the detailed argument for M , as it is immediately adapted to N .

Notice that the support of \mathbb{P}^∞ is separable under the Skorokhod distance d_{SK} , as a subspace of the separable metric space Ω . Then it follows from the Skorokhod's representation

theorem, see Billingsley [10, Theorem 6.7], that there exists a probability space $(\Omega^0, \mathcal{F}^0, \mathbb{P}^0)$ and processes $\{Y^n := (X^n, I^n)\}_{n \geq 1}$ and $Y^\infty := (X^\infty, I^\infty)$ defined on this space such that,

$$\mathbb{P}_Y^n = \mathbb{P}_{Y^n}^0 \text{ for all } n \leq \infty, \text{ and } d_{SK}(Y^n, Y^\infty) \xrightarrow[n \rightarrow \infty]{} 0, \mathbb{P}^0 - \text{a.s.} \quad (\text{III.7.2})$$

For all $n \geq 1$, the \mathbb{P}^n -martingale property of M translates to:

$$\mathbb{E}^{\mathbb{P}^0}[(M_s^n - M_t^n)\psi(Y_{\cdot \wedge t}^n)] = 0 \quad \text{for all } \psi \in C_b(\Omega) \text{ and } 0 \leq t \leq s \leq T, \quad (\text{III.7.3})$$

with $M_s^n = X_t^n - \int_t^s b(r, X_r^n, \mathbb{P}_{Y_r^n}^0) I_r^n dr$ and $C_b(\Omega)$ the set of \mathbb{R}^d -valued bounded continuous functions on Ω . Moreover, for $r \in [t, T]$, by the Lipschitz continuity of b we have

$$|b(r, X_r^n, \mathbb{P}_{Y_r^n}^0) - b(r, X_r^\infty, \mathbb{P}_{Y_r^\infty}^0)| \leq C[|X_r^n - X_r^\infty| + \mathcal{W}_2(\mathbb{P}_{Y_r^n}^0, \mathbb{P}_{Y_r^\infty}^0)].$$

Send $n \rightarrow \infty$, by (III.7.2) we have $|X_r^n - X_r^\infty| \rightarrow 0, \mathbb{P}^0$ -a.s. and $\mathbb{P}_{Y_r^n}^0 \rightarrow \mathbb{P}_{Y_r^\infty}^0$ weakly. Then, by the 2-uniform integrability (III.7.1) of $\{\mathbb{P}_{Y_r^n}^0\}_{n \geq 1}$, we have $\mathcal{W}_2(\mathbb{P}_{Y_r^n}^0, \mathbb{P}_{Y_r^\infty}^0) \rightarrow 0$, see Carmona & Delarue [21, Vol. I, Theorem 5.5]. Thus

$$b(r, X_r^n, \mathbb{P}_{Y_r^n}^0) \xrightarrow[n \rightarrow \infty]{} b(r, X_r^\infty, \mathbb{P}_{Y_r^\infty}^0), \mathbb{P}^0\text{-a.s.}$$

Moreover, as b is Lipschitz and $\{X^n\}_{n \geq 1}$ are uniformly integrable (as the 2-uniform integrability of (III.7.1) implies the 1-uniform integrability), then $\{M^n\}_{n \geq 1}$ are uniformly integrable. The convergence for the Skorokhod distance also implies the convergence of $I_{\cdot \wedge t}^n$ to $I_{\cdot \wedge t}^\infty$. This allows to take the limit in (III.7.3) as $\psi \in C_b(\Omega)$, hence $\mathbb{E}^{\mathbb{P}^0}[(M_s^\infty - M_t^\infty)\psi(Y_{\cdot \wedge t}^\infty)] = 0$. By the arbitrariness of $\psi \in C_b(\Omega)$, this proves M^∞ is a \mathbb{P}^0 -martingale, or equivalently that M is a \mathbb{P}^∞ -martingale.

Step 3. We now show that $\mathcal{P}(0, m)$ is compact under \mathcal{W}_2 . Let $\{\mathbb{P}^n\}_{n \geq 1} \subset \mathcal{P}(0, m)$. First, by the first estimate in (III.7.1) and noticing that I is bounded by 1, one can easily obtain a uniform bound for the conditional variation of Y under all \mathbb{P}^n , then by Meyer & Zheng [59, Theorem 4] we see that $\{\mathbb{P}^n\}_{n \geq 1}$ is relatively weakly compact, namely there exists a weakly convergent subsequence. By Step 2, without loss of generality we assume the whole sequence $\mathbb{P}^n \rightarrow \mathbb{P}^\infty \in \mathcal{P}(0, m)$ weakly. Moreover, by the second estimate in (III.7.1) $\{\mathbb{P}^n\}_{n \geq 1}$ is 2-uniformly integrable, then it follows from Carmona & Delarue [21, Vol. I, Theorem 5.5] again that $\lim_{n \rightarrow \infty} \mathcal{W}_2(\mathbb{P}^n, \mathbb{P}^\infty) = 0$. This proves the compactness of $\mathcal{P}(0, m)$.

Finally, since g is upper-semicontinuous, the above compactness implies the existence of optimal \mathbb{P}^* for the mean field optimal stopping problem (III.2.6). \blacksquare

III.8 Proof of Theorem III.3.2

Let $\Xi_{\mathbf{m}}$ denote the convex hull of $\{m_s, m_{s-} : 0 \leq s \leq T\}$:

$$\Xi_{\mathbf{m}} := \left\{ \lambda m_{s'} + (1 - \lambda) m_{t'} : 0 \leq \lambda \leq 1, 0 \leq s \leq t \leq T, s' = s, s-, t' = t, t- \right\} \subset \mathcal{P}_2(\mathbb{R}^d).$$

We first show that $\Xi_{\mathbf{m}}$ is compact. Indeed, for any (λ_n, s'_n, t'_n) , there exists a convergent subsequence and we may assume without loss of generality that $(\lambda_n, s_n, t_n) \rightarrow (\lambda, s, t)$. By considering different cases, one can easily show that, possibly along a subsequence, for some s', t' we have $\lambda_n m_{s'_n} + (1 - \lambda) m_{t'_n} \rightarrow \lambda m_{s'} + (1 - \lambda) m_{t'} \in \Xi_{\mathbf{m}}$, thus $\Xi_{\mathbf{m}}$ is compact.

Denote $\Delta Y_s := Y_s - Y_{s-}$ and $Y_t^D := \sum_{0 < s \leq t} \Delta Y_s$. By (III.3.3) it is clear that

$$\mathbb{E}[|Y_T^*|^2 + \|Y^D\|_T^2] < \infty, \quad \text{where } Y_T^* := \sup_{0 \leq s \leq T} |Y_s|, \quad \|Y^D\|_t := \sum_{0 < s \leq t} |\Delta Y_s|. \quad (\text{III.8.1})$$

For $n \geq 1$, set $\Delta t := \frac{T}{n}$, $t_i := i\Delta t$, $i = 0, \dots, n$. Then, for each i ,

$$u(t_{i+1}, m_{t_{i+1}}) - u(t_i, m_{t_i}) = \int_{t_i}^{t_{i+1}} \partial_t u(s, m_{t_{i+1}}) ds + \int_0^1 \mathbb{E}[\xi_{t_{i+1}}^\lambda] d\lambda, \quad (\text{III.8.2})$$

where $\xi_{t_{i+1}}^\lambda := \delta_m u(t_i, m_{t_i}^\lambda, Y_{t_{i+1}}) - \delta_m u(t_i, m_{t_i}^\lambda, Y_{t_i})$, $m_{t_i}^\lambda := \lambda m_{t_i} + [1 - \lambda] m_{t_{i+1}}$.

By the standard Itô's formula:

$$\xi_{t_{i+1}}^\lambda = \int_{t_n}^{t_{n+1}} [\Gamma_s^{2,\lambda} \cdot dY_s^c + \frac{1}{2} \Gamma_s^{3,\lambda} : d\langle Y^c \rangle_s] + \int_{(t_n, t_{n+1}]} \Gamma_s^{4,\lambda} dY_s^D,$$

where $\Gamma_s^1 := \partial_t u(s, m_{t_{i+1}})$, $\Gamma_s^{2,\lambda} := \partial_y \delta_m u(t_i, m_{t_i}^\lambda, Y_s)$, $\Gamma_s^{3,\lambda} := \partial_{yy}^2 \delta_m u(t_i, m_{t_i}^\lambda, Y_s)$,

$$\Gamma_s^{4,\lambda} := \int_0^1 \partial_y \delta_m u(t_i, m_{t_i}^\lambda, \theta Y_s + [1 - \theta] Y_{s-}) d\theta.$$

Note that $m_{t_{i+1}}, m_{t_i}^\lambda \in \Xi_{\mathbf{m}}$, by the growth conditions in Definition III.3.1 we have

$$|\Gamma_s^1| \leq C, \quad |\Gamma_s^{2,\lambda}| \leq C[1 + |Y_s|], \quad |\Gamma_s^{3,\lambda}| \leq C, \quad |\Gamma_s^{4,\lambda}| \leq C[1 + |Y_s| + |Y_{s-}|]. \quad (\text{III.8.3})$$

Then

$$\begin{aligned} \mathbb{E} \left[\left(\int_{t_n}^{t_{n+1}} \Gamma_s^{2,\lambda} (\Gamma_s^{2,\lambda})^\top : d\langle M^c \rangle_s \right)^{\frac{1}{2}} \right] &\leq C \mathbb{E} \left[\left([1 + |Y_T^*|^2] \langle M^c \rangle_T \right)^{\frac{1}{2}} \right] \\ &\leq C \mathbb{E} \left[1 + |Y_T^*|^2 + \langle M^c \rangle_T \right] < \infty. \end{aligned}$$

This implies $\int_0^1 \mathbb{E} \left[\int_{t_n}^{t_{n+1}} \Gamma_s^{2,\lambda} \cdot dM_s^c \right] d\lambda = 0$, and thus

$$\begin{aligned} u(T, m_T) &= u(0, m_0) + \int_0^T \Gamma_s^1 ds \\ &\quad + \int_0^1 \mathbb{E} \left[\int_0^T [\Gamma_s^{2,\lambda} \cdot dA_s^c + \Gamma_s^{3,\lambda} : d\langle Y^c \rangle_s] + \int_{(0,T]} \Gamma_s^{4,\lambda} dY_s^D \right] d\lambda. \end{aligned}$$

Fix λ, s , and send $n \rightarrow \infty$. By the regularity of u we have: denoting $m_s^\lambda := \lambda m_{s-} + [1-\lambda]m_s$,

$$\begin{aligned}\Gamma_s^1 &\rightarrow \partial_t u(s, m_s), & \Gamma_s^{2,\lambda} &\rightarrow \partial_y \delta_m u(s, m_s^\lambda, Y_s), & \Gamma_s^{3,\lambda} &\rightarrow \partial_{yy}^2 \delta_m u(s, m_s^\lambda, Y_s), \\ \Gamma_s^{4,\lambda} &\rightarrow \int_0^1 \partial_y \delta_m u(s, m_s^\lambda, \theta Y_s + [1-\theta]Y_{s-}) d\theta, & & a.s.\end{aligned}$$

By (III.3.3), (III.8.1), and (III.8.3), we may apply the dominated convergence theorem to obtain

$$\begin{aligned}u(T, m_T) &= u(0, m_0) + \int_0^T \partial_t u(s, m_s) ds + \int_0^1 \mathbb{E} \left[\int_0^T [\partial_y \delta_m u(s, m_s^\lambda, Y_s) \cdot dA_s^c \right. \\ &\quad \left. + \partial_{yy}^2 \delta_m u(s, m_s^\lambda, Y_s) : d\langle Y^c \rangle_s] + \int_{(0,T]} \int_0^1 \partial_y \delta_m u(s, m_s^\lambda, \theta Y_s + [1-\theta]Y_{s-}) d\theta dY_s^D \right] d\lambda.\end{aligned}$$

Since A^c and $\langle Y^c \rangle$ are continuous, and m_s has at most countably many jumps, then

$$\begin{aligned}u(T, m_T) &= u(0, m_0) + \int_0^T \partial_t u(s, m_s) ds + \mathcal{J}_D \\ &\quad + \mathbb{E} \left[\int_0^T [\partial_y \delta_m u(s, m_s, Y_s) \cdot dA_s^c + \partial_{yy}^2 \delta_m u(s, m_s, Y_s) : d\langle Y^c \rangle_s] \right],\end{aligned}\tag{III.8.4}$$

$$\text{where } \mathcal{J}_D := \mathbb{E} \left[\int_0^1 \int_{(0,T]} \int_0^1 \partial_y \delta_m u(s, m_s^\lambda, \theta Y_s + [1-\theta]Y_{s-}) d\theta dY_s^D d\lambda \right].$$

It remains to compute \mathcal{J}_D . First, by Fubini's theorem,

$$\begin{aligned}\mathcal{J}_D &= \mathbb{E} \left[\sum_{s \in (0,T]} \int_0^1 \int_0^1 \partial_y \delta_m u(s, m_s^\lambda, \theta Y_s + [1-\theta]Y_{s-}) d\theta d\lambda \Delta Y_s \right] = \mathbb{E} \left[\sum_{s \in (0,T]} \Delta \delta_m u_s \right] \\ &\quad \text{where } \Delta \delta_m u_s := \int_0^1 [\delta_m u(s, m_s^\lambda, Y_s) - \delta_m u(s, m_s^\lambda, Y_{s-})] d\lambda.\end{aligned}\tag{III.8.5}$$

Note that $(0, T] = J_{(0,T]}(\mathbf{m}) \cup J_{(0,T]}^c(\mathbf{m})$. Since $J_{(0,T]}(\mathbf{m})$ is countable, then

$$\mathbb{E} \left[\sum_{s \in J_{(0,T]}(\mathbf{m})} \Delta \delta_m u_s \right] = \sum_{s \in J_{(0,T]}(\mathbf{m})} \mathbb{E} [\Delta \delta_m u_s] = \sum_{s \in J_{(0,T]}(\mathbf{m})} [\delta_m u(s, m_s) - \delta_m u(s, m_{s-})],\tag{III.8.6}$$

where the second equality is due to (III.3.2). Next, for $s \in J_{(0,T]}^c(\mathbf{m})$, we have $m_s^\lambda = m_s$, $0 \leq \lambda \leq 1$. Then

$$\mathbb{E} \left[\sum_{s \in J_{(0,T]}^c(\mathbf{m})} \Delta \delta_m u_s \right] = \mathbb{E} \left[\sum_{s \in J_{(0,T]}^c(\mathbf{m})} [\delta_m u(s, m_s, Y_s) - \delta_m u(s, m_s, Y_{s-})] \right].\tag{III.8.7}$$

We emphasize that, since $J_{(0,T]}^c(\mathbf{m})$ is uncountable, unlike in (III.8.6) we cannot switch the order of \mathbb{E} and $\sum_{s \in J_{(0,T]}^c(\mathbf{m})}$ at above. Now plug (III.8.6), (III.8.7) into (III.8.5), and then plug (III.8.5) into (III.8.4), we complete the proof. \blacksquare

Chapter IV

Viscosity solutions for obstacle problems on Wasserstein space

IV.1 Introduction

In Chapter III, we characterized the so-called mean field optimal stopping problem by a dynamic programming equation on the Wasserstein space, that we call obstacle equation on Wasserstein space by analogy with the equation corresponding to the standard optimal stopping problem (see e.g. El Karoui [41] or Shiryaev [70]). More precisely, we proved that the value function of our optimization problem is the unique solution of the obstacle equation on the Wasserstein space, provided it has $C^{1,2}$ regularity (in an appropriate sense). We note that, besides its obvious connection with multiple stopping problems over a large interacting particle system, this obstacle equation provides a convenient tool for many time inconsistent optimal stopping problems. We also remark that, our mean field optimal stopping problem has quite different structure than the mean field games of optimal stopping.

However, as in the case of the standard optimal stopping problems, one can rarely expect a classical solution for the obstacle equations. In particular, the infinite dimensionality of the space of measures makes the regularity requirement even harder to meet. Our goal of this chapter is thus to develop a viscosity solution theory for the obstacle problem on the Wasserstein space, which as well-known requires much weaker regularities.

There have been some serious efforts on viscosity solutions of nonlinear partial differential equations on the Wasserstein space. We first mention the paper by Cardaliaguet & Quin-

campoix [18], which considered a first order Hamilton-Jacobi-Isaacs equation on Wasserstein space arising from deterministic zero-sum games with random initial conditions. The comparison principle for viscosity solutions was established by combining the doubling variables argument with Ekeland’s variational principle. In the context of mean field control problems in a path dependent setting, Wu & Zhang [77] proposed a notion of viscosity solutions for parabolic equations on the Wasserstein space, inspired from Ekren, Keller, Ren, Touzi & Zhang [37, 39, 40, 68]. Note that the natural idea which consists in taking Wasserstein balls for the viscosity neighborhood (as in Carmona & Delarue [20]) leads to serious difficulties as the Wasserstein ball is in general not compact. Instead, [77] restricted the viscosity neighborhood of some point (t, μ) (where t is a time and μ a measure) to the compact set of all possible laws of the controlled state process starting from this point. Another remarkable work by Burzoni, Ignazio, Reppen & Soner [15], in the context of mean field control of jump-diffusions, restricted the viscosity neighborhood in another way, so as to guarantee compactness. They proved a comparison result by the doubling variables argument. To do this, they succeeded in constructing a smooth metric which serves as a test function, but unfortunately restricts the scope of the method to the case when the coefficients of the controlled dynamics do not depend on the space variable.

We shall follow the approach of [77]. We consider the joint law of $(X_{\tau \wedge t}, \mathbf{1}_{\{\tau \geq t\}})$ as the variable of the value function, where X is the state process and τ is the stopping time. As in [77] we define viscosity solutions by using the set of such laws over all stopping times τ . This neighborhood set of laws, for a given initial condition, is compact under Wasserstein distance and thus is desirable for the viscosity theory. We show that, under natural conditions, the value function of the mean field optimal stopping problem is indeed the unique viscosity solution of the corresponding obstacle equation on Wasserstein space. We shall also establish the stability and the comparison principle for the viscosity solutions. To prove the latter, one key ingredient is a smooth mollifier for continuous functions on the Wasserstein space, introduced by Mou & Zhang [61]. However, to obtain some uniform estimates of the smooth mollifier which are needed in our proof of comparison principle, as in [61] we require the coefficients to be Lipschitz continuous under the 1-Wasserstein distance, rather than the more natural 2-Wasserstein distance.

As applications of our viscosity theory, we invest several time inconsistent optimal stopping problems, including problems related to mean variance, probability distortion, and expected shortfall. By considering the law (instead of the value) of the stopped state pro-

cess as the variable, we show that the value functions are indeed the unique viscosity solution to the corresponding obstacle equation on the Wasserstein space. Moreover, our results can be easily extended to the infinite horizon case.

The chapter is organized as follows. In §IV.2, we present the mean field optimal stopping problem, the corresponding dynamic programming equation and some of its elementary properties. §IV.3 is the main section, where we propose our definition of viscosity solutions and prove the main results. §IV.4 is devoted to several applications. Finally, we prove some technical results in the appendix.

IV.2 The obstacle problem on Wasserstein space

IV.2.1 Formulation

Let $T < \infty$ be fixed, and $\Omega := C^0([-1, T], \mathbb{R}^d) \times \mathbb{I}^0([-1, T])$ the canonical space, where:

- $C^0([-1, T], \mathbb{R}^d)$ is the set of continuous paths from $[-1, T]$ to \mathbb{R}^d , constant on $[-1, 0)$;
- $\mathbb{I}^0([-1, T])$ is the set of non-increasing and càdlàg maps from $[-1, T]$ to $\{0, 1\}$, constant on $[-1, 0)$, and ending with value 0 at T .

We equip Ω with the Skorokhod distance, under which it is a Polish space. Note that the choice of the extension to -1 is arbitrary, the extension of time to the left of the origin is only needed to allow for an immediate stop at time $t = 0$.

We denote $Y := (X, I)$ the canonical process, with state space $\mathbf{S} := \mathbb{R}^d \times \{0, 1\}$, its canonical filtration $\mathbb{F} = \mathbb{F}^Y = (\mathcal{F}_t)_{t \in [-1, T]}$, and the corresponding jump time of the survival process I :

$$\tau := \inf\{t \geq 0 : I_t = 0\}, \text{ so that } I_t := I_{0-} \mathbf{1}_{t < \tau} \text{ for all } t \in [-1, T].$$

By the càdlàg property of I , τ is an \mathbb{F} -stopping time.

Let $(b, \sigma, f) : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbf{S}) \rightarrow \mathbb{R}^d \times \mathbb{R}^{d \times d} \times \mathbb{R}$ with σ taking values in non-negative matrices, and $g : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$. In the following assumption, which will always be in force throughout the chapter, $\mathcal{P}_2(\mathbf{S})$ is equipped with the \mathcal{W}_2 distance.

Assumption IV.2.1 (i) b, σ are continuous in t , and uniformly Lipschitz continuous in (x, m) .

(ii) f is Borel measurable and has quadratic growth in $x \in \mathbb{R}^d$, and

$$F(t, m) := \int_{\mathbb{R}^d} f(t, x, m) m(dx, 1) \text{ is continuous on } [0, T] \times \mathcal{P}_2(\mathbf{S}). \quad (\text{IV.2.1})$$

(iii) g is upper-semicontinuous and locally bounded; and extended to $\mathcal{P}_2(\mathbf{S})$ by $g(m) := g(m(\cdot, \{0, 1\}))$.

Introduce the dynamic value function

$$V(t, m) := \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \int_t^T F(r, \mathbb{P}_{Y_r}) dr + g(\mathbb{P}_{Y_T}), \quad (t, m) \in [0, T] \times \mathcal{P}_2(\mathbf{S}). \quad (\text{IV.2.2})$$

Here $\mathcal{P}(t, m)$ is the set of probability measures \mathbb{P} on (Ω, \mathcal{F}_T) s.t. $\mathbb{P}_{Y_{t-}} = m$, the paths $s \in [-1, t) \rightarrow Y_s$ are constants, \mathbb{P} -a.s., and the processes:

$$M := X - \int_t^\cdot b(r, X_r, \mathbb{P}_{Y_r}) I_r dr \quad \text{and} \quad M \cdot M^\top - \int_t^\cdot \sigma^2(r, X_r, \mathbb{P}_{Y_r}) I_r dr \quad (\text{IV.2.3})$$

are \mathbb{P} -martingales on $[t, T]$, that is, for some \mathbb{P} -Brownian motion $W^\mathbb{P}$,

$$X_s = X_t + \int_t^s b(r, X_r, \mathbb{P}_{Y_r}) I_r dr + \sigma(r, X_r, \mathbb{P}_{Y_r}) I_r dW_r^\mathbb{P}, \quad I_s = I_{t-} \mathbf{1}_{s < \tau}, \quad \mathbb{P} - \text{a.s.}$$

A special element of $\mathcal{P}(t, m)$ is $\bar{\mathbb{P}} = \bar{\mathbb{P}}^{t, m}$ under which X is unstopped. That is,

$$X_s = X_t + \int_t^s b(r, X_r, \bar{\mathbb{P}}_{Y_r}) I_r dr + \int_t^s \sigma(r, X_r, \bar{\mathbb{P}}_{Y_r}) I_r dW_r^{\bar{\mathbb{P}}}, \quad I_s = I_{t-} \mathbf{1}_{[t, T)}(s), \quad \bar{\mathbb{P}}\text{-a.s.} \quad (\text{IV.2.4})$$

Note that $Y = Y_{\wedge \tau}$, and in particular $Y_T = Y_\tau$, \mathbb{P} -a.s. Moreover, from the definition of F in (IV.2.1), we have $\int_t^T F(r, \mathbb{P}_{Y_r}) dr = \mathbb{E}^\mathbb{P} \int_t^\tau f(r, X_r, \mathbb{P}_{Y_r}) dr$.

We recall from Chapter III that $\mathcal{P}(t, m)$ is compact under the Wasserstein distance \mathcal{W}_2 , and thus existence holds for the mean field optimal stopping problem (IV.2.2). Furthermore, we have the dynamic programming principle (DPP for short): for any $s \in [t, T]$,

$$V(t, m) = \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \int_t^s F(r, \mathbb{P}_{Y_r}) dr + V(s, \mathbb{P}_{Y_{s-}}) = \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \int_t^s F(r, \mathbb{P}_{Y_r}) dr + V(s, \mathbb{P}_{Y_s}). \quad (\text{IV.2.5})$$

IV.2.2 Differential calculus

We next recall some differential calculus tools on the Wasserstein space. We say that a function $u : \mathcal{P}_2(\mathbf{S}) \rightarrow \mathbb{R}$ has a functional linear derivative $\delta_m u : \mathcal{P}_2(\mathbf{S}) \times \mathbf{S} \rightarrow \mathbb{R}$ if

$$u(m') - u(m) = \int_0^1 \int_{\mathbf{S}} \delta_m u(\lambda m' + (1 - \lambda)m, y) (m' - m)(dy) d\lambda \quad \text{for all } m, m' \in \mathcal{P}_2(\mathbf{S}),$$

and $\delta_m u(m, \cdot)$ has quadratic growth, locally uniformly in m , so as to guarantee integrability in the last expression. As in [74], we denote

$$\delta_m u_i(t, m, x) := \delta_m u(t, m, x, i) \quad \text{for } i \in \{0, 1\}, \quad D_I u := \delta_m u_1 - \delta_m u_0, \quad (\text{IV.2.6})$$

and we introduce the measure flow generator of X

$$\begin{aligned} \mathbb{L}u(t, m) &:= \partial_t u(t, m) + \int_{\mathbb{R}^d} \mathcal{L}_x \delta_m u_1(t, m, x) m(dx, 1), \\ \text{where } \mathcal{L}_x \delta_m u_1 &:= b \cdot \partial_x \delta_m u_1 + \frac{1}{2} \sigma^2 : \partial_{xx}^2 \delta_m u_1. \end{aligned} \quad (\text{IV.2.7})$$

Throughout this chapter, we denote by

$$\mathbf{Q}_t := [t, T] \times \mathcal{P}_2(\mathbf{S}), \quad \text{and} \quad \overline{\mathbf{Q}}_t := [t, T] \times \mathcal{P}_2(\mathbf{S}), \quad t \in [0, T].$$

Definition IV.2.2 *Let $C_2^{1,1}(\overline{\mathbf{Q}}_t)$ be the set of functions $u : \overline{\mathbf{Q}}_t \rightarrow \mathbb{R}$ s.t.*

- $\partial_t u, \delta_m u, \partial_x \delta_m u_1, \partial_{xx}^2 \delta_m u_1$ exist, and are continuous in all variables,
- $\partial_{xx}^2 \delta_m u_1$ is bounded in x , locally uniformly in (t, m) .

The following Itô's formula is due to [74, §3]: for any $u \in C_2^{1,1}(\overline{\mathbf{Q}}_0)$ and $\mathbb{P} \in \mathcal{P}(0, m)$:

$$\begin{aligned} u(T, m_{T-}) &= u(0, m) + \int_0^T \mathbb{L}u(s, m_s) ds \\ &+ \sum_{s \in J_{[0,T]}(\mathbf{m})} [u(s, m_s) - u(s, m_{s-})] + \mathbb{E}^{\mathbb{P}} \left[\int_{J_{[0,T]}^c(\mathbf{m})} D_I u(s, m_s, X_s) dI_s \right], \end{aligned} \quad (\text{IV.2.8})$$

where $\mathbf{m} := \{m_s := \mathbb{P}_{Y_s}\}_{s \in [-1, T]}$, $J_{\mathbb{T}}(\mathbf{m}) := \{s \in \mathbb{T} : m_s \neq m_{s-}\}$, for all subset $\mathbb{T} \subset [0, T]$, and $J_{\mathbb{T}}^c(\mathbf{m})$ its complement set in \mathbb{T} .

IV.2.3 The dynamic programming equation

Given two probability measures $m, m' \in \mathcal{P}_2(\mathbf{S})$, we say that $m' \preceq m$ if

$$m'(dx, 1) = p(x)m(dx, 1), \quad \text{and} \quad m'(dx, 0) = [1 - p(x)]m(dx, 1) + m(dx, 0), \quad (\text{IV.2.9})$$

for some measurable $p : \mathbb{R}^d \rightarrow [0, 1]$, i.e. $m'(dx, 1)$ is obtained from m by randomly stopping a proportion $1 - p(x)$ of the surviving particles. In our context, $m_{t-} = \mathbb{P}_{(X_t, I_{t-})}$ and $m_t = \mathbb{P}_{(X_t, I_t)}$, with $\mathbb{P} \in \mathcal{P}(t, m)$, so that $m_t \preceq m_{t-}$ with conditional transition probability $p(x) = p_t(x) := \mathbb{P}(I_t = 1 \mid X_t = x, I_{t-} = 1)$.

The following property (proved in Appendix IV.5) will be often used in this chapter:

Lemma IV.2.3 *For an arbitrary $m \in \mathcal{P}_2(\mathbf{S})$:*

- (i) *the set $\{m' : m' \preceq m\}$ is compact,*
- (ii) *any compact subset $\mathcal{K}(m) \subset \{m' : m' \preceq m\}$ has a smallest element for \preceq , i.e., there exists $\bar{m} \in \mathcal{K}(m)$ such that for all $m' \in \mathcal{K}(m)$, we have $m' \preceq \bar{m}$ implies that $m' = \bar{m}$.*

The dynamic programming equation corresponding to our mean field optimal stopping problem is the infinitesimal counterpart of the dynamic programming principle (IV.2.5), and is defined by

$$\min \left\{ \min_{m' \in C_u(t, m)} [-(\mathbb{L}u + F)(t, m')], (\mathbb{D}_I u)_*(t, m) \right\} = 0, \quad (t, m) \in \mathbf{Q}_0, \quad (\text{IV.2.10})$$

with boundary condition $u|_{t=T} = g$. Here the function $(\mathbb{D}_I u)_*$ is the LSC envelope of

$$\mathbb{D}_I u : (t, m) \mapsto \inf_{x \in \text{Supp}(m(\cdot, 1))} D_I u(t, m, x),$$

which is upper semicontinuous, but not continuous, in general, and the set

$$C_u(t, m) := \{m' \preceq m : u(t, m') \geq u(t, m)\}, \quad (t, m) \in \mathbf{Q}_0,$$

indicates the set of positions in \mathbf{Q}_0 which improve u by stopping the corresponding particles.

For the purpose of the present chapter, we note that this equation is slightly different from the obstacle equation (III.4.3) introduced in Chapter III:

- if u is a classical solution of (IV.2.10), then it is nondecreasing for \preceq (see [74, Lemma 4.3]), and thus $C_u(t, m)$ is characterized by an equality, as in [74];

- despite the remaining differences, the two equations define the same solution, but this does not seem to have an immediate proof; we emphasize however that the equivalence between these two equations is a direct consequence of our uniqueness result in [74, Theorem 4.5], and the comparison Theorem IV.3.11 below.

Our objective in this chapter is to develop a notion of viscosity solution for this equation which bypasses the strong regularity requirements of classical solutions. As usual, we start by introducing the notions of the sub- and supersolutions.

Definition IV.2.4 *Let $u \in C_2^{1,1}(\overline{\mathbf{Q}_0})$.*

(i) *u is a classical supersolution of (IV.2.10) if*

$$\min \{ -(\mathbb{L}u + F), \mathbb{D}_I u \}(t, m) \geq 0, \quad \forall (t, m) \in \mathbf{Q}_0. \quad (\text{IV.2.11})$$

(ii) *u is a classical subsolution of (IV.2.10) if*

$$\min \{ -(\mathbb{L}u + F), (\mathbb{D}_I u)_* \}(t, m) \leq 0, \quad \forall (t, m) \in \mathbf{Q}_0 \text{ s.t. } C_u(t, m) = \{m\}. \quad (\text{IV.2.12})$$

(iii) *u is a classical solution of (IV.2.10) if it is a classical supersolution and subsolution.*

IV.3 Viscosity solutions

IV.3.1 Definition and consistency

For $\delta > 0$ and $(t, m) \in \mathbf{Q}_0$, we introduce the neighborhood

$$\mathcal{N}_\delta(t, m) := \{(s, \tilde{m}) : s \in [t, t + \delta], \mathbb{P} \in \mathcal{P}(t, m), \tilde{m} \in \{\mathbb{P}_{Y_{s-}}, \mathbb{P}_{Y_s}\}\}.$$

Note that, as the closure of a càdlàg $\mathcal{P}_2(\mathbf{S})$ -valued graph, $\mathcal{N}_\delta(t, m)$ is compact, by the compactnesses of $[t, t + \delta]$, $\mathcal{P}(t, m)$ and $\{(\mathbb{P}_{Y_{s-}}, \mathbb{P}_{Y_s})\}_{s \in [t, t + \delta]}$ for any $\mathbb{P} \in \mathcal{P}(t, m)$.

For a locally bounded function $u : \overline{\mathbf{Q}}_0 \rightarrow \mathbb{R}$, we introduce its LSC and USC envelopes relatively to $\mathcal{P}(t, m)$, u_* and u^* respectively:

$$u_*(t, m) := \liminf_{(s, \tilde{m}) \rightarrow (t, m)} u(s, \tilde{m}), \quad u^*(t, m) := \limsup_{(s, \tilde{m}) \rightarrow (t, m)} u(s, \tilde{m}), \quad \text{for all } (t, m) \in \overline{\mathbf{Q}}_0,$$

where the limits are taken on all sequences $\{(t_n, m_n)\}_{n \geq 1}$ converging to (t, m) , with $(t_n, m_n) \in \mathcal{N}_{T-t}(t, m)$ for all n . We then introduce the sets of test functions

$$\begin{aligned} \overline{\mathcal{A}}u(t, m) &:= \left\{ \varphi \in C_2^{1,1}(\overline{\mathbf{Q}}_t) : (\varphi - u_*)(t, m) = \max_{\mathcal{N}_\delta(t, m)} (\varphi - u_*) \text{ for some } \delta > 0 \right\}, \\ \underline{\mathcal{A}}u(t, m) &:= \left\{ \varphi \in C_2^{1,1}(\overline{\mathbf{Q}}_t) : (\varphi - u^*)(t, m) = \min_{\mathcal{N}_\delta(t, m)} (\varphi - u^*) \text{ for some } \delta > 0 \right\}. \end{aligned}$$

Definition IV.3.1 *Let $u : \mathbf{Q}_0 \rightarrow \mathbb{R}$ be locally bounded.*

(i) *u is a viscosity supersolution of (IV.2.10) if, for any $(t, m) \in \mathbf{Q}_0$,*

$$u_*(t, m) \geq u_*(t, m'), \quad \forall m' \preceq m, \quad \text{and} \quad -(\mathbb{L}\varphi + F)(t, m) \geq 0, \quad \forall \varphi \in \overline{\mathcal{A}}u(t, m). \quad (\text{IV.3.1})$$

(ii) *u is a viscosity subsolution of (IV.2.10) if, for any $(t, m) \in \mathbf{Q}_0$ s.t. $C_{u^*}(t, m) = \{m\}$,*

$$\min\{-(\mathbb{L}\varphi + F), (\mathbb{D}_I\varphi)_*\}(t, m) \leq 0, \quad \forall \varphi \in \underline{\mathcal{A}}u(t, m). \quad (\text{IV.3.2})$$

(iii) *u is a viscosity solution of (IV.2.10) if it is a viscosity supersolution and subsolution.*

Remark IV.3.2 Without loss of generality, we may assume that the maximum in the definition of $\overline{\mathcal{A}}u(t, m)$ is strict. Indeed, for $\varphi \in \overline{\mathcal{A}}u(t, m)$, we set

$$\tilde{\varphi}(s, \tilde{m}) := \varphi(s, \tilde{m}) - (s - t)^2 - (\tilde{m}(\mathbb{R}^d, 1) - m(\mathbb{R}^d, 1))^2 \quad \text{for all } (s, \tilde{m}) \in \mathbf{Q}_t.$$

It is obvious that $\tilde{\varphi} \in C_2^{1,1}(\overline{\mathbf{Q}}_t)$. As $\tilde{\varphi}(s, \tilde{m}) = \tilde{\varphi}(t, m)$ if and only if $s = t$ and $\tilde{m} = m$ (since $\tilde{m}(\mathbb{R}^d, 1) = m(\mathbb{R}^d, 1)$, and observing that in this case $\tilde{m} = \mathbb{P}_{Y_t} \preceq m$ for some $\mathbb{P} \in \mathcal{P}(t, m)$),

we deduce that $\tilde{\varphi} \in \overline{\mathcal{A}}u(t, m)$ and the maximum is strict. Moreover, simple computations show that $\partial_t \tilde{\varphi}(t, m) = \partial_t \varphi(t, m)$ and $\mathbb{L}\tilde{\varphi}(t, m) = \mathbb{L}\varphi(t, m)$. An analogous statement holds for $\underline{\mathcal{A}}u(t, m)$. \blacksquare

Our first result shows the consistency between classical and viscosity solutions.

Theorem IV.3.3 *Let $u \in C_2^{1,1}(\overline{\mathbf{Q}}_0)$. Then u is a classical sub- (resp. super)solution of (IV.2.10) if, and only if, it is a viscosity sub- (resp. super)solution of (IV.2.10).*

Proof (i) Let $(t, m) \in \mathbf{Q}_0$. If u is a viscosity super/subsolution, then given its smoothness we have $u \in \overline{\mathcal{A}}u(t, m) \cap \underline{\mathcal{A}}u(t, m)$, and we immediately deduce that u is a classical super/subsolution. In particular, by [74, Lemma 4.3], u is nondecreasing for \preceq implies that $D_I u \geq 0$.

(ii) Assume u is a classical supersolution of (IV.2.10). By (IV.2.11) we see that $D_I u \geq 0$, then by [74, Lemma 4.3] again we see that u is nondecreasing for \preceq . Now let $\varphi \in \overline{\mathcal{A}}u(t, m)$ with corresponding δ . Introduce $\psi := \varphi - u$ and let $\overline{\mathbb{P}} \in \mathcal{P}(t, m)$ be defined by (IV.2.4) s.t. X is unstopped under $\overline{\mathbb{P}}$. By definition of $\overline{\mathcal{A}}u(t, m)$, we have $\psi(t, m) \geq \psi(s, \overline{\mathbb{P}}_{Y_s})$ for all $s \in [t, t + \delta]$. Applying Itô's formula (IV.2.8), since the jump terms are equal to zero under $\overline{\mathbb{P}}$, we obtain: $-\frac{1}{\delta} \int_t^{t+\delta} \mathbb{L}\psi(s, \overline{\mathbb{P}}_{Y_s}) ds \geq 0$. Send $\delta \rightarrow 0$, by the continuity of $s \mapsto \overline{\mathbb{P}}_{Y_s}$ we have $-\mathbb{L}\psi(t, m) \geq 0$, hence $-(\mathbb{L}\varphi + F)(t, m) \geq -(\mathbb{L}u + F)(t, m) \geq 0$ by the supersolution property of u .

Assume now that u is a classical subsolution. Let $\varphi \in \underline{\mathcal{A}}u(t, m)$ with corresponding δ , and assume that $(\mathbb{D}_I)_* \varphi(t, m) > 0$ and $C_u(t, m) = \{m\}$. By definition of $\underline{\mathcal{A}}u(t, m)$, we have

$$[\varphi - u](t, m) \leq [\varphi - u](s, \mathbb{P}_{Y_s}), \quad \text{for all } s \in [t, t + \delta], \mathbb{P} \in \mathcal{P}(t, m). \quad (\text{IV.3.3})$$

Set $s = t$ in (IV.3.3), then it follows from the arbitrariness of $\mathbb{P} \in \mathcal{P}(t, m)$ that $[\varphi - u](t, m) \leq [\varphi - u](t, m')$ for all $m' \preceq m$. Following the arguments of [74, Lemma 4.3], we deduce from above that $D_I[\varphi - u](t, m, \cdot) \leq 0$, and therefore $(\mathbb{D}_I u)_*(t, m) \geq (\mathbb{D}_I \varphi)_*(t, m) > 0$. The subsolution property of u then implies that $-(\mathbb{L}u + F)(t, m) \leq 0$. Using Itô's formula (IV.2.8) under $\overline{\mathbb{P}}$ again on $[t, t + \delta]$, we get from (IV.3.3) that $-(\mathbb{L}\varphi + F)(t, m) \leq -(\mathbb{L}u + F)(t, m) \leq 0$. \blacksquare

IV.3.2 Some regularity results

In this subsection, we present some regularity results which will be used in the rest of this section. Since our main focus is the viscosity properties, we postpone their proofs to

Appendix [IV.5](#).

Lemma IV.3.4 *Under Assumption [IV.2.1](#), the value function V is USC.*

Theorem IV.3.5 (i) *Assume f and g are uniformly continuous in (t, x, m) , under \mathcal{W}_2 for m , then V is continuous on $\overline{\mathbf{Q}_0}$, under \mathcal{W}_2 for m .*

(ii) *Assume further that b, σ are uniformly Lipschitz continuous in m under \mathcal{W}_1 , and f, g are uniformly continuous in m under \mathcal{W}_1 , then V is also continuous in m under \mathcal{W}_1 .*

Even for the standard optimal stopping problems, one can hardly expect the value function to be smooth. We next establish a regularity result for the value function when X is unstopped. For $(t, m) \in \mathbf{Q}_0$, let $\bar{\mathbb{P}}^{t,m} \in \mathcal{P}(t, m)$ be as by [\(IV.2.4\)](#), and define

$$U(t, m) := g(\bar{\mathbb{P}}_{Y_T}^{t,m}) + \int_t^T F(r, \bar{\mathbb{P}}_{Y_r}^{t,m}) dr. \quad (\text{IV.3.4})$$

Lemma IV.3.6 *For $\varphi = b, \sigma, f, g$, assume φ is continuous in t and $\partial_x \varphi, \delta_m \varphi, \partial_x \delta_m \varphi, \partial_{xx}^2 \delta_m \varphi$ exist and are continuous and bounded and that, for $\varphi = b, \sigma$, all their derivatives of φ are Lipschitz up to order 2. Then $U \in C^{1,2}(\overline{\mathbf{Q}_0})$ with bounded $\partial_x \delta_m U, \partial_{xx} \delta_m U$ and in particular $U \in C_2^{1,1}(\overline{\mathbf{Q}_0})$. Moreover, if b, σ, f, g are uniformly Lipschitz continuous in m under \mathcal{W}_1 with a Lipschitz constant L , then U is uniformly Lipschitz continuous in m under \mathcal{W}_1 with a Lipschitz constant C_L .*

Finally, we introduce a smooth mollifier for functions on the Wasserstein space.

Lemma IV.3.7 (i) *Let $U : \mathcal{P}_2(\mathbf{S}) \rightarrow \mathbb{R}$ be continuous. There exists $\{U_n\}_{n \geq 1}$ in $C^\infty(\mathcal{P}_2(\mathbf{S}))$ such that $\lim_{n \rightarrow \infty} \sup_{m \in \mathcal{M}} |U_n(m) - U(m)| = 0$ for any compact set $\mathcal{M} \subset \mathcal{P}_2(\mathbf{S})$;*

(ii) *Let $U : \mathcal{P}_1(\mathbf{S}) \rightarrow \mathbb{R}$ be continuous under \mathcal{W}_1 . There exists $\{U_n\}_{n \geq 1}$ in $C^\infty(\mathcal{P}_2(\mathbf{S})) \cap C^0(\mathcal{P}_1(\mathbf{S}))$ such that $\lim_{n \rightarrow \infty} \sup_{m \in \mathcal{M}} |U_n(m) - U(m)| = 0$ for any compact set $\mathcal{M} \subset \mathcal{P}_1(\mathbf{S})$;*

(iii) *Assume further that U is Lipschitz continuous under \mathcal{W}_1 , then we may choose $\{U_n\}_{n \geq 1}$ to be Lipschitz continuous under \mathcal{W}_1 , uniformly in n .*

The mollifier is adopted from Mou & Zhang [\[61\]](#). Note that the extension of the state space from \mathbb{R}^d in [\[61\]](#) to \mathbf{S} here is straightforward. We remark that if U is Lipschitz continuous under \mathcal{W}_2 , in general the Lipschitz continuity of U_n under \mathcal{W}_2 is not uniform in n .

IV.3.3 Viscosity property

We first need a simple lemma whose proof is postponed to Appendix IV.5.

Lemma IV.3.8 (i) *Let $v : \mathcal{P}_2(\mathbf{S}) \rightarrow \mathbb{R}$ be LSC, and $m \in \mathcal{P}_2(\mathbf{S})$ s.t. $v(m) \geq v(m')$ for all $m' \preceq m$ with continuous conditional transition probability. Then $v(m) \geq v(m')$ for all $m' \preceq m$.*

(ii) *Let $\varphi \in C^0(\overline{\mathbf{Q}}_0, \mathbb{R})$ admit a continuous linear functional derivative $\delta_m \varphi$. Assume we have $(\mathbb{D}_I \varphi)_*(t, m) > 0$ for some $(t, m) \in \mathbf{Q}_0$. Then φ is nondecreasing for \preceq in a neighborhood of (t, m) .*

Theorem IV.3.9 *The value function V is a viscosity solution of (IV.2.10).*

Proof First, by Lemma IV.3.4, V inherits the local boundedness of g .

(i) We first verify the viscosity supersolution property. Fix (t, m) and $\varphi \in \overline{AV}(t, m)$. We may assume w.l.o.g. that $[V_* - \varphi](t, m) = 0$. Let $\delta > 0$ and $(t_n, m_n)_{n \geq 1} \in \mathcal{N}_\delta(t, m)$ converging to (t, m) s.t. $V(t_n, m_n) \xrightarrow{n \rightarrow \infty} V_*(t, m)$, and denote $\eta_n := [V - \varphi](t_n, m_n) \geq 0$, as $V \geq V_*$. Thus, we have $\eta_n \xrightarrow{n \rightarrow \infty} 0$. By the DPP (IV.2.5), we have

$$\begin{aligned} \eta_n + \varphi(t_n, m_n) &= V(t_n, m_n) \geq \int_{t_n}^{s_n} F(r, \bar{\mathbb{P}}_{Y_r}^{m_n}) dr + V(s_n, \bar{\mathbb{P}}_{Y_{s_n}}^{m_n}) \\ &\geq \int_{t_n}^{s_n} F(r, \bar{\mathbb{P}}_{Y_r}^{m_n}) dr + V_*(s_n, \bar{\mathbb{P}}_{Y_{s_n}}^{m_n}) \geq \int_{t_n}^{s_n} F(r, \bar{\mathbb{P}}_{Y_r}^{m_n}) dr + \varphi(s_n, \bar{\mathbb{P}}_{Y_{s_n}}^{m_n}), \end{aligned}$$

where $\bar{\mathbb{P}}^{m_n} := \bar{\mathbb{P}}^{t_n, m_n} \in \mathcal{P}(t_n, m_n)$ is defined by (IV.2.4) such that X is unstopped, and $s_n := t_n + h_n$ with $h_n := \sqrt{\eta_n} \vee n^{-1}$. Thus, by Itô's formula, the above gives $h_n + \frac{1}{h_n} \int_{t_n}^{s_n} -(\mathbb{L}\varphi + F)(r, \bar{\mathbb{P}}_{Y_r}^{m_n}) dr \geq 0$. Send $n \rightarrow \infty$, since $h_n \rightarrow 0$, we obtain $-(\mathbb{L}\varphi + F)(t, m) \geq 0$.

We now prove the remaining part of the supersolution property. Let $m' \preceq m$ with transition probability p . By Lemma IV.3.8 (i), we may assume without loss of generality that p is continuous. For all $n \geq 1$, define $m'_n \preceq m_n$ as the measure obtained from m_n by applying the same p . Given the continuity of p and the compactness $\mathcal{N}_\delta(t, m)$, we see by (IV.2.9) that $\mathcal{W}_2(m'_n, m') \xrightarrow{n \rightarrow \infty} 0$. Let $\bar{\mathbb{P}}^{m_n, m'_n} \in \mathcal{P}(t_n, m_n)$ be s.t. $\bar{\mathbb{P}}_{Y_{t_n}}^{m_n, m'_n} = m'_n$, and $I_s = I_{t_n}$, $\bar{\mathbb{P}}^{m_n, m'_n}$ -a.s. for all $s \geq t_n$. By (IV.2.5),

$$V(t_n, m_n) \geq \int_{t_n}^s F(r, \bar{\mathbb{P}}_{Y_r}^{m_n, m'_n}) dr + V(s, \bar{\mathbb{P}}_{Y_s}^{m_n, m'_n}), \text{ for all } s \geq t_n \text{ and } n \geq 1. \quad (\text{IV.3.5})$$

Take $s = t_n$ and $\liminf_{n \rightarrow \infty}$ in (IV.3.5), we obtain $V_*(t, m) \geq V_*(t, m')$ as $V(t_n, m_n) \rightarrow V_*(t, m)$.

(ii) We next verify the viscosity subsolution property. Let (t, m) and $\varphi \in \underline{\mathcal{A}}V(t, m)$ be s.t. $C_{V^*}(t, m) = \{m\}$ and $(\mathbb{D}_I\varphi)_*(t, m) > 0$. We may assume w.l.o.g. that $[V^* - \varphi](t, m) = 0$. Let $\delta > 0$ and $(t_n, m_n)_{n \geq 1} \in \mathcal{N}_\delta(t, m)$ converging to (t, m) such that $V(t_n, m_n) \xrightarrow[n \rightarrow \infty]{} V^*(t, m)$, and denote $-\eta_n := [V - \varphi](t_n, m_n) \leq 0$. Thus $\eta_n \xrightarrow[n \rightarrow \infty]{} 0$. For $n \geq 1$, since g is USC and $\mathcal{P}(t_n, m_n)$ is compact, there exists $\mathbb{P}^{n,*} \in \mathcal{P}(t_n, m_n)$ s.t. $V(t_n, m_n) = \int_{t_n}^T F(r, \mathbb{P}_{Y_r}^{n,*}) dr + g(\mathbb{P}_{Y_T}^{n,*})$. By DPP, we have

$$V(t_n, m_n) \geq \int_{t_n}^{s_n} F(r, \mathbb{P}_{Y_r}^{n,*}) dr + V(s_n, \mathbb{P}_{Y_{s_n}}^{n,*}) \geq \int_{t_n}^T F(r, \mathbb{P}_{Y_r}^{n,*}) dr + g(\mathbb{P}_{Y_T}^{n,*}),$$

where $s_n := t_n + h_n$ with $h_n := \sqrt{\eta_n} \vee n^{-1}$, and thus,

$$V(t_n, m_n) = \int_{t_n}^{s_n} F(r, \mathbb{P}_{Y_r}^{n,*}) dr + V(s_n, \mathbb{P}_{Y_{s_n}}^{n,*}). \quad (\text{IV.3.6})$$

Note that $\mathbb{P}^{n,*} \in \mathcal{P}(t_n, m_n) \subset \mathcal{P}(t, m)$ for all n , and $\mathcal{P}(t, m)$ is compact, we may extract a subsequence (still denoted the same) s.t. $\mathbb{P}^{n,*} \xrightarrow[n \rightarrow \infty]{} \mathbb{P}^*$, for some $\mathbb{P}^* \in \mathcal{P}(t, m)$. As the trajectories $r \mapsto \mathbb{P}_{Y_r}^{n,*}$ are càdlàg and $s_n \downarrow t$, this implies $\mathcal{W}_2(\mathbb{P}_{Y_{s_n}}^{n,*}, m^*) \xrightarrow[n \rightarrow \infty]{} 0$, where $m^* := \mathbb{P}_{Y_t}^* \preceq m$ as $\mathbb{P}^* \in \mathcal{P}(t, m)$. Thus, taking the $\limsup_{n \rightarrow \infty}$ in (IV.3.6) and recalling $V(t_n, m_n) \rightarrow V^*(t, m)$, we have $V^*(t, m) \leq V^*(t, m^*)$. As $C_{V^*}(t, m) = \{m\}$, we obtain $m^* = m$. Moreover, (IV.3.6) also implies that

$$-\eta_n + \varphi(t_n, m_n) \leq \int_{t_n}^{s_n} F(r, \mathbb{P}_{Y_r}^{n,*}) dr + \varphi(s_n, \mathbb{P}_{Y_{s_n}}^{n,*}) \quad \text{for all } n \geq 1. \quad (\text{IV.3.7})$$

Let $\mathcal{B}_{\mathcal{W}_2}(m, \delta)$ denote the \mathcal{W}_2 ball centered in m , with radius δ . By Lemma IV.3.8 (ii), the fact that $(\mathbb{D}_I\varphi)_*(t, m) > 0$ implies that φ is (strictly) increasing for \preceq on $[t, t+\delta) \times \mathcal{B}_{\mathcal{W}_2}(m, \delta)$, for a possibly smaller $\delta > 0$. By convergence to (t, m) , we have $\{\mathbb{P}_{Y_r}^{n,*}, t \leq r \leq s_n\} \subset \mathcal{B}_{\mathcal{W}_2}(m, \delta)$ for n large. Then $D_I\varphi(r, \mathbb{P}_{Y_r}^{n,*}, \cdot) \geq 0$ and $\varphi(r, \mathbb{P}_{Y_r}^{n,*}) \geq \varphi(r, \mathbb{P}_{Y_r}^{n,*})$ for $t \leq r \leq s_n$. Using the fact that the trajectories are càdlàg, by applying Itô's formula on (IV.3.7) we obtain $-(\mathbb{L}\varphi + F)(t, m) \leq 0$. \blacksquare

IV.3.4 Stability

Theorem IV.3.10 *Let $\{u_\varepsilon\}_{\varepsilon>0}$ and $\{v_\varepsilon\}_{\varepsilon>0}$ be two families of viscosity subsolutions and supersolutions of (IV.2.10), respectively. Assume that the following relaxed semilimits are finite*

$$\bar{u}(t, m) := \limsup_{(\varepsilon, s, \tilde{m}) \rightarrow (0, t, m)} u_\varepsilon(s, \tilde{m}), \quad \text{and} \quad \underline{v}(t, m) := \liminf_{(\varepsilon, s, \tilde{m}) \rightarrow (0, t, m)} v_\varepsilon(s, \tilde{m}), \quad (t, m) \in \overline{\mathbf{Q}}_0,$$

where the limits are sequences $(\varepsilon_n, t_n, m_n) \rightarrow (0, t, m)$, with $(t_n, m_n) \in \mathcal{N}_{T-t}(t, m)$. Then \bar{u} (resp. \underline{v}) is a USC (resp. LSC) viscosity subsolution (resp. supersolution) of (IV.2.10).

Proof (i) We prove the stability of the supersolution first. Observe that we may assume without loss of generality that v_ε is LSC as $\underline{v}(t, m) = \liminf_{(\varepsilon, s, \tilde{m}) \rightarrow (0, t, m)} (v_\varepsilon)_*(s, \tilde{m})$.

Fix $(t, m) \in \mathbf{Q}_0$, and $\varphi \in \overline{\mathcal{A}}\underline{v}(t, m)$ with corresponding δ , and s.t. (t, m) is a strict maximizer of $\varphi - \underline{v}$ on $\mathcal{N}_\delta(t, m)$, see Remark IV.3.2. By definition, there exists a sequence $(\varepsilon_n, t_n, m_n) \rightarrow (0, t, m)$ s.t. $v_{\varepsilon_n}(t_n, m_n) \rightarrow \underline{v}(t, m)$. Note that $(t_n, m_n) \in \mathcal{N}_\delta(t, m)$ for all n large, then we can find $\delta' < \delta$ s.t. $\mathcal{N}_{\delta'}(t_n, m_n) \subset \mathcal{N}_\delta(t, m)$. Let (\hat{t}_n, \hat{m}_n) be a maximizer of $\varphi - v_{\varepsilon_n}$ on $\mathcal{N}_{\delta'}(t_n, m_n)$. We first note that

$$(\hat{t}_n, \hat{m}_n) \xrightarrow{n \rightarrow \infty} (t, m). \quad (\text{IV.3.8})$$

Indeed, $(\hat{t}_n, \hat{m}_n) \in \mathcal{N}_{\delta'}(t_n, m_n) \subset \mathcal{N}_\delta(t, m)$ for all n . Thus, by compactness, there exists a subsequence (still named \hat{m}_n) converging to some $(\hat{t}, \hat{m}) \in \mathcal{N}_\delta(t, m)$. Observing that

$$\begin{aligned} [\varphi - \underline{v}](t, m) &= \lim_{n \rightarrow \infty} [\varphi - v_{\varepsilon_n}](t_n, m_n) \leq \liminf_{n \rightarrow \infty} [\varphi - v_{\varepsilon_n}](\hat{t}_n, \hat{m}_n) \\ &\leq \limsup_{n \rightarrow \infty} [\varphi - v_{\varepsilon_n}](\hat{t}_n, \hat{m}_n) \leq [\varphi - \underline{v}](\hat{t}, \hat{m}), \end{aligned}$$

we conclude from the fact that (t, m) is a strict maximizer of $\varphi - \underline{v}$ on $\mathcal{N}_\delta(t, m)$ that $(\hat{t}, \hat{m}) = (t, m)$, and thus (IV.3.8) holds true. Then, given that (t_n, m_n) and (\hat{t}_n, \hat{m}_n) have the same limit, we have $\mathcal{N}_{\delta''}(\hat{t}_n, \hat{m}_n) \subset \mathcal{N}_{\delta'}(t_n, m_n)$ for some $\delta'' < \delta'$ and n large enough. Then, as (\hat{t}_n, \hat{m}_n) is also a maximizer on $\mathcal{N}_{\delta''}(\hat{t}_n, \hat{m}_n)$, the supersolution property implies $-(\mathbb{L}\varphi + F)(\hat{t}_n, \hat{m}_n) \geq 0$, for n large enough, and we derive the first part of the supersolution property of \underline{v} by sending $n \rightarrow \infty$.

We now prove that \underline{v} is increasing for \preceq . By Lemma IV.3.8 (i), it suffices to prove that $\underline{v}(t, m) \geq \underline{v}(t, m')$ for a given $m' \preceq m$ with continuous conditional transition probability p . We define for all n the measure $m'_n \preceq m_n$, obtained from m_n by applying p . As $\mathcal{W}_2(m_n, m) \xrightarrow{n \rightarrow \infty} 0$ and p is continuous, similarly to the proof of Theorem IV.3.9, we see that $\mathcal{W}_2(m'_n, m') \xrightarrow{n \rightarrow \infty} 0$. Moreover, by the supersolution property of v_{ε_n} , we have $v_{\varepsilon_n}(t_n, m_n) \geq v_{\varepsilon_n}(t_n, m'_n)$, for all $n \geq 1$, and we conclude by taking the \liminf that $\underline{v}(t, m) \geq \underline{v}(t, m')$, as the l.h.s. of the inequality converges.

(ii) We now prove the stability of the subsolution. Similarly to (i), we may assume that $\{u_\varepsilon\}_{\varepsilon > 0}$ is a family of USC viscosity subsolutions of (IV.2.10). Let (t, m) and $\varphi \in \underline{\mathcal{A}}\bar{u}(t, m)$

be such that (t, m) is a strict local minimizer of $\varphi - \bar{u}$. Assume that $C_{\bar{u}}(t, m) = \{m\}$ and $(\mathbb{D}_I \varphi)_*(t, m) > 0$. Following the same argument as in the previous step, replacing maximizers with minimizers, we may construct (\hat{t}_n, \hat{m}_n) , converging to some (\hat{t}, \hat{m}) , and satisfying the inequalities

$$[\varphi - \bar{u}](t, m) \geq \limsup_{n \rightarrow \infty} [\varphi - u_{\varepsilon_n}](\hat{t}_n, \hat{m}_n) \geq \liminf_{n \rightarrow \infty} [\varphi - u_{\varepsilon_n}](\hat{t}_n, \hat{m}_n) \geq [\varphi - \bar{u}](\hat{t}, \hat{m}).$$

By the strict minimum property of (t, m) , this again implies that $(\hat{t}, \hat{m}) = (t, m)$, and $\lim_{n \rightarrow \infty} u_{\varepsilon_n}(\hat{t}_n, \hat{m}_n) = \bar{u}(t, m)$. By Lemma IV.2.3, we may now take $m_n^* \in \operatorname{argmin}_{\leq} C_{u_{\varepsilon_n}}(\hat{t}_n, \hat{m}_n)$. By compactness, there is a subsequence $\{m_n^*\}_{n \geq 1}$ converging to some m^* . As $u_{\varepsilon_n}(\hat{t}_n, \hat{m}_n) \leq u_{\varepsilon_n}(\hat{t}_n, m_n^*)$ for all n , by definition of $C_{u_{\varepsilon_n}}(\hat{t}_n, \hat{m}_n)$, taking the lim sup implies $\bar{u}(t, m) \leq \bar{u}(t, m^*)$, hence $m^* = m$ as $C_{\bar{u}}(t, m) = \{m\}$. As $(\mathbb{D}_I \varphi)_*(t, m) > 0$, \hat{m}_n and m_n^* are both in a neighborhood where φ is strictly increasing for n large enough, and thus $[\varphi - u_{\varepsilon_n}](\hat{t}_n, \hat{m}_n) \geq [\varphi - u_{\varepsilon_n}](\hat{t}_n, m_n^*)$, which implies equality by definition of (\hat{t}_n, \hat{m}_n) and the fact that $(\hat{t}_n, m_n^*) \in \mathcal{N}_{\delta'}(\hat{t}_n, \hat{m}_n)$. Then $\varphi \in \underline{\mathcal{A}}_{u_{\varepsilon_n}}(\hat{t}_n, m_n^*)$. As $C_{u_{\varepsilon_n}}(\hat{t}_n, m_n^*) = \{m_n^*\}$, the viscosity subsolution implies $-(\mathbb{L}\varphi + F)(\hat{t}_n, m_n^*) \leq 0$ for n large enough, and we conclude by letting $n \rightarrow \infty$. ■

IV.3.5 Comparison

Theorem IV.3.11 (i) *Let u be an USC viscosity subsolution of (IV.2.10) satisfying $u|_{t=T} \leq g$. Assume further that f is uniformly continuous in (t, x, m) under \mathcal{W}_2 . Then $u \leq V$.*

(ii) *Let v be a LSC viscosity supersolution of (IV.2.10) satisfying $v|_{t=T} \geq g$. Assume further that b, σ, f, g can be extended to $\mathcal{P}_1(\mathbf{S})$ under \mathcal{W}_1 continuously; b is uniformly Lipschitz continuous in (x, m) under \mathcal{W}_1 ; and σ has the regularity required in Lemma IV.3.6. Then $v \geq V$.*

Proof (i) We first compare V and u . Assume by contradiction that $u(t, m) > V(t, m)$ for some (t, m) . Then, for $\varepsilon > 0$ small enough,

$$u(t, m) - \varphi_\varepsilon(t, m) > \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \int_t^T F(r, \mathbb{P}_{Y_r}) dr + g(\mathbb{P}_{Y_T}), \quad (\text{IV.3.9})$$

where $\varphi_\varepsilon(s, \tilde{m}) := \varepsilon[(T - t) + m(\mathbb{R}^d, 1)]$. Let (t^*, \mathbb{P}^*) be s.t.

$$\begin{aligned} & (u - \varphi_\varepsilon)(t^*, m^*) + \int_t^{t^*} F(r, \mathbb{P}_{Y_r}^*) dr \\ &= \max_{\substack{(s, m, \mathbb{P}) \in \mathcal{N}_{T-t}(t, m) \times \mathcal{P}(t, m) : \\ m \in \{\mathbb{P}_{Y_{s-}}, \mathbb{P}_{Y_s}\}}} (u - \varphi_\varepsilon)(s, m) + \int_t^s F(r, \mathbb{P}_{Y_r}) dr. \end{aligned} \quad (\text{IV.3.10})$$

where m^* is the optimal argument in $\{\mathbb{P}_{Y_{t^*-}}^*, \mathbb{P}_{Y_{t^*}}^*\}$. Clearly $t^* < T$. Indeed, if $t^* = T$, then $(T, m^*) \in \mathcal{N}_{T-t}(t, m)$, and by (IV.3.10) and (IV.3.9) we have

$$\begin{aligned} u(T, m^*) - \varepsilon m^*(\mathbb{R}^d, 1) + \int_t^T F(r, \mathbb{P}_{Y_r}^*) dr &\geq (u - \varphi_\varepsilon)(t, m) \\ &> \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \int_t^T F(r, \mathbb{P}_{Y_r}) dr + g(\mathbb{P}_{Y_T}) \geq \int_t^T F(r, \mathbb{P}_{Y_r}^*) dr + u(T, m^*), \end{aligned}$$

as $u(T, \cdot) \leq g$. This is the desired contradiction. Moreover, by Lemma IV.2.3, we may choose m^* to be the smallest one which keeps the same value $(u - \varphi_\varepsilon)(t^*, m^*)$. Note that this change is only at t^* , and thus has no impact on the value of $\int_t^{t^*} F(r, \mathbb{P}_{Y_r}^*) dr$. Then

$$(u - \varphi_\varepsilon)(t^*, m^*) > (u - \varphi_\varepsilon)(t^*, m'), \quad \text{for all } m^* \neq m' \preceq m^*. \quad (\text{IV.3.11})$$

Furthermore, we note that, since $m \mapsto m(\mathbb{R}^d, 1)$ is increasing, by (IV.3.11) actually we have

$$u(t^*, m^*) > u(t^*, m'), \quad \text{for all } m^* \neq m' \preceq m^*, \quad \text{namely } C_u(t^*, m^*) = \{m^*\}. \quad (\text{IV.3.12})$$

Next, let f^+, f^- denote the positive and negative part of f , respectively, and ρ_0 the modulus of continuity function of f . Introduce:

$$\begin{aligned} \underline{f}^+(s, x, \tilde{m}) &:= f^+(s, x, \tilde{m}) - \rho_0(|\tilde{m}(\mathbb{R}^d, 1) - m^*(\mathbb{R}^d, 1)|^{\frac{1}{2}}); \\ \bar{f}^-(s, x, \tilde{m}) &:= f^-(s, x, \tilde{m}) + \rho_0(|\tilde{m}(\mathbb{R}^d, 1) - m^*(\mathbb{R}^d, 1)|^{\frac{1}{2}}). \end{aligned}$$

It is clear that $\underline{f}^+, \bar{f}^-$ are also uniformly continuous in (s, x, \tilde{m}) (under \mathcal{W}_2). For $\varepsilon > 0$, by Lemma IV.3.7 (i) let $\underline{f}_n^+, \bar{f}_n^-$ be a smooth mollifier (under \mathcal{W}_2) such that

$$|\underline{f}_n^+ - \underline{f}^+| \leq \frac{\varepsilon}{6}, \quad |\bar{f}_n^- - \bar{f}^-| \leq \frac{\varepsilon}{6}, \quad \text{on } \mathcal{P}(t^*, m^*).$$

Then, for all $(s, \tilde{m}) \in \mathcal{N}_{T-t^*}(t^*, m^*)$ with corresponding $\mathbb{P} \in \mathcal{P}(t^*, m^*)$, and $t^* \leq r < s$, considering the case $\tilde{m} = \mathbb{P}_{Y_{s-}}$, we have

$$\begin{aligned} \mathbb{E}^\mathbb{P} \left[\left| f^+(s, X_s, \mathbb{P}_{(X_s, I_r)}) I_{s-} - f^+(s, X_s, \tilde{m}) I_{s-} \right| \right] &\leq \rho_0 \left(\mathcal{W}_2(\mathbb{P}_{(X_s, I_r)}, \mathbb{P}_{(X_s, I_{s-})}) \right) \\ &\leq \rho_0 \left(\sqrt{\mathbb{E}^\mathbb{P}[|I_r - I_{s-}|^2]} \right) \leq \rho_0 \left(\sqrt{\mathbb{E}^\mathbb{P}[|I_{t^*} - I_{s-}|^2]} \right) = \rho_0(|\tilde{m}(\mathbb{R}^d, 1) - m^*(\mathbb{R}^d, 1)|^{\frac{1}{2}}). \end{aligned}$$

Then

$$\mathbb{E}^\mathbb{P} \left[f^+(s, X_s, \mathbb{P}_{(X_s, I_r)}) I_{s-} \right] \geq \mathbb{E}^\mathbb{P} \left[\underline{f}_n^+(s, X_s, \tilde{m}) I_{s-} \right]$$

and, similarly,

$$\mathbb{E}^{\mathbb{P}} \left[f^-(t^*, X_{t^*}, \mathbb{P}_{(X_s, I_r)}) I_{t^*} \right] \leq \mathbb{E}^{\mathbb{P}} \left[\bar{f}^-(t^*, X_{t^*}, \tilde{m}) I_{t^*} \right].$$

Thus, by (IV.2.1) and the regularity of f , we have

$$\begin{aligned} F(r, \mathbb{P}_{Y_r}) &= \mathbb{E}^{\mathbb{P}} \left[f(r, X_r, \mathbb{P}_{Y_r}) I_r \right] = \mathbb{E}^{\mathbb{P}} \left[f^+(r, X_r, \mathbb{P}_{Y_r}) I_r - f^-(r, X_r, \mathbb{P}_{Y_r}) I_r \right] \\ &\geq \mathbb{E}^{\mathbb{P}} \left[f^+(r, X_r, \mathbb{P}_{Y_r}) I_{s-} - f^-(r, X_r, \mathbb{P}_{Y_r}) I_{t^*} \right] \\ &\geq \mathbb{E}^{\mathbb{P}} \left[f^+(s, X_s, \mathbb{P}_{(X_s, I_r)}) I_{s-} - f^-(t^*, X_{t^*}, \mathbb{P}_{(X_s, I_r)}) I_{t^*} \right] - \rho(s - t^*) \\ &\geq \mathbb{E}^{\mathbb{P}} \left[\underline{f}_n^+(s, X_s, \tilde{m}) I_{s-} - \bar{f}_n^-(t^*, X_{t^*}, \tilde{m}) I_{t^*} \right] - \rho(s - t^*) \\ &\geq \mathbb{E}^{\mathbb{P}} \left[\underline{f}_n^+(s, X_s, \tilde{m}) I_{s-} - \bar{f}_n^-(t^*, X_{t^*}, \tilde{m}) I_{t^*} \right] - \frac{\varepsilon}{3} - \rho(s - t^*), \end{aligned}$$

for some modulus of continuity ρ which can be chosen to be smooth on $(0, \infty)$. That is,

$$F(r, \mathbb{P}_{Y_r}) \geq \int \underline{f}_n^+(s, x, \tilde{m}) i \tilde{m}(dx, di) - \int \bar{f}_n^-(t^*, x, \tilde{m}) i m^*(dx, di) - \frac{\varepsilon}{3} - \rho(s - t^*).$$

In the case $\tilde{m} = \mathbb{P}_{Y_s}$, following similar arguments we see the above still holds true. Denote

$$\begin{aligned} \phi_\varepsilon^n(s, \tilde{m}) &:= \int f_n^+(s, x, \tilde{m}) i \tilde{m}(dx, di) - \int f_n^-(t^*, x, \tilde{m}) i m^*(dx, di), \\ \psi_\varepsilon^n(s, \tilde{m}) &:= \varphi_\varepsilon(s, \tilde{m}) - (s - t^*) \left[\phi_\varepsilon^n(s, \tilde{m}) - \frac{\varepsilon}{3} - \rho(s - t^*) \right], \end{aligned}$$

which are obviously in $C_2^{1,1}(\bar{\mathbf{Q}}_0)$. Then, by (IV.3.10),

$$\begin{aligned} (u - \psi_\varepsilon^n)(t^*, m^*) &= (u - \varphi_\varepsilon)(t^*, m^*) \geq (u - \varphi_\varepsilon)(s, \tilde{m}) + \int_{t^*}^s F(r, \mathbb{P}_{Y_r}) dr \\ &\geq (u - \varphi_\varepsilon)(s, \tilde{m}) + (s - t^*) \left[\phi_\varepsilon^n(s, \tilde{m}) - \frac{\varepsilon}{3} - \rho(s - t^*) \right] = (u - \psi_\varepsilon^n)(s, \tilde{m}). \end{aligned}$$

Thus $\psi_\varepsilon^n \in \underline{\mathcal{A}}u(t^*, m^*)$. Note that

$$\begin{aligned} [\mathbb{L}\psi_\varepsilon^n + F](t^*, m^*) &= \left[\mathbb{L}\varphi_\varepsilon - \phi_\varepsilon^n + \frac{\varepsilon}{3} + F \right](t^*, m^*) = -\varepsilon - \phi_\varepsilon^n(t^*, m^*) + \frac{\varepsilon}{3} + F(t^*, m^*) \\ &= -\frac{2\varepsilon}{3} + \int \left[\underline{f}_n^+(t^*, x, m^*) - \bar{f}_n^-(t^*, x, m^*) \right] i m^*(dx, di) + F(t^*, m^*) \\ &\leq -\frac{\varepsilon}{3} + \int \left[\underline{f}^+(t^*, x, m^*) - \bar{f}^-(t^*, x, m^*) \right] i m^*(dx, di) + F(t^*, m^*) \\ &\leq -\frac{\varepsilon}{3} + \int \left[f^+(t^*, x, m^*) - f^-(t^*, x, m^*) \right] i m^*(dx, di) + F(t^*, m^*) = -\frac{\varepsilon}{3} < 0; \\ (D_I \psi_\varepsilon^n)(s, \tilde{m}) &= \varepsilon - (s - t^*) (D_I \phi_\varepsilon^n)(s, \tilde{m}), \quad \text{and thus} \quad (D_I \psi_\varepsilon^n)_*(t^*, m^*) = \varepsilon > 0. \end{aligned}$$

Recall (IV.3.12), this contradicts the viscosity subsolution property of u .

(ii) We next compare V and v . Fix $\varepsilon > 0$. For each $n \geq 1$, denote $t_i := t_i^{(n)} := \frac{iT}{n}$, $0 \leq i \leq n$. First, note that, for $(t, m) \in \mathbf{Q}_0$, it follows from the continuity of the coefficients that

$$V(t, m) := \lim_{n \rightarrow \infty} V_n(t, m), \quad \text{where} \quad V_n(t, m) := \sup_{\mathbb{P} \in \mathcal{P}_n(t, m)} \int_t^T F(r, \mathbb{P}_{Y_r}) dr + g(\mathbb{P}_{Y_T}),$$

$$\mathcal{P}_n(t, m) := \left\{ \mathbb{P} \in \mathcal{P}(t, m) : \tau \text{ takes values in } \{t_1, \dots, t_n\} \cap [t, T], \mathbb{P}\text{-a.s.} \right\}.$$

Step 1: We show that $(V_n - v)(t_{n-1}, \cdot) \leq \frac{\varepsilon}{n}$. Assume to the contrary that there exists m_{n-1} such that $(V_n - v)(t_{n-1}, m_{n-1}) > \frac{\varepsilon}{n}$. By the definition of $\mathcal{P}_n(t_{n-1}, m_{n-1})$, we have $V_n(t, m) = \int_t^T F(r, \bar{\mathbb{P}}_{Y_r}^{t, m}) dr + g(\bar{\mathbb{P}}_{Y_T}^{t, m})$, $t \in (t_{n-1}, T]$, where $\bar{\mathbb{P}}^{t, m} \in \mathcal{P}(t, m)$ is defined by (IV.2.4).

Let $\delta_1, \delta_2 > 0$ be small numbers which will be specified later. Applying Lemma IV.3.7 (ii), (iii), let (g_k, f_k, b_k) be the smooth mollifier of (g, f, b) (under \mathcal{W}_1), where b_k is also mollified in (t, x) in a standard way, such that $\|g_k - g\|_\infty + \|f_k - f\|_\infty \leq \delta_1$, $\|b_k - b\|_\infty \leq \delta_2$, and g_k is Lipschitz continuous under \mathcal{W}_1 with a Lipschitz constant L_k depending on k , and b_k is uniformly Lipschitz continuous in (x, m) under \mathcal{W}_1 with a Lipschitz constant L independent of k . By otherwise choosing a larger L we assume σ is also uniformly Lipschitz continuous in (x, m) under \mathcal{W}_1 with Lipschitz constant L . Let U^{k_1, k_2} be defined by (IV.3.4) corresponding to $(b_{k_2}, \sigma, g_{k_1}, f_{k_1})$. Then, by Lemma IV.3.6,

$$\partial_t U^{k_1, k_2}(t, m) + \int_{\mathbb{R}^d} \left[b_{k_2} \cdot \partial_x \delta_m U_1^{k_1, k_2} + \frac{1}{2} \sigma^2 : \partial_{xx}^2 \delta_m U_1^{k_1, k_2} + f_{k_1} \right] m(dx, 1) = 0, \quad (\text{IV.3.13})$$

and U^{k_1, k_2} is Lipschitz continuous in m under \mathcal{W}_1 with a Lipschitz constant $C_{L, L_{k_1}}$ independent of k_2 . Here, $\delta_m U_1^{k_1, k_2}$ is in the sense of (IV.2.6). This, in particular, implies $|\partial_x \delta_m U_1^{k_1, k_2}| \leq C_{L, L_{k_1}}$ for all k_2 . Then, we deduce from (IV.3.13) that

$$\left| (\mathbb{L} U^{k_1, k_2} + F_{k_1})(t, m) \right| = \left| \int (b - b_{k_2}) \cdot \partial_x \delta_m U_1^{k_1, k_2} m(dx, 1) \right| \leq C_{L, L_{k_1}} \delta_2, \quad \forall k_2 \geq \mathbb{I}, \quad (\text{IV.3.14})$$

where $F_{k_1}(t, m) := \int_{\mathbb{R}^d} f_{k_1}(t, x, m) m(dx, 1)$ as in (IV.2.1). Moreover, since

$$U^{k_1, k_2}(t, m) = g_{k_1}(\bar{\mathbb{P}}_{Y_T}^{t, m, k_2}) + \int_t^T F_{k_1}(r, \bar{\mathbb{P}}_{Y_r}^{t, m, k_2}) dr,$$

where $\bar{\mathbb{P}}^{t, m, k_2}$ is s.t. X is unstopped with drift coefficient b_{k_2} instead of b , one can easily show that

$$\left| U^{k_1, k_2}(t, m) - (g(\bar{\mathbb{P}}_{Y_T}^{t, m}) + \int_t^T F(r, \bar{\mathbb{P}}_{Y_r}^{t, m}) dr) \right| \leq C[\delta_1 + \delta_2] \leq \frac{\varepsilon}{4n},$$

for δ_1, δ_2 small enough. Then

$$\begin{aligned} V_n(t_{n-1}, m_{n-1}) &= \sup_{m' \preceq m_{n-1}} \int_{t_{n-1}}^T F(r, \bar{\mathbb{P}}_{Y_r}^{t_{n-1}, m'}) dr + g(\bar{\mathbb{P}}_{Y_T}^{t_{n-1}, m'}) \\ &\leq \sup_{m' \preceq m_{n-1}} U^{k_1, k_2}(t_{n-1}, m') + \frac{\varepsilon}{4n}. \end{aligned}$$

By the supersolution property, v is nondecreasing for \preceq , hence

$$\frac{\varepsilon}{n} \leq (V_n - v)(t_{n-1}, m_{n-1}) \leq \sup_{m' \preceq m_{n-1}} (U_{k_1, k_2} - v)(t_{n-1}, m') + \frac{\varepsilon}{4n}.$$

This implies that,

$$\max_{(s, \tilde{m}) \in \mathcal{N}_{\frac{T}{n}}(t_{n-1}, m_{n-1})} \left[(U^{k_1, k_2} - v)(s, \tilde{m}) - \frac{T-s}{n} \right] \geq \frac{3\varepsilon}{4n} - \frac{T}{n^2} \geq \frac{\varepsilon}{2n}, \quad (\text{IV.3.15})$$

for n sufficiently large. Note that $(U^{k_1, k_2} - v)(T, \mathbb{P}_{Y_T}) \leq (g_{k_1} - g)(\mathbb{P}_{Y_T}) \leq \frac{\varepsilon}{4n}$ for all $\mathbb{P} \in \mathcal{P}(t_{n-1}, m_{n-1})$ and v is LSC, then by compactness of $\mathcal{N}_{\frac{T}{n}}(t_{n-1}, m_{n-1})$ there exists an optimal argument (t^*, m^*) , $t^* < T$, to the above maximum. Thus $\varphi(s, \tilde{m}) := U^{k_1, k_2}(s, \tilde{m}) - \frac{T-s}{n} \in \bar{\mathcal{A}}v(t^*, m^*)$, and therefore,

$$\begin{aligned} 0 &\leq -(\mathbb{L}\varphi + F)(t^*, m^*) = -(\mathbb{L}U^{k_1, k_2} + F_{k_1})(t^*, m^*) + (F_{k_1} - F)(t^*, m^*) - \frac{1}{n} \\ &\leq C_{L, L_{k_1}} \delta_2 + (F_{k_1} - F)(t^*, m^*) - \frac{1}{n}, \end{aligned}$$

where the last inequality thanks to (IV.3.14). Fix k_1 so that $(F_{k_1} - F)(t^*, m^*) \leq \frac{1}{2n}$ and set δ_2 small enough, we obtain the desired contradiction.

Step 2: We show that $(V_n - v)(t_{n-2}, \cdot) \leq \frac{2\varepsilon}{n}$. Assume to the contrary that there exists m_{n-2} such that $(V_n - v)(t_{n-2}, m_{n-2}) > \frac{2\varepsilon}{n}$. By the DPP, we have

$$V_n(t_{n-2}, m_{n-2}) = \sup_{\mathbb{P} \in \mathcal{P}_n(t_{n-2}, m_{n-2})} \int_{t_{n-2}}^{t_{n-1}} F(r, \mathbb{P}_{Y_r}) dr + V_n(t_{n-1}, \mathbb{P}_{Y_{(t_{n-1})^-}}),$$

Observe the fact that v is a viscosity supersolution of (IV.2.10) also implies that $v + \frac{\varepsilon}{n}$ is a viscosity supersolution. Moreover, by *Step 1*, we have $(v + \frac{\varepsilon}{n})(t_{n-1}, \cdot) \geq V_n(t_{n-1}, \cdot)$. Thus, using the same procedure as *Step 1* (where V_n replaces g on $(t_{n-2}, t_{n-1}]$), it follows that

$$(V_n - (v + \frac{\varepsilon}{n}))(t_{n-2}, \cdot) \leq \frac{\varepsilon}{n}.$$

Finally, by backwards induction, we have $(V_n - v)(t_{n-j}, \cdot) \leq \frac{j\varepsilon}{n}$ for all $j \in \{0, \dots, n\}$, and thus $(V_n - v)(t, \cdot) \leq \varepsilon$, which implies by the arbitrariness of n and ε that $v \geq V$. \blacksquare

IV.3.6 Infinite horizon case

As in [74, §6.1], we may formulate the problem in infinite horizon (i.e., in the case $T = \infty$), by replacing Assumption IV.2.1 with the following conditions:

Assumption IV.3.12 (i) *Assumption IV.2.1 holds true on $[0, \infty)$;*

(ii) $\int_0^\infty \sup_{m \in \mathcal{P}_2(\mathbf{S})} |F(t, m)| dt < \infty$;

(iii) *For any (t, m) and $\mathbb{P} \in \mathcal{P}(t, m)$, $X_\infty := \lim_{t \rightarrow \infty} X_t$ exists, \mathbb{P} -a.s.*

We remark that one sufficient condition of (ii) above is that $|f(t, x, m)| \leq Ce^{-\lambda t}$ for some constants $C, \lambda > 0$, and a special case of (iii) is:

$$d = 1, \quad b = b_0 x, \quad \sigma = \sigma_0 x, \quad b_0 - \frac{1}{2}\sigma_0^2 < 0. \quad (\text{IV.3.16})$$

That is, under the $\bar{\mathbb{P}}^{t, m}$ in (IV.2.4), the unstopped process X is a Geometric Brownian motion and $X_\infty = 0$. All the previous results still hold true in this case, after some obvious modifications.

IV.4 Examples

In this section we revisit the three examples studied in [74], and add a new example concerning probability distortion. Note that in [74] we assumed that the value functions are smooth, which is hard to verify. In this section we show that they are the unique continuous viscosity solution of the corresponding obstacle problem. Note that we shall allow both $T < \infty$ and $T = \infty$, and correspondingly we always assume Assumption IV.2.1 or IV.3.12, and we shall report the detailed arguments in the case $T < \infty$ only. Moreover, for simplicity in this section we always assume $f = 0$.

IV.4.1 Connection with standard optimal stopping

Assume for this example that b and σ do not depend on the measure variable m . For a measurable function $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$, we define the optimal stopping problem

$$V(t, m) := \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \mathbb{E}^{\mathbb{P}} \left[\psi(X_T) \right], \quad (t, m) \in \bar{\mathbf{Q}}_0. \quad (\text{IV.4.1})$$

That is, $g(\mu) := \int_{\mathbb{R}^d} \psi(x) \mu(dx)$ for $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. We also introduce $v(t, x) := V(t, \delta_{(x, 1)})$, which is related to the standard obstacle problem: recalling (IV.2.7),

$$\min\{-(\partial_t + \mathcal{L})v, v - \psi\} = 0, \quad v(T, \cdot) = \psi, \quad \text{where } \mathcal{L}v := b \cdot \partial_x v + \frac{1}{2}\sigma^2 : \partial_{xx}^2 v. \quad (\text{IV.4.2})$$

Proposition IV.4.1 *Assume b, σ do not depend on m , σ satisfies the regularities required in Lemma IV.3.6, and ψ is uniformly continuous. Then V is the unique continuous viscosity solution of the corresponding obstacle equation (IV.2.10), and it holds*

$$V(t, m) = \int_{\mathbf{S}} [v(t, x)i + \psi(x)(1 - i)] m(dx, di). \quad (\text{IV.4.3})$$

Moreover, there exists a pure strategy optimal stopping time.

Proof First, by the uniform continuity of ψ one can easily show that g is uniformly continuous in m under \mathcal{W}_1 . Then by Theorem IV.3.5 V is continuous in t and uniformly continuous in m under \mathcal{W}_1 . Thus it follows from Theorems IV.3.9 and IV.3.11 that V is the unique viscosity solution of (IV.2.10).

It remains to verify (IV.4.3). Let $\mathbb{P}^* \in \mathcal{P}(t, m)$ be such that,

$$\tau = \inf\{s \geq t : v(s, X_s) = \psi(X_s)\}, \quad \mathbb{P}^* - \text{a.s. on } \{I_{t-} = 1\}. \quad (\text{IV.4.4})$$

By the standard optimal stopping problem, see e.g. Karatzas & Shreve [49, Appendix D], v is continuous and \mathbb{P}^* is optimal. Then by (IV.4.1) we derive (IV.4.3):

$$\begin{aligned} V(t, m) &= \int_{\mathbb{R}^d} \psi(x) m(dx, 0) + \int_{\mathbb{R}^d} \mathbb{E}^{\mathbb{P}^*} [\psi(X_T) | X_t = x] m(dx, 1) \\ &= \int_{\mathbb{R}^d} \psi(x) m(dx, 0) + \int_{\mathbb{R}^d} v(t, x) m(dx, 1). \end{aligned}$$

Moreover, clearly the optimal stopping time determined by (IV.4.4) is a pure strategy. \blacksquare

We remark that, by utilizing (IV.4.3), it is possible to prove the uniqueness of viscosity solution under weaker requirement on σ . We leave the details to the interested readers.

IV.4.2 A generalization of the mean variance problem

Consider the optimal stopping problem:

$$V(t, m) := \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \varphi \left(\mathbb{E}^{\mathbb{P}} [\psi(X_T)] \right), \quad (\text{IV.4.5})$$

where $\psi : \mathbb{R}^d \rightarrow \mathbb{R}^k$ for some $k \geq 1$ and $\varphi : \mathbb{R}^k \rightarrow \mathbb{R}$. That is, $g(\mu) = \varphi \left(\int_{\mathbb{R}^d} \psi(x) \mu(dx) \right)$.

Proposition IV.4.2 *Let b, σ satisfy the conditions in Theorem IV.3.11 (ii), ψ be uniformly continuous and φ be continuous. Assume further that either ψ is bounded or φ is uniformly continuous. Then V is the unique continuous viscosity solution of the corresponding obstacle equation (IV.2.10).*

Proof Note that, when $|\psi| \leq C$, we have $|\mathbb{E}^{\mathbb{P}}[\psi(X_T)]| \leq C$ and thus in (IV.4.5) we may replace φ with the truncated function $\varphi_C(z) := \varphi(\frac{C}{|z|\sqrt{C}}z)$, $z \in \mathbb{R}^k$, which is uniformly continuous. Then in both cases, we may assume w.l.o.g. that φ is uniformly continuous, and therefore, g is uniformly continuous in μ under \mathcal{W}_1 . Then the results follows from Theorems IV.3.5, IV.3.9, and IV.3.11. \blacksquare

Remark IV.4.3 (i) In the case that φ is convex: $\varphi(z) := \sup_{\alpha} [\alpha z - \varphi^*(\alpha)]$, we have

$$V(t, m) = \sup_{\alpha} [V_{\alpha}(t, m) - \varphi^*(\alpha)], \quad \text{where} \quad V_{\alpha}(t, m) := \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \mathbb{E}^{\mathbb{P}}[\alpha \cdot \psi(X_T)].$$

Let $\alpha^*(t, m)$ be the optimal argument, then the optimal \mathbb{P}^* for $V_{\alpha^*(t, m)}(t, m)$ is also optimal for $V(t, m)$, and thus by Proposition IV.4.1 there exists a pure optimal strategy for $V(t, m)$.

Moreover, let \mathbb{P}^* be the optimal control for $V(0, m)$ and $V_{\alpha^*(0, m)}(0, m)$ as above, and denote $m_t^* := \mathbb{P}_{Y_t}^*$. Then, by the DPP for V and for $V_{\alpha^*(0, m)}$, we have

$$V(t, m_{t-}^*) = V(0, m) = V_{\alpha^*(0, m)}(0, m) - \varphi^*(\alpha^*(0, m)) = V_{\alpha^*(0, m)}(t, m_{t-}^*) - \varphi^*(\alpha^*(0, m)).$$

That is, $\alpha^*(0, m)$ is optimal for $\sup_{\alpha} [V_{\alpha}(t, m_{t-}^*) - \varphi^*(\alpha)]$, or say, $\alpha^*(t, m_{t-}^*) = \alpha^*(0, m)$ for all t .

(ii) A more special case is the mean variance problem: for some constant $\lambda > 0$,

$$d = 1, \quad k = 2, \quad \psi_1(x) = x, \quad \psi_2(x) = x^2, \quad \varphi(z_1, z_2) = z_1 + \frac{\lambda}{2}z_1^2 - \frac{\lambda}{2}z_2. \quad (\text{IV.4.6})$$

In the homogeneous case (IV.3.16) with $T = \infty$, Pedersen & Peskir [64] solved the problem $V(\delta_{(x,1)})$ and the optimal stopping time is a pure strategy. We are in a much more general framework. However, we should point out that (IV.4.6) does not satisfies the technical conditions in Proposition IV.4.2. \blacksquare

IV.4.3 Expected shortfall

Let $d = 1$, and fix some $\alpha \in (0, 1)$, we consider the mean field optimal stopping problem

$$V(t, m) := \inf_{\mathbb{P} \in \mathcal{P}(t, m)} \text{ES}_{\alpha}^{\mathbb{P}}(X_T) \quad \text{for all } (t, m) \in \overline{\mathbf{Q}}_0,$$

where $\text{ES}_{\alpha}^{\mathbb{P}}$ denotes the expected shortfall under \mathbb{P} , i.e., for any r.v. Z with law μ ,

$$g(\mu) := \text{ES}_{\alpha}^{\mathbb{P}}(Z) := \frac{1}{\alpha} \int_0^{\alpha} q_{\gamma}(Z) d\gamma = \inf_{\beta \in \mathbb{R}} \left\{ \beta + \frac{1}{1-\alpha} \int_{\mathbb{R}} (x - \beta)^+ \mu(dx) \right\}, \quad (\text{IV.4.7})$$

where $q_{\gamma}(Z) := \inf\{z : \mu(Z \leq z) > \gamma\}$.

Here the second equality has been established by Rockafellar & Uryasev [69].

Proposition IV.4.4 *V is the unique continuous viscosity solution of the corresponding equation.*

Proof Clearly, $x \mapsto (x - \beta)^+$ is Lipschitz continuous with Lipschitz constant 1. By (IV.4.7), this implies that g is Lipschitz continuous, and, given our assumptions on the coefficients, we conclude similarly to Proposition IV.4.2 that the required claim follows. ■

Note further that

$$V(t, m) = \inf_{\beta \in \mathbb{R}} \left[\beta + \frac{1}{1-\alpha} V_\beta(t, m) \right], \quad \text{where} \quad V_\beta(t, m) := \inf_{\mathbb{P} \in \mathcal{P}(t, m)} \mathbb{E}^\mathbb{P}[(X_T - \beta)^+].$$

One can easily show that $\lim_{\beta \rightarrow \infty} \left[\beta + \frac{1}{1-\alpha} V_\beta(t, m) \right] = \lim_{\beta \rightarrow -\infty} \left[\beta + \frac{1}{1-\alpha} V_\beta(t, m) \right] = \infty$, where the second equality is due to $\alpha \in (0, 1)$. Then there exists optimal $\beta^* = \beta^*(t, m) \in \mathbb{R}$ such that $V(t, m) = \beta^* + \frac{1}{1-\alpha} V_{\beta^*}(t, m)$. Therefore, similar to Remark IV.4.3 (i), $V(t, m)$ and $V_{\beta^*}(t, m)$ share an optimal $\mathbb{P}^* \in \mathcal{P}(t, m)$, which is a pure optimal strategy as in Proposition IV.4.1.

Moreover, in the homogeneous case with (IV.3.16) and $T = \infty$, one can easily show that V and V_β are independent of t , and $V_\beta(m) = \beta^- m(\mathbb{R}_+, 1) + \int_0^\infty (x - \beta)^+ m(dx, 0)$ whenever $m(\mathbb{R}_+, \{0, 1\}) = 1$.

IV.4.4 Probability distortion

Consider the following optimal stopping problem under probability distortion:

$$V(t, m) := \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \int_0^\infty \varphi(\mathbb{P}(\psi(X_T) \geq z)) dz, \quad (\text{IV.4.8})$$

where $\psi : \mathbb{R}^d \rightarrow [0, \infty)$ is a utility function, and $\varphi : [0, 1] \rightarrow [0, 1]$ is a probability distortion function: $\varphi(0) = 0, \varphi(1) = 1$, and φ is strictly increasing. That is, $g(\mu) = \int_0^\infty \varphi(\mu(\{\psi \geq z\})) dz$.

Proposition IV.4.5 *Let b, σ satisfy the conditions in Theorem IV.3.11 (ii), φ be a uniformly Lipschitz continuous probability distortion function, and ψ be uniformly continuous. Then V is the unique continuous viscosity solution of the corresponding obstacle equation (V.4.2).*

Proof As in the previous examples, it suffices to show that g is uniformly continuous in m under \mathcal{W}_1 . Given arbitrary $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$ and, for $i = 1, 2$, let ξ_i be a random variable

on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P}_{\xi_i} = \mu_i$ and $\mathbb{E}^{\mathbb{P}}[|\xi_1 - \xi_2|] = \mathcal{W}_1(\mu_1, \mu_2)$. Then

$$\begin{aligned} |g(\mu_1) - g(\mu_2)| &\leq \int_0^\infty \left| \varphi(\mathbb{P}(\psi(\xi_1) \geq z)) - \varphi(\mathbb{P}(\psi(\xi_2) \geq z)) \right| dz \\ &\leq C \int_0^\infty \mathbb{E}^{\mathbb{P}} \left[\left| \mathbf{1}_{\{\psi(\xi_1) \geq z\}} - \mathbf{1}_{\{\psi(\xi_2) \geq z\}} \right| \right] dz = C \mathbb{E}^{\mathbb{P}} \left[|\psi(\xi_1) - \psi(\xi_2)| \right]. \end{aligned}$$

Since ψ is uniformly continuous, we see that g is uniformly continuous in μ under \mathcal{W}_1 . ■

Remark IV.4.6 (i) In the homogeneous case (IV.3.16) with $T = \infty$, Xu & Zhou [78] solved the optimal stopping problem $V(\delta_{(x,1)})$ for appropriate φ, ψ , and the optimal stopping time is a pure strategy.

(ii) The mean variance and probability distortion problems are typically viewed as time inconsistent, as the DPP does not hold for value function $v(t, x) := V(t, \delta_{(x,1)})$. However, we emphasize that, by viewing m as our variable, V satisfies the DPP and the problem is hence time consistent. ■

IV.5 Technical results

Proof of Lemma IV.2.3. (i) The set $\{m' : m' \preceq m\}$ is in continuous bijection with the compact set $\{\hat{m} \in \mathcal{P}_2(\mathbf{S} \times \{0, 1\}) : \hat{m} \circ (\mathbf{x}, \mathbf{i})^{-1} = m\}$, with $(\mathbf{x}, \mathbf{i}, \mathbf{i}')$ the projection coordinates on $\mathbf{S} \times \{0, 1\}$. This shows the compactness of $\{m' : m' \preceq m\}$.

(ii) As the map $m' \in \mathcal{K}(m) \mapsto m'(\mathbb{R}^d, 1)$ is continuous and $\mathcal{K}(m)$ is compact, there exists $\bar{m} \in \mathcal{K}(m)$ s.t. $\bar{m} \in \operatorname{argmin}_{m' \in \mathcal{K}(m)} m'(\mathbb{R}^d, 1)$. Let $m' \in \mathcal{K}(m)$ be such that $m' \preceq \bar{m}$ with some corresponding transition probability p , see the definition in (IV.2.9). Then, clearly $m'(\mathbb{R}^d, 1) \leq \bar{m}(\mathbb{R}^d, 1)$ and thus equality holds by minimality of $\bar{m}(\mathbb{R}^d, 1)$. As $p \leq 1$, we conclude that $m' = \bar{m}$. ■

Proof of Lemma IV.3.4. For each $(t, m) \in \bar{\mathbf{Q}}_0$ and $\mathbb{P} \in \mathcal{P}(t, m)$, we extend \mathbb{P} to (Ω, \mathcal{F}_T) as follows: denoted as $\hat{\mathbb{P}} \in \hat{\mathcal{P}}(t, m)$, $X_s = X_r$, $I_s = I_{t-}$, $s \in [-1, t)$, $\hat{\mathbb{P}}$ -a.s. We prove the lemma in two steps.

Step 1. For any compact $\mathcal{M} \subset \mathcal{P}_2(\mathbf{S})$, denote $\hat{\mathcal{P}}_{\mathcal{M}} := \bigcup_{(t,m) \in [0,T] \times \mathcal{M}} \hat{\mathcal{P}}(t, m)$. For each (t, m) , $\mathbb{P} \in \mathcal{P}(t, m)$, and $R > 1$, following the proof of [74, Proposition 2.2] we have

$$\begin{aligned} \mathbb{E}^{\hat{\mathbb{P}}} [|X_T^*|^2] &\leq C \int_{\mathbb{R}^d} |x|^2 m(dx, \{0, 1\}); \\ \mathbb{E}^{\hat{\mathbb{P}}} [|X_T^*|^2 \mathbf{1}_{\{|X_T^*| \geq R\}}] &\leq C \int_{\mathbf{S}} [1 + |x|^2] \left[\mathbf{1}_{\{|x| \geq \sqrt{R-1}\}} + \frac{1}{\sqrt{R}} \right] m(dx, \{0, 1\}). \end{aligned}$$

where $X_T^* := \sup_{0 \leq s \leq T} |X_s|$. By the compactness of \mathcal{M} , one can easily see that

$$\sup_{\hat{\mathbb{P}} \in \hat{\mathcal{P}}_{\mathcal{M}}} \mathbb{E}^{\hat{\mathbb{P}}} [|X_T^*|^2] < \infty, \quad \lim_{R \rightarrow \infty} \sup_{\hat{\mathbb{P}} \in \hat{\mathcal{P}}_{\mathcal{M}}} \mathbb{E}^{\hat{\mathbb{P}}} [|X_T^*|^2 \mathbf{1}_{\{|X_T^*| \geq R\}}] = 0.$$

Then the set $\hat{\mathcal{P}}_{\mathcal{M}}$ is compact. That is, for any $(t_n, m_n) \in [0, T] \times \mathcal{M}$ and $\mathbb{P}^n \in \mathcal{P}(t_n, m_n)$, there exists a subsequence, still denoted the same, such that $\hat{\mathbb{P}}^n \rightarrow \hat{\mathbb{P}}^*$ under \mathcal{W}_2 , for some $\mathbb{P}^* \in \mathcal{P}_2(\Omega, \mathcal{F}_T)$.

We may assume w.l.o.g. that $(t_n, m_n) \rightarrow (t^*, m^*)$ under \mathcal{W}_2 for some $(t^*, m^*) \in [0, T] \times \mathcal{M}$. We next show that $\hat{\mathbb{P}}^* \in \hat{\mathcal{P}}(t^*, m^*)$. Indeed, for any $\delta > 0$, we have $t^* - \delta < t_n < t^* + \delta$ for all n large enough. By the required convergence, it is obvious that $X_s = X_{t^* - \delta}, I_s = I_{t^* - \delta}, s \leq t^* - \delta, \hat{\mathbb{P}}^*$ -a.s., and $\hat{\mathbb{P}}_{Y_{t^* - \delta}}^* = m^*$. Thus, by sending $\delta \rightarrow 0$, $X_s = X_{t^*}, I_s = I_{t^*}, s < t^*, \hat{\mathbb{P}}^*$ -a.s. and $\hat{\mathbb{P}}_{Y_{t^* -}}^* = m^*$. Here we used the fact that X has continuous paths. Moreover, following the arguments in [74, Proposition 2.2] again, we see that the processes M and MM^\top in (IV.2.3) are $\hat{\mathbb{P}}^*$ -martingales, on $[t^* + \delta, T]$ for all $\delta > 0$, and hence also on $[t^*, T]$ (again since X is continuous). That is, $\hat{\mathbb{P}}^* \in \hat{\mathcal{P}}(t^*, m^*)$.

Step 2. We now show that V is USC. Fix (t, m) and choose $(t_n, m_n) \rightarrow (t, m)$ such that $\lim_{n \rightarrow \infty} V(t_n, m_n) = \limsup_{(\tilde{t}, \tilde{m}) \rightarrow (t, m)} V(\tilde{t}, \tilde{m})$. For each n , let $\mathbb{P}^n \in \mathcal{P}(t_n, m_n)$ be optimal: $V(t_n, m_n) = \int_{t_n}^T F(r, \mathbb{P}_{Y_r}^n) dr + g(\mathbb{P}_{Y_T}^n)$. Note that $\mathcal{M} := \{m, m_n, n \geq 1\} \subset \mathcal{P}_2(\mathbf{S})$ is compact. By Step 1, we may assume without loss of generality that $\hat{\mathbb{P}}^n \rightarrow \hat{\mathbb{P}} \in \hat{\mathcal{P}}(t, m)$. Then, since F is continuous and g is USC in m , we have

$$\lim_{n \rightarrow \infty} V(t_n, m_n) = \lim_{n \rightarrow \infty} \left[\int_{t_n}^T F(r, \hat{\mathbb{P}}_{Y_r}^n) dr + g(\hat{\mathbb{P}}_{Y_T}^n) \right] \leq \int_t^T F(r, \hat{\mathbb{P}}_{Y_r}) dr + g(\hat{\mathbb{P}}_{Y_T}) \leq V(t, m).$$

This means that V is USC. ■

Proof of Theorem IV.3.5. (i) follows similar but easier arguments than (ii), so we prove (ii) only. Let ρ_0 denote the modulus of continuity of f, g under \mathcal{W}_1 . We proceed in two steps.

Step 1. Fix $t \in [0, T]$ and $m, \tilde{m} \in \mathcal{P}_2(\mathbf{S})$. For any $\mathbb{P} \in \mathcal{P}(t, m)$, by possibly enlarging the space, there exists $(\tilde{X}_t, \tilde{I}_{t-})$ on the space $(\Omega, \mathcal{F}, \mathbb{P})$ such that

$$\mathbb{P}_{(\tilde{X}_t, \tilde{I}_{t-})} = \tilde{m}, \quad \mathbb{E}^{\mathbb{P}} \left[|\tilde{X}_t - X_t| + |\tilde{I}_{t-} - I_{t-}| \right] = \mathcal{W}_1(m, \tilde{m}).$$

Consider the following SDE on the space $(\Omega, \mathcal{F}, \mathbb{P})$: for $\tilde{Y} := (\tilde{X}, \tilde{I})$,

$$\tilde{X}_s = \tilde{X}_t + \int_t^s b(r, \tilde{X}_r, \mathbb{P}_{\tilde{Y}_r}) \tilde{I}_r dr + \int_t^s \sigma(r, \tilde{X}_r, \mathbb{P}_{\tilde{Y}_r}) \tilde{I}_r dW_r^{\mathbb{P}}, \quad \tilde{I}_r := I_r \tilde{I}_{t-}, \quad \mathbb{P}\text{-a.s. (IV.5.1)}$$

Denote $\Delta Y := \tilde{Y} - Y$. Note that $I_r = I_r I_{t-}$, then

$$\sup_{t \leq r \leq T} |\Delta I_r| = I_r |\Delta I_{t-}| \leq |\Delta I_{t-}|, \quad \text{and thus} \quad \mathbb{E}^{\mathbb{P}} \left[\sup_{t \leq r \leq T} |\Delta I_r| \right] \leq \mathcal{W}_1(m, \tilde{m}). \quad (\text{IV.5.2})$$

Moreover, for $\varphi = b, \sigma$, by the desired Lipschitz continuity under \mathcal{W}_1 , we have

$$\left| \varphi(r, \tilde{X}_r, \mathbb{P}_{\tilde{Y}_r}) \tilde{I}_r - \varphi(r, X_r, \mathbb{P}_{Y_r}) I_r \right| \leq C \left[|\Delta X_r| + \mathcal{W}_1(\mathbb{P}_{\tilde{Y}_r}, \mathbb{P}_{Y_r}) \right] + C[1 + |X_r|] |\Delta I_{t-}|.$$

By standard estimates, one can show that

$$\begin{aligned} \mathbb{E}_t^{\mathbb{P}} \left[\sup_{t \leq s \leq T} |X_s|^2 \right] &\leq C[1 + |X_t|^2]; \\ \mathbb{E}_t^{\mathbb{P}} \left[|\Delta X_s|^2 \right] &\leq C \int_s^T \mathcal{W}_1^2(\mathbb{P}_{\tilde{Y}_r}, \mathbb{P}_{Y_r}) dr + C |\Delta X_t|^2 + C \sup_{t \leq s \leq T} \mathbb{E}_t^{\mathbb{P}} [1 + |X_s|^2] |\Delta I_{t-}|^2; \\ \mathbb{E}_t^{\mathbb{P}} \left[|\Delta X_s| \right] &\leq C \left(\int_s^T \mathcal{W}_1^2(\mathbb{P}_{\tilde{Y}_r}, \mathbb{P}_{Y_r}) dr \right)^{\frac{1}{2}} + C |\Delta X_t| + C[1 + |X_t|] |\Delta I_{t-}|. \end{aligned} \quad (\text{IV.5.3})$$

This implies that

$$\begin{aligned} \mathbb{E}^{\mathbb{P}} \left[|\Delta X_s| \right] &\leq C \left(\int_s^T \mathcal{W}_1^2(\mathbb{P}_{\tilde{Y}_r}, \mathbb{P}_{Y_r}) dr \right)^{\frac{1}{2}} + C \mathbb{E}^{\mathbb{P}} \left[|\Delta X_t| + [1 + |X_t|] |\Delta I_{t-}| \right]; \\ \mathcal{W}_1^2(\mathbb{P}_{\tilde{Y}_s}, \mathbb{P}_{Y_s}) &\leq C \int_s^T \mathcal{W}_1^2(\mathbb{P}_{\tilde{Y}_r}, \mathbb{P}_{Y_r}) dr + C \left(\mathbb{E}^{\mathbb{P}} \left[|\Delta X_t| + [1 + |X_t|] |\Delta I_{t-}| \right] \right)^2. \end{aligned}$$

By Grownwall inequality we have, for any $R > 0$,

$$\begin{aligned} \sup_{t \leq s \leq T} \mathcal{W}_1(\mathbb{P}_{\tilde{Y}_s}, \mathbb{P}_{Y_s}) &\leq C \mathbb{E}^{\mathbb{P}} \left[|\Delta X_t| + [1 + |X_t|] |\Delta I_{t-}| \right] \\ &\leq CR \mathcal{W}_1(m, \tilde{m}) + C \mathbb{E}^{\mathbb{P}} \left[|X_t| \mathbf{1}_{\{|X_t| \geq R\}} \right] =: \delta_R. \end{aligned} \quad (\text{IV.5.4})$$

Notice that $\tilde{\mathbb{P}} := \mathbb{P} \circ (\tilde{X}, I)^{-1} \in \mathcal{P}(t, \tilde{m})$. Then

$$\begin{aligned} &\int_t^T \hat{F}(r, \mathbb{P}_{Y_r}) dr + g(\mathbb{P}_{Y_T}) - V(t, \tilde{m}) \\ &\leq \int_t^T \mathbb{E}^{\mathbb{P}} \left[f(r, X_r, \mathbb{P}_{Y_r}) I_r - f(r, \tilde{X}_r, \mathbb{P}_{\tilde{Y}_r}) \tilde{I}_r \right] dr + [g(\mathbb{P}_{Y_T}) - g(\mathbb{P}_{\tilde{Y}_T})] \\ &\leq \rho_0 \left(\mathcal{W}_1(\mathbb{P}_{\tilde{Y}_T}, \mathbb{P}_{Y_T}) \right) + \int_t^T \mathbb{E}^{\mathbb{P}} \left[\rho_0(|\Delta X_r|) + \rho_0(\mathcal{W}_1(\mathbb{P}_{\tilde{Y}_r}, \mathbb{P}_{Y_r})) + |f(r, X_r, \mathbb{P}_{Y_r})| |\Delta I_r| \right]. \end{aligned} \quad (\text{IV.5.5})$$

The uniform regularity of f implies that

$$|f(r, X_r, \mathbb{P}_{Y_r})| \leq |f(r, 0, \mathbb{P}_{Y_r})| + C|X_r| \leq C_m[1 + |X_r|],$$

where the constant C_m may depend on m . Then, by (IV.5.2), (IV.5.3), and (IV.5.4), we have

$$\begin{aligned} \mathbb{E}^{\mathbb{P}} \left[|f(r, X_r, \mathbb{P}_{Y_r})| |\Delta I_r| \right] &\leq C_m \mathbb{E}^{\mathbb{P}} \left[[1 + |X_r|] |\Delta I_{t-}| \right] \\ &\leq C_m \mathbb{E}^{\mathbb{P}} \left[|\Delta I_{t-}| + \rho_0(\delta_R) + |\Delta X_t| + [1 + |X_t|] |\Delta I_{t-}| \right] \leq C_m [\mathcal{W}_1(m, \tilde{m}) + \rho_0(\delta_R)]. \end{aligned}$$

Plug this into (IV.5.5), we have

$$\int_t^T \hat{F}(r, \mathbb{P}_{Y_r}) dr + g(\mathbb{P}_{Y_T}) - V(t, \tilde{m}) \leq C_m [\mathcal{W}_1(m, \tilde{m}) + \rho_0(\delta_R)] + \int_t^T \mathbb{E}^{\mathbb{P}} [\rho_0(|\Delta X_r|)] dr.$$

Since $\mathbb{P} \in \mathcal{P}(t, m)$ is arbitrary, for some appropriate modulus of continuity ρ we have $V(t, m) - V(t, \tilde{m}) \leq C_m \rho(\delta_R)$. Switching m, \tilde{m} , and noticing that we may still use X_t in δ_R , we have

$$|V(t, m) - V(t, \tilde{m})| \leq C_m \rho(\delta_R). \quad (\text{IV.5.6})$$

Fix m and send $\tilde{m} \rightarrow m$ under \mathcal{W}_1 , we see that

$$\limsup_{\tilde{m} \rightarrow m} |V(t, m) - V(t, \tilde{m})| \leq C_m \rho \left(C \mathbb{E}^{\mathbb{P}} [|X_t| \mathbf{1}_{\{|X_t| \geq R\}}] \right)$$

for any $R > 0$. Now send $R \rightarrow \infty$, we see that $\lim_{\tilde{m} \rightarrow m} V(t, \tilde{m}) = V(t, m)$.

Step 2. Let $t < \tilde{t}$ and $m \in \mathcal{P}_2(\mathbf{S})$. By DPP we have

$$\begin{aligned} V(t, m) &= \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \int_t^{\tilde{t}} F(r, \mathbb{P}_{Y_r}) dr + V(\tilde{t}, \mathbb{P}_{Y_{\tilde{t}}}) = \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \int_t^{\tilde{t}} F(r, \mathbb{P}_{Y_r}) dr + V(\tilde{t}, \mathbb{P}_{Y_{\tilde{t}}}), \\ V(\tilde{t}, m) &= \sup_{m' \preceq m} V(\tilde{t}, m'). \end{aligned} \quad (\text{IV.5.7})$$

First, for any $\mathbb{P} \in \mathcal{P}(t, m)$, note that $m' := \mathbb{P} \circ (X_t, I_{\tilde{t}-})^{-1} \preceq m$, then

$$V(\tilde{t}, \mathbb{P}_{Y_{\tilde{t}}}) - V(\tilde{t}, m) \leq V(\tilde{t}, \mathbb{P}_{(X_{\tilde{t}}, I_{\tilde{t}-})}) - V(\tilde{t}, \mathbb{P}_{(X_t, I_{\tilde{t}-})}) \leq C_m \rho(\delta_R),$$

thanks to (IV.5.6) and (IV.5.4), where, following similar arguments as in Step 1,

$$\begin{aligned} \delta_R &:= C R \mathbb{E}^{\mathbb{P}} [|X_{\tilde{t}} - X_t|] + C \mathbb{E}^{\mathbb{P}} \left[|X_t| \mathbf{1}_{\{|X_t| \geq R\}} \right] \\ &\leq C R \mathbb{E}^{\mathbb{P}} [1 + |X_t|] \sqrt{\tilde{t} - t} + C \mathbb{E}^{\mathbb{P}} [|X_t| \mathbf{1}_{\{|X_t| \geq R\}}]. \end{aligned} \quad (\text{IV.5.8})$$

Since $\mathbb{P} \in \mathcal{P}(t, m)$ is arbitrary, by (IV.5.7) we have

$$V(t, m) - V(\tilde{t}, m) \leq \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \int_t^{\tilde{t}} F(r, \mathbb{P}_{Y_r}) dr + C_m \rho(\delta_R) \leq C_m \rho(\delta_R).$$

Next, for $m' \preceq m$, choose $\mathbb{P} \in \mathcal{P}(t, m)$ s.t. $I_s = I_{t-}$, $t \leq s < \tilde{t}$, and $\mathbb{P} \circ (X_t, I_{\tilde{t}})^{-1} = m'$. Then

$$V(\tilde{t}, m') - V(t, m) \leq V(\tilde{t}, \mathbb{P}_{(X_t, I_{\tilde{t}})}) - V(\tilde{t}, \mathbb{P}_{(X_{\tilde{t}}, I_{\tilde{t}})}) - \int_t^{\tilde{t}} F(r, \mathbb{P}_{Y_r}) dr \leq C_m \rho(\delta_R).$$

Since $m' \preceq m$ is arbitrary, by (IV.5.7) we have

$$V(\tilde{t}, m) - V(t, m) \leq C_m \rho(\delta_R), \quad \text{and thus} \quad |V(t, m) - V(\tilde{t}, m)| \leq C_m \rho(\delta_R).$$

This, together with (IV.5.8), implies the desired regularity immediately. \blacksquare

Proof of Lemma IV.3.6. We shall apply the results in Buckdahn, Li, Peng & Rainer [14]. For this purpose, we extend functions on $\mathcal{P}_2(\mathbf{S})$ to $\mathcal{P}_2(\mathbb{R}^d \times \mathbb{R})$. Let $\phi : \mathbb{R} \rightarrow \mathbb{R}$ be a smooth function with bounded derivatives s.t. $0 \leq \phi \leq 1$, $\phi(0) = 0$, $\phi(1) = 1$, and $\Phi : \hat{m} \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}) \mapsto m \in \mathcal{P}_2(\mathbf{S})$, with

$$m(A, 1) := \int_{\mathbb{R}} \phi(y) \hat{m}(A, dy), \quad m(A, 0) := \int_{\mathbb{R}} [1 - \phi(y)] \hat{m}(A, dy), \quad \text{for all } A \in \mathcal{B}(\mathbb{R}^d).$$

Now for $\varphi = b, \sigma, f, g$, define $\hat{\varphi}(t, x, \hat{m}) := \varphi(t, x, \Phi(\hat{m}))$. $\hat{\varphi}$ inherits the regularity of φ on $\mathcal{P}_2(\mathbb{R}^d \times \mathbb{R})$.

Next, fix a filtered probability space $(\hat{\Omega}, \hat{\mathcal{F}}_T, \hat{\mathbb{P}}, \hat{\mathbb{P}})$ on which is defined a d -dimensional Brownian motion W . For any (t, \hat{m}) , let $\xi \in \mathbb{L}^2(\mathcal{F}_t; \mathbb{R}^d)$, $\eta \in \mathbb{L}^2(\mathcal{F}_t; \mathbb{R})$ be such that $\hat{\mathbb{P}}_{(\xi, \eta)} = \hat{m}$. Consider the following SDE on $[t, T]$ with solution $\hat{Y} = (\hat{X}, \hat{I})$:

$$\hat{X}_s = \xi + \int_t^s \hat{b}(r, \hat{X}_r, \hat{\mathbb{P}}_{\hat{Y}_r}) \phi(\hat{I}_r) dr + \int_t^s \hat{\sigma}(r, \hat{X}_r, \hat{\mathbb{P}}_{\hat{Y}_r}) \phi(\hat{I}_r) dW_r; \quad \hat{I}_s = \eta \mathbf{1}_{[t, T]}(s), \quad \hat{\mathbb{P}}\text{-a.s.}$$

We then define, recalling (IV.2.1),

$$\hat{U}(t, \hat{m}) := \hat{g}(\hat{\mathbb{P}}_{\hat{Y}_T}) + \int_t^T \hat{F}(r, \hat{\mathbb{P}}_{Y_r}) dr, \quad \text{where} \quad \hat{F}(r, \hat{m}) := \int_{\mathbb{R}^{d+1}} \hat{f}(r, x, \hat{m}) i \hat{m}(dx, di).$$

We remark that, since b and σ are not necessarily bounded, the coefficients of the SDE for \hat{X} is not Lipschitz continuous in \hat{I} . However, since \hat{I} is already given, such Lipschitz continuity is not needed. In particular, we can apply [14, Lemmas 6.2 and 7.1] so that $\partial_t \hat{U}, \partial_{\hat{m}} \hat{U}, \partial_{\hat{y}\hat{m}} \hat{U}$ exist and are continuous and bounded. Here $\partial_{\hat{m}} \hat{U}$ is the Lions derivative and satisfies: $\partial_{\hat{m}} \hat{U}(t, \hat{m}, \hat{y}) := \partial_{\hat{y}} \delta_{\hat{m}} \hat{U}(t, \hat{m}, \hat{y})$, see e.g. Carmona & Delarue [21, Vol. 1, Chapter 5]. We also remark that in [14] the function \hat{U} takes the form $\hat{U}(t, x, \hat{m})$ while here \hat{U} does not have the x -variable. Moreover, note that each $m \in \mathcal{P}(\mathbf{S})$ can be viewed

as an element of $\mathcal{P}(\mathbb{R}^d \times \mathbb{R})$ with support included in \mathbf{S} . Since $\phi(0) = 0, \phi(1) = 1$, one can easily see that $U(t, m) = \hat{U}(t, m)$. Then clearly $U \in C_2^{1,1}(\mathbf{Q}_0)$. Finally, the \mathcal{W}_1 Lipschitz continuity of U follows arguments similar to that of Theorem IV.3.5, we thus omit the proof here. \blacksquare

Proof of Lemma IV.3.7 (ii) and (iii) follow directly from [61, Theorem 3.1], after the straightforward extension to $\mathcal{P}_2(\mathbf{S})$, as we will do next. Thus we shall only prove (i). For the ease of presentation, we assume $d = 1$.

Fix $n \geq 1$, we construct U_n as follows. First, let $H_n, \phi_j^n \in C^\infty(\mathbb{R}), j \in \mathbb{Z}$, satisfy:

$$\begin{aligned} 0 \leq H_n \leq 1, \quad \text{supp}(H_n) \subset [-\frac{3n}{2}, \frac{3n}{2}], \quad H_n(x) = 1 \text{ for } |x| \leq n, \quad |\partial_x H_n| \leq \frac{3}{n}; \\ 0 \leq \phi_j^n \leq 1, \quad \text{supp}(\phi_j^n) \subset [\frac{j-1}{n}, \frac{j+1}{n}], \quad \phi_j^n(x) + \phi_{j+1}^n(x) = 1 \text{ for all } x \in [\frac{j}{n}, \frac{j+1}{n}]. \end{aligned}$$

See [77, (3.3)] for a construction of ϕ_j^n . Next, for each $j \in \mathbb{Z}$, define

$$\psi_j^n(\mu) := \int_{\mathbb{R}} \phi_j^n(x) H_n(x) \mu(dx) + \mathbf{1}_{\{j=0\}} \int_{\mathbb{R}} [1 - H_n(x)] \mu(dx), \quad (\text{IV.5.9})$$

for all finite measure μ on \mathbb{R} . We emphasize that, slightly different from [77], here the μ will be $m(\cdot, i)$ whose total measure is less than 1 and thus it is not a probability measure. Note that $\psi_j^n \geq 0$ and $\sum_{j \in \mathbb{Z}} \psi_j^n = \mu(\mathbb{R})$. Moreover, denote $\mathbb{Z}_n := \{j \in \mathbb{Z} : |j| \leq 2n^2\}$ with size $N_n := 4n^2 + 1$, and

$$\Delta_n := \left\{ \vec{z} = \{z_j\}_{j \in \mathbb{Z}_n} : |z_j| \leq N_n^{-3} \text{ for all } j \neq 0, \quad \text{and} \quad z_0 := - \sum_{j \in \mathbb{Z}_n \setminus \{0\}} z_j \right\}.$$

We now define, for each $\vec{z} \in \Delta_n$ and $m \in \mathcal{P}_2(\mathbf{S}), i = 0, 1$,

$$\begin{cases} m_n(dx, i, \vec{z}) := \sum_{j \in \mathbb{Z}_n} \hat{\psi}_j^n(m(\cdot, i), \vec{z}) \delta_{\frac{j}{n}}(dx), \\ \hat{\psi}_j^n(\mu, \vec{z}) := \frac{N_n}{N_n + 1} \left[\psi_j^n(\mu) + \mu(\mathbb{R}) \left[\frac{1}{N_n^2} + z_j \right] \right]. \end{cases} \quad (\text{IV.5.10})$$

Note that $|z_0| \leq N_n^{-2}$, and thus $\hat{\psi}_j^n(\mu, \vec{z}) \geq 0$. One may easily verify that

$$\begin{aligned} \sum_{j \in \mathbb{Z}_n} \hat{\psi}_j^n(\mu, \vec{z}) &= \frac{N_n}{N_n + 1} \left[\sum_{j \in \mathbb{Z}_n} \psi_j^n(\mu) + \frac{\mu(\mathbb{R})}{N_n} \right] = \frac{N_n}{N_n + 1} \left[\mu(\mathbb{R}) + \frac{\mu(\mathbb{R})}{N_n} \right] = \mu(\mathbb{R}); \\ m_n(\mathbb{R}, i, \vec{z}) &= \sum_{j \in \mathbb{Z}_n} \hat{\psi}_j^n(m(\cdot, i), \vec{z}) = m(\mathbb{R}, i), \quad \text{and thus} \quad m_n(\mathbf{S}, \vec{z}) = 1. \end{aligned}$$

In particular, this implies that $m_n(\cdot, \vec{z}) \in \mathcal{P}_2(\mathbf{S})$ for every $\vec{z} \in \Delta_n$, where the square integrability follows from the fact that $\text{supp}(m_n(\cdot, \vec{z}))$ is finite. Finally, let ζ_n be a smooth density function with support Δ_n , and we construct

$$U_n(m) := \int_{\Delta_n} U(m_n(\cdot, \vec{z})) \zeta_n(\vec{z}) d\vec{z}, \quad m \in \mathcal{P}_2(\mathbf{S}). \quad (\text{IV.5.11})$$

The smoothness of U_n follows from the same arguments as in [61, Theorem 3.1]. However, we note that [61] uses the \mathcal{W}_1 -distance and requires \mathcal{M} to be a compact subset of $\mathcal{P}_1(\mathbf{S})$. This is mainly for the uniform Lipschitz continuity of U_n which holds only under \mathcal{W}_1 . Here we provide a proof for the uniform convergence of U_n under \mathcal{W}_2 . We first show that

$$\overline{\mathcal{M}} := \left\{ m_n(\cdot, \vec{z}) : m \in \mathcal{M}, n \geq 1, \vec{z} \in \Delta_n \right\} \subset \mathcal{P}_2(\mathbf{S}) \quad \text{is compact.} \quad (\text{IV.5.12})$$

Indeed, fix $R > 0$. Denote $\mathbb{Z}_n^R := \{j \in \mathbb{Z}_n : |j| \geq nR\}$ for $n > \frac{R}{2}$. Then

$$\begin{aligned} \int_{\{|x|>R\}} |x|^2 m_n(dx, i, \vec{z}) &= \sum_{j \in \mathbb{Z}_n^R} \frac{j^2}{n^2} \hat{\psi}_j^n(m(\cdot, i), \vec{z}) \\ &= \sum_{j \in \mathbb{Z}_n^R} \frac{j^2}{n^2} \frac{N_n}{N_n + 1} \left[\psi_j^n(m(\cdot, i)) + m(\mathbb{R}, i) \left[\frac{1}{N_n^2} + z_j \right] \right]. \end{aligned}$$

From the construction of ψ_j^n , one can easily verify that

$$\sum_{j \in \mathbb{Z}_n^R} \frac{j^2}{n^2} \psi_j^n(m(\cdot, i)) \leq 2 \int_{\{|x|>R\}} |x|^2 m(dx, i).$$

Moreover, note that $|z_j| \leq N_n^{-3}$ for all $j \in \mathbb{Z}_n^R$. Then, for $n > \frac{R}{2}$,

$$\begin{aligned} \int_{\{|x|>R\}} |x|^2 m_n(dx, i, \vec{z}) &\leq 2 \int_{\{|x|>R\}} |x|^2 m(dx, i) + \sum_{j \in \mathbb{Z}_n^R} \frac{j^2}{n^2} m(\mathbb{R}, i) \frac{C}{N_n^2} \\ &\leq 2 \int_{\{|x|>R\}} |x|^2 m(dx, i) + \frac{Cm(\mathbb{R}, i)}{N_n} \leq 2 \int_{\{|x|>R\}} |x|^2 m(dx, i) + \frac{Cm(\mathbb{R}^d, i)}{R^2}. \end{aligned}$$

On the other hand, when $n < \frac{R}{2}$, we have $\int_{\{|x|>R\}} |x|^2 m_n(dx, i, \vec{z}) = 0$. Thus,

$$\sup_{m \in \mathcal{M}, n \geq 1, \vec{z} \in \Delta_n} \sum_{i=0,1} \int_{\{|x|>R\}} |x|^2 m_n(dx, i, \vec{z}) \leq 2 \sup_{m \in \mathcal{M}} \sum_{i=0,1} \int_{\{|x|>R\}} |x|^2 m(dx, i) + \frac{C}{R^2}.$$

Since $\mathcal{M} \subset \mathcal{P}_2(\mathbf{S})$ is compact, we have $\lim_{R \rightarrow \infty} \sup_{m \in \mathcal{M}} \sum_{i=0,1} \int_{\{|x|>R\}} |x|^2 m(dx, i) = 0$. Then

$$\lim_{R \rightarrow \infty} \sup_{m \in \mathcal{M}, n \geq 1, \vec{z} \in \Delta_n} \sum_{i=0,1} \int_{\{|x|>R\}} |x|^2 m_n(dx, i, \vec{z}) = 0.$$

This proves that $\overline{\mathcal{M}}$ is uniformly square integrable, and therefore compact in $\mathcal{P}_2(\mathbf{S})$.

Next, note that \mathcal{M} is also compact in $\mathcal{P}_1(\mathbf{S})$, by [61, (3.15)] we have

$$\lim_{n \rightarrow \infty} \sup_{m \in \mathcal{M}, \vec{z} \in \Delta_n} \mathcal{W}_1(m_n(\cdot, \vec{z}), m) = 0. \quad (\text{IV.5.13})$$

Then, for any $R > 0$,

$$\mathcal{W}_2^2(m_n(\cdot, \vec{z}), m) \leq R \mathcal{W}_1(m_n(\cdot, \vec{z}), m) + C \sum_{i=0,1} \int_{|x| \geq \frac{R}{2}} |x|^2 [m_n(dx, i, \vec{z}) + m(dx, i)].$$

This, together with the uniform integrability of $\overline{\mathcal{M}}$ and (IV.5.13), implies immediately that

$$\lim_{n \rightarrow \infty} \sup_{m \in \mathcal{M}, \vec{z} \in \Delta_n} \mathcal{W}_2(m_n(\cdot, \vec{z}), m) = 0. \quad (\text{IV.5.14})$$

Finally, by the compactness (IV.5.12), we see that U is uniformly continuous on $\overline{\mathcal{M}}$. Then it follows from (IV.5.11) and (IV.5.14) that $\lim_{n \rightarrow \infty} \sup_{m \in \mathcal{M}} |U_n(m) - U(m)| = 0$. \blacksquare

Remark IV.5.1 While not used in the thesis, the following property is interesting in its own right: if U is monotone under \preceq , then so is the U_n constructed in (IV.5.11). Indeed, assume U is increasing, and let $m' \preceq m$ with transition probability p . For each $\vec{z} \in \Delta_n$, by (IV.5.9) and (IV.5.10), it is clear that

$$0 < \hat{\psi}_j^n(m'(\cdot, 1), \vec{z}) \leq \hat{\psi}_j^n(m(\cdot, 1), \vec{z}), \quad \text{and thus} \quad \hat{p}_j(\vec{z}) := \hat{\psi}_j^n(m'(\cdot, 1), \vec{z}) / \hat{\psi}_j^n(m(\cdot, 1), \vec{z}) \in (0, 1].$$

Since $m'(dx, \{0, 1\}) = m(dx, \{0, 1\})$, it is also obvious that

$$\sum_{i=0,1} \hat{\psi}_j^n(m'(\cdot, i), \vec{z}) = \sum_{i=0,1} \hat{\psi}_j^n(m(\cdot, i), \vec{z}).$$

Then $m'_n(\cdot, \vec{z}) \leq m_n(\cdot, \vec{z})$ for each $\vec{z} \in \Delta_n$, with transition probability $\hat{p}(\cdot, \vec{z})$ satisfying $\hat{p}(\frac{j}{n}, \vec{z}) = \hat{p}_j(\vec{z})$ for all $j \in \mathbb{Z}_n$. Then, since U is increasing, by (IV.5.11) we see that $U_n(m') \leq U_n(m)$. \blacksquare

Proof of Lemma IV.3.8. (i) Let $m' \preceq m$ with transition probability p . As m is a probability measure on $(\mathbf{S}, \mathcal{B}(\mathbf{S}))$, it is a Radon measure. Then, by Lusin's theorem (see Folland [43, 7.10]), we may find for all $k \geq 1$ a continuous $p_k : \mathbb{R}^d \rightarrow [0, 1]$ s.t.

$$m(\{x : p(x) \neq p_k(x)\}, \{0, 1\}) \leq \frac{1}{k}.$$

Let $\{m'_k\}_{k \geq 1}$ be the measures obtained from m by applying the transition probabilities $\{p_k\}_{k \geq 1}$, and ϕ a bounded and continuous function. Then

$$\left| \int_{\mathbb{R}^d} \phi(x) p_k(x) m(dx, 1) - \int_{\mathbb{R}^d} \phi(x) p(x) m(dx, 1) \right| \leq \frac{2}{k} \|\phi\|_\infty,$$

and thus $m'_k(dx, 1)$ converges weakly to $m'(dx, 1)$. We do similarly with $m'_k(dx, 0)$, and thus m'_k converges weakly to m' . As $\{m'_k\}_{k \geq 1}$ is uniformly integrable, we have $\lim_{k \rightarrow \infty} \mathcal{W}_2(m'_k, m') = 0$. As v is nondecreasing for \preceq , we have $v(m) \geq v(m'_k)$ for all $k \geq 1$. Then, as v is LSC, we have $v(m) \geq \liminf_{k \rightarrow \infty} v(m'_k) \geq v(m')$.

(ii) As $(\mathbb{D}_I \varphi)_*$ is LSC, there exists $\delta > 0$ s.t. $(\mathbb{D}_I \varphi)_* \geq 0$ on $[t, t + \delta] \times \mathcal{B}_{\mathcal{W}_2}(m, \delta)$. Let $(s, m_0), (s, m_1)$ be in this neighborhood, s.t. $m_1 \preceq m_0$ with transition probability p . Then, we have

$$\varphi(s, m_0) - \varphi(s, m_1) = \int_0^1 \int_{\mathbb{R}^d} D_I \varphi(t, \lambda m_0 + (1 - \lambda) m_1, x) (1 - p(x)) m(dx, 1) d\lambda.$$

By convexity of $\mathcal{B}_{\mathcal{W}_2}(m, \delta)$, we have $D_I \varphi(t, \lambda m_0 + (1 - \lambda) m_1, \cdot) \geq 0$, hence the desired result.

■

Chapter V

A finite-dimensional approximation for the obstacle problem on Wasserstein space

V.1 Introduction

Lasry & Lions [53] and Caines, Huang & Malhamé [47] introduced mean field games to study systems of interacting agents as their number goes to infinity. In [17], Cardaliaguet, Delarue, Lasry & Lions indeed proved the convergence of the N -players game to the corresponding mean field game, provided that the value function of the problem is smooth. Since then, an extensive part of the literature has been dedicated to these convergence issues: see e.g. Cardaliaguet [16], Bayraktar, Cecchin, Cohen & Delarue [5, 4], Cecchin & Pelino [25], Cecchin, Dai Pra, Fischer & Pelino [26], Laurière & Tangpi [55], Djete [32] or Doncel, Gast & Gaujal [35].

However, the literature is not as extensive when it comes to the convergence of mean field control problems. We may mention Lacker [52], who proved a propagation of chaos result for optimally controlled McKean-Vlasov dynamics, or Fischer & Livieri [42], who solved the mean-variance control problem by first solving its finite-dimensional approximation and then proving the convergence of the solution as the dimension goes to infinity. We also refer to Cecchin [24], Bayraktar, Cecchin & Chakraborty [3] and Cardaliaguet & Souganidis [19].

The two contributions the most relevant to our work are the ones of Gangbo, Mayorga

& Swiech [44] and Talbi [73] (which correspond to Chapter VI of this thesis). In the first paper, the authors develop a finite-dimensional approximation for Hamilton-Jacobi-Bellman equations with uncontrolled volatility, in the sense of viscosity solutions (defined via lifting on the Hilbert space of square integrable random variables). The second paper introduces an approximation for a class of partial differential equations on Wasserstein space, which covers the scope of mean field control problems.

Our present objective is to find a finite-dimensional approximation of the mean field optimal stopping problem introduced in Chapter III, which we characterized by a dynamic programming equation on Wasserstein space, that we call obstacle problem by analogy with the equation corresponding to the standard optimal stopping problem (see e.g. El Karoui [41] or Shiryaev [70]). More precisely, after developing in Chapter IV an appropriate notion of viscosity solution for our obstacle problem, we now intend to find a finite-dimensional PDE whose viscosity solutions converge to viscosity solutions of the equation on Wasserstein space.

This finite-dimensional PDE (that we call cascade obstacle due to its remarkable structure) naturally turns out to be related to the multiple optimal stopping problem, which consists in optimizing some reward of a system of interacting particles by assigning each particle a stopping time (see Kobylanski, Quenez and Rouy-Mironescu [50] for a review of this problem in a very general setting). In the spirit of Barles & Souganidis [2], we prove that the semi-relaxed limits of the viscosity super/subsolutions of the cascade obstacle are viscosity super/subsolutions of the obstacle on the Wasserstein space as the number of particles goes to infinity. We also complete this result with a propagation of chaos theorem for stopped McKean-Vlasov diffusion (see Snitizman [72] for the case of classical diffusions).

An important feature that ensures the convergence relies in the choice of the set of the test functions for the finite dimensional problem. Similarly to the viscosity theory developed for path-dependent PDEs (see e.g. Ekren, Keller, Touzi & Zhang [37], Ekren, Touzi & Zhang [39, 40], Ekren [36] and Ren, Touzi & Zhang [67]), we only require test functions to be tangent to the super/subsolution through the mean, whereas the tangency is pointwise in the standard literature. This tangency through the mean can be seen as the finite-dimensional counterpart of our set of test functions on the Wasserstein space.

The chapter is organized as follows. In §V.2, we present the main notions of the chapter, i.e. mean field optimal stopping problem, the obstacle equation on Wasserstein space, the multiple optimal stopping problem and the cascade obstacle equation. §V.3 states the main

results of the chapter, namely, the convergence of the value function and the propagation of chaos. §V.4 introduces the dynamic programming equations characterizing the mean field and N particles problems. §V.5 and §V.6 are dedicated to the proofs of our main results. In §V.7, we study the multiple optimal stopping problem and specify our notion of viscosity solutions for the cascade obstacle equation. Finally, we provide a technical lemma in the appendix.

Notations. Given $x \in \mathbb{R}^{d \times N}$, we denote by $m^N(x) := \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \in \mathcal{P}(\mathbb{R}^d)$.

V.2 Presentation of the problem

V.2.1 The mean field optimal stopping

Let $T < \infty$ be fixed, and $\Omega := C^0([-1, T], \mathbb{R}^d) \times \mathbb{I}^0([-1, T])$ the canonical space, where:

- $C^0([-1, T], \mathbb{R}^d)$ is the set of continuous paths from $[-1, T]$ to \mathbb{R}^d , constant on $[-1, 0)$;
- $\mathbb{I}^0([-1, T])$ is the set of non-increasing and càdlàg maps from $[-1, T]$ to $\{0, 1\}$, constant on $[-1, 0)$, and ending with value 0 at T .

We equip Ω with the Skorokhod distance, under which it is a Polish space. We denote $Y := (X, I)$ the canonical process, with state space $\mathbf{S} := \mathbb{R}^d \times \{0, 1\}$, its canonical filtration $\mathbb{F} = (\mathcal{F}_t)_{t \in [-1, T]}$, and the corresponding jump time of the survival process I :

$$\tau := \inf\{t \geq 0 : I_t = 0\}, \text{ so that } I_t := I_{0-\mathbf{1}_{t < \tau}} \text{ for all } t \in [-1, T]. \quad (\text{V.2.1})$$

By the càdlàg property of I , τ is a \mathbb{F} -stopping time. Note that the choice of the extension to -1 is arbitrary, the extension of time to the left of the origin is only needed to allow for an immediate stop at time $t = 0$.

Let $(b, \sigma, f) : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbf{S}) \rightarrow \mathbb{R}^d \times \mathbb{R}^{d \times d} \times \mathbb{R}$ with σ taking values in the set of non-negative matrices, and $g : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$. In the following assumption, which will always be in force throughout the chapter, $\mathcal{P}_2(\mathbf{S})$ is equipped with the \mathcal{W}_2 distance.

Assumption V.2.1 (i) b, σ are continuous in t , and uniformly Lipschitz continuous in (x, m) .

(ii) f is Borel measurable and has quadratic growth in $x \in \mathbb{R}^d$, and

$$F(t, m) := \int_{\mathbb{R}^d} f(t, x, m) m(dx, 1) \text{ is continuous on } [0, T] \times \mathcal{P}_2(\mathbf{S}). \quad (\text{V.2.2})$$

(iii) g is upper-semicontinuous and locally bounded.

We also extend g to $\mathcal{P}_2(\mathbf{S})$ by $g(m) := g(m(\cdot, \{0, 1\}))$. Introduce the dynamic value function

$$V(t, m) := \sup_{\mathbb{P} \in \mathcal{P}(t, m)} \int_t^T F(r, \mathbb{P}_{Y_r}) dr + g(\mathbb{P}_{Y_T}), \quad (t, m) \in [0, T] \times \mathcal{P}_2(\mathbf{S}). \quad (\text{V.2.3})$$

Here $\mathcal{P}(t, m)$ is the set of probability measures \mathbb{P} on (Ω, \mathcal{F}_T) s.t. $\mathbb{P}_{Y_{t-}} = m$ and the processes:

$$M := X - \int_t^\cdot b(r, X_r, \mathbb{P}_{Y_r}) I_r dr \quad \text{and} \quad M \cdot M^\top - \int_t^\cdot \sigma^2(r, X_r, \mathbb{P}_{Y_r}) I_r dr \quad (\text{V.2.4})$$

are \mathbb{P} -martingales on $[t, T]$, that is, for some \mathbb{P} -Brownian motion $W^\mathbb{P}$,

$$X_s = X_t + \int_t^s b(r, X_r, \mathbb{P}_{Y_r}) I_r dr + \sigma(r, X_r, \mathbb{P}_{Y_r}) I_r dW_r^\mathbb{P}, \quad I_s = I_t - \mathbf{1}_{s < \tau}, \quad \mathbb{P} - \text{a.s.} \quad (\text{V.2.5})$$

A special element of $\mathcal{P}(t, m)$ is $\bar{\mathbb{P}} = \bar{\mathbb{P}}^{t, m}$ under which X is unstopped. That is,

$$X_s = X_t + \int_t^s b(r, X_r, \bar{\mathbb{P}}_{Y_r}) I_r dr + \int_t^s \sigma(r, X_r, \bar{\mathbb{P}}_{Y_r}) I_r dW_r^{\bar{\mathbb{P}}}, \quad I_s = I_t - \mathbf{1}_{[t, T)}(s), \quad \bar{\mathbb{P}}\text{-a.s.} \quad (\text{V.2.6})$$

Note that $Y_\cdot = Y_{\cdot \wedge \tau}$, and in particular $Y_T = Y_\tau$, \mathbb{P} -a.s. Moreover, from the definition of F in (V.2.2), we have $\int_t^T F(r, \mathbb{P}_{Y_r}) dr = \mathbb{E}^\mathbb{P} \int_t^T f(r, X_r, \mathbb{P}_{Y_r}) dr$.

We recall from Chapter III that $\mathcal{P}(t, m)$ is compact under the Wasserstein distance \mathcal{W}_2 , and thus existence holds for the mean field optimal stopping problem (V.2.3).

V.2.2 Differential calculus on Wasserstein space

We recall some differential calculus tools on the Wasserstein space. We say that a function $u : \mathcal{P}_2(\mathbf{S}) \rightarrow \mathbb{R}$ has a functional linear derivative $\delta_m u : \mathcal{P}_2(\mathbf{S}) \times \mathbf{S} \rightarrow \mathbb{R}$ if

$$u(m') - u(m) = \int_0^1 \int_{\mathbf{S}} \delta_m u(\lambda m' + (1 - \lambda)m, y) (m' - m)(dy) d\lambda \quad \text{for all } m, m' \in \mathcal{P}_2(\mathbf{S}),$$

and $\delta_m u(m, \cdot)$ has quadratic growth, locally uniformly in m , so as to guarantee integrability in the last expression. As in Chapter III, we denote

$$\delta_m u_i(t, m, x) := \delta_m u(t, m, x, i) \quad \text{for } i \in \{0, 1\}, \quad D_I u := \delta_m u_1 - \delta_m u_0,$$

and we introduce the measure flow generator of X

$$\mathbb{L}u(t, m) := \partial_t u(t, m) + \int_{\mathbb{R}^d} \mathcal{L}_x \delta_m u_1(t, m, x) m(dx, 1), \quad (\text{V.2.7})$$

where \mathcal{L}_x is the generator of X :

$$\mathcal{L}_x \delta_m u_1 := b \cdot \partial_x \delta_m u_1 + \frac{1}{2} \sigma^2 : \partial_{xx}^2 \delta_m u_1.$$

We also write $\partial_{\mu\mu}^2 u_{1,1}(t, m, x, x') := \partial_{x'} \delta_m \partial_x \delta_m u(t, m, x, 1, x', 1)$. Throughout this chapter, we denote by

$$\mathbf{Q}_t := [t, T] \times \mathcal{P}_2(\mathbf{S}), \quad \text{and} \quad \overline{\mathbf{Q}}_t := [t, T] \times \mathcal{P}_2(\mathbf{S}), \quad t \in [0, T].$$

V.2.3 The multiple optimal stopping problem

Let $(\Omega^0, \mathcal{F}_T^0, \mathbb{F}^0, \mathbb{P}^0)$ be a filtered probability space, equipped with a sequence of d -dimensional Brownian motions $\{W^N\}_{N \geq 1}$. We denote \mathbb{F}^N the canonical completed filtration of $\{W^k\}_{1 \leq k \leq N}$. We also denote $\mathcal{T}_{t,T}^N$ the set of the $[t, T]$ -valued \mathbb{F}^N -stopping times for all $t \in [0, T]$, and $[N] := \{1, \dots, N\}$, and $\mathcal{T}_{t,T}^N := (\mathcal{T}_{t,T}^N)^N$.

We recall that $m^N(\mathbf{x}) := \frac{1}{N} \sum_{k=1}^N \delta_{x_k}$. For $\tau \in \mathcal{T}_{t,T}^N$, let \mathbf{X}^τ be the solution of the stochastic differential equation

$$\begin{cases} X_t^{k,\tau} = x_k + \int_0^t b(s, X_s^{k,\tau}, m^N(\mathbf{X}_s^\tau)) I_s^k ds + \int_0^t \sigma(s, X_s^{k,\tau}, m^N(\mathbf{X}_s^\tau)) I_s^k dW_s^k, \\ I_s^k = i_k \mathbf{1}_{\tau_k > t}, \end{cases} \quad (\text{V.2.8})$$

for all $k \in [N]$. We define the multiple optimal stopping problem

$$V^N(t, \mathbf{y}) := \sup_{\tau \in \mathcal{T}_{t,T}^N} \mathbb{E}_{t,\mathbf{y}} \left[\frac{1}{N} \sum_{k=1}^N \int_t^{\tau} f(r, X_r^{k,\tau}, m^N(\mathbf{X}_r^\tau)) I_r^k dr + g(m^N(\mathbf{X}_T^\tau)) \right]. \quad (\text{V.2.9})$$

V.3 Main results

Our first main result states the convergence of the value function of the multiple optimal stopping problem to the corresponding mean field problem. We recall the assumptions needed for our comparison principle Theorem IV.3.11:

Assumption V.3.1 *b, σ, f and g can be extended to $\mathcal{P}_1(\mathbf{S})$ under \mathcal{W}_1 continuously, b, g are uniformly Lipschitz continuous in (x, m) under \mathcal{W}_1 and all the derivatives of σ up to order 2 are bounded and Lipschitz-continuous.*

Theorem V.3.2 *Let Assumption V.3.1 hold, and V^N and V be respectively the value functions of the multiple and mean field optimal stopping problems (V.2.9) and (V.2.3). Then V^N converges to V , i.e.*

$$V(t, m) := \lim_{\substack{N \rightarrow \infty, s \rightarrow t \\ m^N(\mathbf{y}^N) \xrightarrow{\mathcal{W}_2} m}} V^N(s, m^N(\mathbf{y}^N)) \quad \text{for all } (t, m) \in \overline{\mathbf{Q}}_0.$$

This result is proved in §V.5 by using the dynamic programming equations, which characterize the finite population and the mean field optimal stopping problems. In particular, our proof relies heavily on Theorem V.5.1, which is stated and proved in §V.5.

Our next main result states the convergence of optimal stopping strategies:

Theorem V.3.3 (i) *Let the assumptions of Theorem V.3.2 hold and $\hat{\tau}^N \in \mathcal{T}_{0,T}^N$ be ε_N -optimal for (V.7.2) for all $N \geq 1$, with $\varepsilon_N \searrow 0$. Assume $m^N(\mathbf{Y}_{0-}^{\hat{\tau}^N}) \xrightarrow{W_2} m$ for some $m \in \mathcal{P}_2(\mathbf{S})$, and that f and g are bounded. Then, $\{\mathbb{P}^0 \circ (m^N(\mathbf{Y}^{\hat{\tau}^N}))^{-1}\}_{N \geq 1}$ is tight, and all its accumulation points are supported on the set of the optimal controls for (V.2.3).*

(ii) *In particular, if (V.2.3) has a unique solution \mathbb{P}^* and $\{\hat{\tau}^N\}_{N \geq 1}$ is a sequence of ε_N -optimal stopping strategies for (V.2.9), then $\{m^N(\mathbf{Y}^{\hat{\tau}^N})\}_{N \geq 1}$ converges weakly to \mathbb{P}^* .*

We finally prove that any control for the mean field problem may be approximated by controls for the N -particles problem. In particular, even when (V.2.3) has multiple solutions, optimal controls may be approximated by a sequence of ε_N -optimal controls for the N -particles problem, with $\varepsilon_N \searrow 0$.

Theorem V.3.4 *Fix $m \in \mathcal{P}_2(\mathbf{S})$, and \mathbf{y}^N , $N \geq 1$, such that $m^N(\mathbf{y}^N) \xrightarrow{W_2} m$.*

(i) *Let $\mathbb{P} \in \mathcal{P}(0, m)$. There exists $\tau^N \in \mathcal{T}_{0,T}^N$, $N \geq 1$, s.t. the process \mathbf{Y}^{τ^N} started from $\mathbf{Y}_{t^N}^{\tau^N} = \mathbf{y}^N$ satisfies*

$$m^N(\mathbf{Y}^{\tau^N}) \xrightarrow{W_2} \mathbb{P}, \mathbb{P}^0\text{-a.s.}$$

(ii) *Let $\mathbb{P}^* \in \mathcal{P}(0, m)$ be optimal for (V.2.3), and let Assumption V.3.1 hold. Assume further than g and F are bounded. Then there exists $\varepsilon_N \searrow 0$, $\hat{\tau}^N \in \mathcal{T}_{0,T}^N$ that is ε_N -optimal for (V.2.9) s.t. the process $\mathbf{Y}^{\hat{\tau}^N}$ started from $\mathbf{Y}_{t^N}^{\hat{\tau}^N} = \mathbf{y}^N$ satisfies*

$$m^N(\mathbf{Y}^{\hat{\tau}^N}) \xrightarrow{W_2} \mathbb{P}^*, \mathbb{P}^0\text{-a.s.}$$

Theorems V.3.3 and V.3.4 are proved in §V.6.

V.4 Dynamic programming equations

In this section, we define the two equations that will be crucial to prove our main convergence result Theorem V.3.2.

V.4.1 Obstacle equation on Wasserstein space

Definition V.4.1 Let $C_2^{1,2}(\mathbf{Q}_0)$ be the set of functions $u : \overline{\mathbf{Q}}_0 \rightarrow \mathbb{R}$ s.t.

- $\partial_t u, \delta_m u, \partial_x \delta_m u_1, \partial_{xx}^2 \delta_m u_1$ exist, and are continuous in all variables,
- $\partial_{xx}^2 \delta_m u_1$ is bounded in x , locally uniformly in (t, m) ,
- $\partial_{\mu\mu}^2 u_{1,1}$ exists and $(t, m, x) \mapsto \partial_{\mu\mu}^2 u_{1,1} u(t, m, x, x)$ is continuous and has quadratic growth in x , locally uniformly in (t, m) .

Remark V.4.2 In the present chapter, the set $C_2^{1,2}(\mathbf{Q}_0)$ has a different definition from the one we used in the previous chapters, as we additionally require the existence of the second order derivative $\partial_{\mu\mu}^2 u_{1,1}$. However, we can see that the viscosity theory implied by requiring such regularity for the test functions is equivalent to the one we developed in Chapter IV; in particular, the test functions involved in the proof of our comparison result Theorem IV.3.11 are infinitely differentiable, so that the validity of this result does not depend on the definition of $C_2^{1,2}(\mathbf{Q}_0)$ we set. ■

Introduce a partial order \preceq on $\mathcal{P}_2(\mathbf{S})$: we say that $m' \preceq m$ if

$$m'(dx, 1) = p(x)m(dx, 1), \text{ and } m'(dx, 0) = [1 - p(x)]m(dx, 1) + m(dx, 0), \quad (\text{V.4.1})$$

for some measurable $p : \mathbb{R}^d \rightarrow [0, 1]$. In our context, $m_{t-} = \mathbb{P}_{(X_t, I_{t-})}$ and $m_t = \mathbb{P}_{(X_t, I_t)}$, with $\mathbb{P} \in \mathcal{P}(t, m)$, so that $m_t \preceq m_{t-}$ with conditional transition probability $p(x) = p_t(x) := \mathbb{P}(I_t = 1 \mid X_t = x, I_{t-} = 1)$. When $p = \mathbf{1}_A$, with $A \in \mathcal{B}(\mathbb{R}^d)$, we denote m^A the measure defined by (V.4.1). We recall from the previous chapters that the set $\{m' : m' \preceq m\}$ is compact.

In Chapter IV, we proved that V is the unique viscosity solution of the obstacle problem on Wasserstein space

$$\min \left\{ \min_{m' \in C_u(t, m)} [- (\mathbb{L}u + F)(t, m')], (\mathbb{D}_I u)_*(t, m) \right\} = 0, \quad (t, m) \in \mathbf{Q}_0, \quad (\text{V.4.2})$$

with boundary condition $u|_{t=T} = g$. The function $(\mathbb{D}_I u)_*$ is the LSC envelope of

$$\mathbb{D}_I u : (t, m) \mapsto \inf_{x \in \text{Supp}(m(\cdot, 1))} D_I u(t, m, x),$$

which is upper semicontinuous, but not continuous, in general, and the set

$$C_u(t, m) := \{m' \preceq m : u(t, m') \geq u(t, m)\}, \quad (t, m) \in \mathbf{Q}_0,$$

indicates the set of positions in \mathbf{Q}_0 which improve u by stopping the corresponding particles.

We recall from Chapter IV that, under appropriate assumptions, the value function V is the unique viscosity solution of (V.4.2).

V.4.2 The cascade obstacle equation

One essential observation to study (V.2.9) is the following result, which is proved in [50]:

Proposition V.4.3 *For all $(t, \mathbf{y}) \in [0, T] \times \mathbf{S}^N$, we have*

$$V^N(t, \mathbf{y}) = \sup_{\theta \in \mathcal{T}_{t,T}^N} \mathbb{E}_{t, \mathbf{y}} \left[\frac{1}{N} \sum_{k=1}^N \int_t^\theta f^k(r, \bar{\mathbf{X}}_r) I_r^k dr + \max_{\mathbf{i}' < \mathbf{i}} V^N(\theta, \bar{\mathbf{X}}_\theta, \mathbf{i}') \right], \quad (\text{V.4.3})$$

where $\mathbf{i} < \mathbf{i}'$ means $i_k \leq i'_k$ for all $k \in [N]$ and $\mathbf{i} \neq \mathbf{i}'$.

Consider the partial differential equation

$$\min\{-\partial_t u(t, \mathbf{y}) - \mathbf{i} \cdot (\mathcal{L}u + \mathbf{f})(t, \mathbf{y}), u(t, \mathbf{y}) - \max_{\mathbf{i}' < \mathbf{i}} u(t, \mathbf{x}, \mathbf{i}')\} = 0, u|_{t=T} = u|_{\mathbf{i}=0} = g^N, \quad (\text{V.4.4})$$

for all $(t, \mathbf{y}) \in [0, T] \times \mathbf{S}^N$, where $\mathbf{i} < \mathbf{i}'$ means $i_k \leq i'_k$ for all $k \in [N]$ and $\mathbf{i} \neq \mathbf{i}'$, and

$$\mathcal{L}\varphi := \left(b_k \partial_{x_k} \varphi + \frac{1}{2} \sigma_k^2 \partial_{x_k x_k}^2 \varphi \right)_{k \in [N]} \quad \text{for all } \varphi \in C^2(\mathbb{R}^{d \times N}), \quad (\text{V.4.5})$$

with $b_k(t, \mathbf{x}) := b(t, x_k, m^N(\mathbf{x}))$, $\sigma_k(t, \mathbf{x}) := \sigma(t, x_k, m^N(\mathbf{x}))$, $\mathbf{f}(\cdot, \mathbf{x}) = (f^1, \dots, f^N)^\top(\cdot, \mathbf{x}) := (f(\cdot, x_1, m^N(\mathbf{x})), \dots, f(\cdot, x_N, m^N(\mathbf{x})))^\top$ and $g^N(\mathbf{x}) := g(m^N(\mathbf{x}))$ for all $\mathbf{x} \in \mathbb{R}^{d \times N}$. We call (V.4.4) *cascade obstacle equation*, for which we define viscosity solutions (see Definition V.7.3). In Section V.7, we prove (in a more general framework which covers the present setting) the following results:

Theorem V.4.4 (Existence) *Assume \mathbf{f} and g^N are Lipschitz-continuous. Then V^N is a continuous viscosity solution of (V.4.4).*

Remark V.4.5 *By standard techniques, a comparison result for this equation can be proved by direct iteration of the comparison result for the standard obstacle problem for single optimal stopping problems.*

These results are proved in §V.7, where we study the finite population multiple optimal stopping problem in a more general framework.

V.5 Convergence of the value function

Introduce $\mathcal{P}^N(\mathbf{S}) := \{m^N(\mathbf{y}) : \mathbf{y} \in \mathbf{S}^N\}$. Let \mathcal{S}^N be the set of functions $w : [0, T] \times \mathbf{S}^N \rightarrow \mathbb{R}$ such that $w(t, \mathbf{y}) = w^N(t, m^N(\mathbf{y}))$ for some $w^N : [0, T] \times \mathcal{P}^N(\mathbf{S})$. If $\{w^N \in \mathcal{S}^N\}_{N \geq 1}$ be a sequence of locally bounded functions, uniformly in N , we denote

$$\underline{w}(t, m) := \liminf_{\substack{N \rightarrow \infty, s \rightarrow t \\ m^N(\mathbf{y}^N) \xrightarrow{\mathcal{W}_2} m}} w^N(s, m^N(\mathbf{y}^N)), \quad \text{and} \quad \bar{w}(t, m) := \limsup_{\substack{N \rightarrow \infty, s \rightarrow t \\ m^N(\mathbf{y}^N) \xrightarrow{\mathcal{W}_2} m}} w^N(s, m^N(\mathbf{y}^N)),$$

for all $(t, m) \in \mathbf{Q}_0$.

Theorem V.5.1 (i) *Let $\{u^N \in \mathcal{S}^N\}_{N \geq 1}$ be a sequence of continuous and locally bounded, uniformly in N , viscosity subsolutions of (V.4.4). Then \bar{u} is a viscosity subsolution of (V.4.2).*

(ii) *Let $\{v^N \in \mathcal{S}^N\}_{N \geq 1}$ be a sequence of continuous and locally bounded, uniformly in N , viscosity supersolutions of (V.4.4). If \underline{v} is continuous, then it is a viscosity supersolution of (V.4.2).*

Proof of Theorem V.5.1

(i) We first prove that \bar{u} is a viscosity subsolution of (V.4.2).

Fix $(t, m) \in \mathbf{Q}_0$ and $\varphi \in \underline{A}\bar{u}(t, m)$ with corresponding $\delta_0 \in (0, T - t)$. We assume w.l.o.g. (see [75, Remark 3.2]) that

$$(t, m) \text{ is a strict minimum of } (\varphi - \bar{u}) \text{ on } \mathcal{N}_\delta(t, m), \text{ with } C_{\bar{u}}(t, m) = \{m\}. \quad (\text{V.5.1})$$

Let (t^N, \mathbf{y}^N) be such that $(t^N, m^N(\mathbf{y}^N)) \rightarrow (t, m)$ and $u^N(t^N, m^N(\mathbf{y}^N)) \rightarrow \bar{u}(t, m)$ as $N \rightarrow \infty$. We also introduce for all $N \geq 1$, the functions $\phi^N(s, \mathbf{y}) := \varphi(s, m^N(\mathbf{y}))$ for all $(s, \mathbf{y}) \in [0, T] \times \mathbf{S}^N$.

Step 1: We prove that, without loss of generality, $\{\mathbf{y}^N = (\mathbf{x}^N, \mathbf{i}^N)\}_{N \geq 1}$ may be taken s.t.

$$u^N(t^N, \mathbf{y}^N) - \max_{\mathbf{i}' < \mathbf{i}^N} u^N(t^N, \mathbf{x}^N, \mathbf{i}') > 0 \quad \text{for all } N \geq 1. \quad (\text{V.5.2})$$

Indeed, either we may extract from $\{\mathbf{y}^N\}_{N \geq 1}$ a subsequence satisfying this property, or we may extract a subsequence (still denoted the same) s.t.

$$u^N(t^N, \mathbf{y}^N) - \max_{\mathbf{i}' < \mathbf{i}^N} u^N(t^N, \mathbf{x}^N, \mathbf{i}') \leq 0 \quad \text{for all } N \geq 1. \quad (\text{V.5.3})$$

Then $\mathbf{i}^N \neq 0$, as otherwise the maximum is equal to $-\infty$, and thus (V.5.2) is trivially satisfied, and there exists $\mathbf{y}^{N,*} := (\mathbf{x}^N, \mathbf{i}^{N,*})$ which attains the above maximum, and s.t.

$$u^N(t^N, \mathbf{y}^{N,*}) - \max_{\mathbf{i}' < \mathbf{i}^{N,*}} u^N(t^N, \mathbf{x}^N, \mathbf{i}') > 0 \quad \text{for all } N \geq 1,$$

that is, the property (V.5.2) is satisfied. As $m^N(\mathbf{y}^N)$ converges, the first marginal of $m^N(\mathbf{y}^{N,*})$ converges and, since its second marginal is a measure on $\{0, 1\}$, the sequence $\{m^N(\mathbf{y}^{N,*})\}_{N \geq 1}$ is tight, and we may thus extract a subsequence (still denoted the same) converging to some $m^* \preceq m$. Then, taking the \liminf as $N \rightarrow \infty$ in (V.5.3), we obtain

$$\bar{u}(t, m) - \bar{u}(t, m^*) \leq 0, \text{ i.e. } m^* \in C_{\bar{u}}(t, m),$$

which implies that $m^* = m$ by (V.5.1). Observing that

$$\bar{u}(t, m) = \lim_{N \rightarrow \infty} u^N(t^N, m^N(\mathbf{y}^N)) \leq \liminf_{N \rightarrow \infty} u^N(t^N, m(\mathbf{y}^{N,*})) \leq \limsup_{N \rightarrow \infty} u^N(t^N, m(\mathbf{y}^{N,*})) \leq \bar{u}(t, m),$$

we deduce that $(t^N, m^N(\mathbf{y}^{N,*}), u^N(t^N, m^N(\mathbf{y}^{N,*}))) \xrightarrow{N \rightarrow \infty} (t, m, \bar{u}(t, m))$, and thus we may replace $\{\mathbf{y}^N\}_{N \geq 1}$ with $\{\mathbf{y}^{N,*}\}_{N \geq 1}$, which satisfies (V.5.2).

Step 2: Fix $\delta \in (0, \delta_0)$. By standard optimal stopping theory, there exists $\theta_\delta^N \in \mathcal{T}_{t^N, T}$ s.t.

$$\mathbb{E}_{t^N, \mathbf{y}^N}^{\mathbb{P}^0} \left[(\phi^N - u^N)(\theta \wedge H_\delta^N, \bar{\mathbf{Y}}_{\theta_\delta^N \wedge H_\delta^N}) \right] = \inf_{\theta \in \mathcal{T}_{t^N, T}} \mathbb{E}_{t^N, \mathbf{y}^N}^{\mathbb{P}^0} \left[(\phi^N - u^N)(\theta \wedge H_\delta^N, \bar{\mathbf{Y}}_{\theta \wedge H_\delta^N}) \right],$$

where $\bar{\mathbf{Y}} := (\bar{X}, \mathbf{i}^N)$ satisfies the unstopped version of (V.2.8), and

$$H_\delta^N := \inf \left\{ s \geq t^N : \mathcal{W}_2(m^N(\bar{\mathbf{Y}}_s), m^N(\mathbf{y}^N)) \geq \delta \right\} \wedge (t^N + \delta).$$

Note that, since $\{u^N\}_{N \geq 1}$ is locally bounded, uniformly in N , we may assume w.l.o.g. that N is large enough and δ small enough so that $\{(\phi^N - u^N)(\theta_\delta^N \wedge H_\delta^N, \bar{\mathbf{Y}}_{\theta_\delta^N \wedge H_\delta^N})\}_{N \geq 1}$ is uniformly bounded.

Step 3: We now justify that

$$\limsup_{N \rightarrow \infty} \mathbb{P}^0 \left(\theta_\delta^N < H_\delta^N \right) > 0. \tag{V.5.4}$$

Indeed, assume to the contrary that $\limsup_{N \rightarrow \infty} \mathbb{P}^0(\theta_\delta^N < H_\delta^N) = 0$. We have:

$$\begin{aligned}
(\varphi - \bar{u})(t, m) &= \lim_{N \rightarrow \infty} (\varphi - u^N)(t^N, m^N(\mathbf{y}^N)) \\
&\geq \limsup_{N \rightarrow \infty} \mathbb{E}_{t^N, \mathbf{y}^N}^{\mathbb{P}^0} \left[(\varphi - u^N)(\theta_\delta^N \wedge H_\delta^N, m^N(\bar{\mathbf{Y}}_{\theta_\delta^N \wedge H_\delta^N})) \right] \\
&\geq \liminf_{N \rightarrow \infty} \mathbb{E}_{t^N, \mathbf{y}^N}^{\mathbb{P}^0} \left[(\varphi - u^N)(\theta_\delta^N \wedge H_\delta^N, m^N(\bar{\mathbf{Y}}_{\theta_\delta^N \wedge H_\delta^N})) \right] \\
&= \liminf_{N \rightarrow \infty} \mathbb{E}_{t^N, \mathbf{y}^N}^{\mathbb{P}^0} \left[\left\{ (\varphi - u^N)(H_\delta^N, m^N(\bar{\mathbf{Y}}_{H_\delta^N})) (1 - \mathbf{1}_{\theta_\delta^N < H_\delta^N}) \right. \right. \\
&\quad \left. \left. + (\varphi - u^N)(\theta_\delta^N, m^N(\bar{\mathbf{Y}}_{\theta_\delta^N})) \mathbf{1}_{\theta_\delta^N < H_\delta^N} \right\} \right] \quad (\text{V.5.5}) \\
&= \liminf_{N \rightarrow \infty} \mathbb{E}_{t^N, \mathbf{y}^N}^{\mathbb{P}^0} \left[(\varphi - u^N)(H_\delta^N, m^N(\bar{\mathbf{Y}}_{H_\delta^N})) \right],
\end{aligned}$$

where we use the uniform boundedness of $\{(\phi^N - u^N)(\theta \wedge H_\delta^N, \bar{\mathbf{Y}}_{\theta \wedge H_\delta^N})\}_{N \geq 1}$ and the fact that $\limsup_{N \rightarrow \infty} \mathbb{P}^0(\theta_\delta^N < H_\delta^N) = \lim_{N \rightarrow \infty} \mathbb{P}^0(\theta_\delta^N < H_\delta^N) = 0$.

Let $\bar{\mathbf{Y}}^{t^N, \mathbf{y}^N}$ be the version of $\bar{\mathbf{Y}}$ starting from $\bar{\mathbf{Y}}_{t^N, \mathbf{y}^N}^{t^N, \mathbf{y}^N} = \mathbf{y}^N$. Since $(t^N, m^N(\mathbf{y}^N)) \xrightarrow{N \rightarrow \infty} (t, m)$ and $\bar{\mathbf{Y}}$ is a unstopped diffusion whose coefficients satisfy the usual Lipschitz conditions, by classical propagation of chaos (see Oelschläger [63]) and compactness of $[t, T]$, there exists a subsequence $\nu^N := \{\mathbb{P}^0 \circ (H_\delta^N, m^N(\bar{\mathbf{Y}}^{t^N, \mathbf{y}^N}))^{-1}\}_{N \geq 1}$ that converges weakly to some element of $\nu \in \mathcal{P}_2([t, T] \times \mathcal{P}_2(\Omega))$, under which Y is an unstopped diffusion starting from the distribution m . Thus, denoting by (τ, λ) the canonical mapping on $[0, T] \times \mathcal{P}_2(\Omega)$, we have by continuity of $\varphi - u^N$ and $(\tau, \lambda) \mapsto \lambda_{Y_\tau}$ (since Y is unstopped under λ , ν -a.s),

$$\begin{aligned}
\liminf_{N \rightarrow \infty} \mathbb{E}_{t^N, \mathbf{y}^N}^{\mathbb{P}^0} \left[(\varphi - u^N)(H_\delta^N, m^N(\bar{\mathbf{Y}}_{H_\delta^N})) \right] &= \liminf_{N \rightarrow \infty} \mathbb{E}^{\nu^N} \left[(\varphi - u^N)(\tau, \lambda_{Y_\tau}) \right] \\
&\geq \mathbb{E}^\nu \left[(\varphi - \bar{u})(\tau, \lambda_{Y_\tau}) \right] \\
&\geq (\varphi - \bar{u})(\tau(\bar{\omega}), \lambda_{Y_{\tau(\bar{\omega})}}(\bar{\omega})),
\end{aligned}$$

for some $\bar{\omega} \in [0, T] \times \mathcal{P}_2(\Omega)$. Yet, given that, by definition of H_δ^N , we have

$$\mathcal{W}_2(\lambda_{Y_\tau}, m^N(\mathbf{y}^N)) \geq \delta \text{ or } \tau = t^N + \delta, \quad \nu^N\text{-a.s.},$$

for all $N \geq 1$, we also have,

$$\mathcal{W}_2(\lambda_{Y_\tau}, m) \geq \delta \text{ or } \tau = t + \delta, \quad \nu\text{-a.s.}$$

$\bar{\omega}$ may then be chosen s.t. $(\tau(\bar{\omega}), \lambda_{Y_{\tau(\bar{\omega})}}(\bar{\omega})) \neq (t, m)$. Since Y is an unstopped diffusion under λ , ν -a.s., with $\lambda_{Y_{t-}} = m$, we deduce that $(\tau(\bar{\omega}), \lambda_{Y_{\tau(\bar{\omega})}}(\bar{\omega})) \in \mathcal{N}_\delta(t, m)$. Since we

have $(\varphi - \bar{u})(t, m) \geq (\varphi - \bar{u})(\tau(\bar{\omega}), \lambda_{Y_{\tau(\bar{\omega})}}(\bar{\omega}))$, this contradicts the fact that (t, m) is a strict minimum on $\mathcal{N}_\delta(t, m)$. Thus, (V.5.4) holds true.

Step 4: We prove that $\phi^N \in C^{1,2}(\Lambda_{t^N})$ (see §V.7.4). More precisely, we extend the results of Carmona & Delarue [21, Vol. 1, Propositions 5.35 & 5.91] to the present setting by proving that, for all $(s, \mathbf{y}) \in [0, T] \times \mathbf{S}^N$ and $k \in [N]$ s.t. $\mathbf{i}_k = 1$,

$$\begin{aligned} \partial_{x_k} \phi^N(s, \mathbf{y}) &= \frac{1}{N} \partial_x \delta_m \varphi_1(s, m^N(\mathbf{y}), x_k), \\ \partial_{x_k x_k}^2 \phi^N(s, \mathbf{y}) &= \frac{1}{N} \partial_{xx}^2 \delta_m \varphi_1(s, m^N(\mathbf{y}), x_k) + \frac{1}{N^2} \partial_{\mu\mu}^2 \varphi_{1,1}(s, m^N(\mathbf{y}), x_k, x_k). \end{aligned} \quad (\text{V.5.6})$$

Let $h > 0$ and $(\mathbf{e}_1, \dots, \mathbf{e}_N)$ be the canonical basis of $\mathbb{R}^{d \times N}$, and k be s.t. $i_k = 1$. Denoting $T_h^k \phi^N(t, \mathbf{y}) := \frac{1}{h} (\phi^N(s, \mathbf{y} + h\mathbf{e}_k) - \phi^N(s, \mathbf{y}))$, with $\mathbf{y} + h\mathbf{e}_k = (\mathbf{x} + h\mathbf{e}_k, \mathbf{i})$, and $m_h^{\lambda,k} := \lambda m^N(\mathbf{y} + h\mathbf{e}_k) + (1 - \lambda)m^N(\mathbf{y})$, we have

$$\begin{aligned} T_h^k \phi^N(s, \mathbf{y}) &= \frac{1}{h} (\varphi(s, m^N(\mathbf{y} + h\mathbf{e}_k)) - \varphi(s, m^N(\mathbf{y}))) \\ &= \frac{1}{h} \int_0^1 \int_{\mathbf{S}} \delta_m \varphi(s, m_h^{\lambda,k}, y) d(m^N(\mathbf{y} + h\mathbf{e}_k) - m^N(\mathbf{y}))(y) d\lambda \\ &= \frac{1}{N} \int_0^1 \frac{1}{h} [\delta_m \varphi(s, m_h^{\lambda,k}, y_k + h) - \delta_m \varphi(s, m_h^{\lambda,k}, y_k)] d\lambda. \end{aligned}$$

As $\varphi \in C_2^{1,1}(\mathbf{Q}_0)$ and $i_k = 1$, the integrand converges as $h \rightarrow 0$, and therefore is bounded for $(h, \lambda) \in [0, h_0] \times [0, 1]$ for some $h_0 > 0$. We may then apply the dominated convergence theorem. As $m_h^{\lambda,k} \xrightarrow{h \rightarrow 0} m^N(\mathbf{y})$, this leads to the first equality. We now compute

$$\begin{aligned} T_h^k \partial_{x_k} \phi(s, \mathbf{y}) &= \frac{1}{hN} \left(\partial_x \delta_m \varphi_1(s, m^N(\mathbf{y} + h\mathbf{e}_k), x_k + h) - \partial_x \delta_m \varphi_1(s, m^N(\mathbf{y}), x_k) \right) \\ &= \frac{1}{hN} \left(\partial_x \delta_m \varphi_1(t, m^N(\mathbf{y} + h\mathbf{e}_k), x_k + h) - \partial_x \delta_m \varphi_1(s, m^N(\mathbf{y} + h\mathbf{e}_k), x_k) \right) \\ &\quad + \frac{1}{hN} \left(\partial_x \delta_m \varphi_1(s, m^N(\mathbf{y} + h\mathbf{e}_k), x_k) - \partial_x \delta_m \varphi_1(s, m^N(\mathbf{y}), x_k) \right) \end{aligned}$$

It is clear that the first term of the right hand side tends to $\frac{1}{N} \partial_{xx}^2 \delta_m \varphi_1(s, m^N(\mathbf{y}), x_k)$ as $h \rightarrow 0$. The second term is equal to

$$\begin{aligned} &\frac{1}{hN} \int_0^1 \int_{\mathbf{S}} \delta_m \partial_\omega \delta_m \varphi_1(s, m_h^{\lambda,k}, x_k, y) d(m^N(\mathbf{y} + h\mathbf{e}_k) - m^N(\mathbf{y}))(y) d\lambda \\ &= \frac{1}{N^2} \int_0^1 \frac{1}{h} [\delta_m \partial_\omega \delta_m \varphi_{1,1}(s, m_h^{\lambda,k}, x_k, x_k + h) - \delta_m \partial_x \delta_m \varphi_{1,1}(s, m_h^{\lambda,k}, x_k, x_k)] d\lambda. \end{aligned}$$

Similarly to the case of the first order derivative, we may apply the dominated convergence theorem to deduce that it converges to

$$\frac{1}{N^2} \partial_x \delta_m \partial_x \delta_m \varphi_{1,1}(s, m^N(\mathbf{y}), x_k, x_k) = \partial_{\mu\mu}^2 \varphi_{1,1}(s, m^N(\mathbf{y}), x_k, x_k).$$

Step 5: We prove that ϕ^N is a test function of u^N in some well chosen point. By (V.5.4), for all $\delta > 0$ and $N \geq 1$, we may find $\omega^{\delta,N} \in \Omega^0$ s.t., after possibly passing to a subsequence, $\theta_\delta^N(\omega^{\delta,N}) < H_\delta^N(\omega^{\delta,N})$. Then, introducing $(t_\delta^N, \mathbf{y}_\delta^N) := (\theta_\delta^N(\omega^{\delta,N}), \bar{\mathbf{Y}}_{\theta_\delta^N(\omega^{\delta,N})}^{t_\delta^N, \mathbf{y}_\delta^N}(\omega^{\delta,N}))$ and the process \mathcal{Z}^N defined by

$$\mathcal{Z}_s^N := \operatorname{ess\,inf}_{\theta \in \mathcal{T}_{t_\delta^N, T}^N} \mathbb{E}^{\mathbb{P}^0} [(\phi^N - u^N)(s, \bar{\mathbf{Y}}_s^{t_\delta^N, \mathbf{y}_\delta^N}) | \mathcal{F}_s^N] \text{ for all } s \in [t_\delta^N, T],$$

i.e., the \mathbb{P}^0 -lower Snell envelope of $s \mapsto (\phi^N - u^N)(s, \bar{\mathbf{Y}}_s^{t_\delta^N, \mathbf{y}_\delta^N})$, which is a \mathbb{P}^0 -submartingale, we have for all $\theta \in \mathcal{T}_{t_\delta^N, T}^N$

$$(\phi^N - u^N)(t_\delta^N, \mathbf{y}_\delta^N) = \mathcal{Z}_{t_\delta^N}^N(\omega^{\delta,N}) \leq \mathbb{E}^{\mathbb{P}^0} [\mathcal{Z}_{\theta \wedge H_\delta^N}^N] \leq \mathbb{E}^{\mathbb{P}^0} [(\phi^N - u^N)(\theta \wedge H_\delta^N, \bar{\mathbf{Y}}_{\theta \wedge H_\delta^N}^{t_\delta^N, \mathbf{y}_\delta^N})],$$

and thus we have

$$(\phi^N - u^N)(t_\delta^N, \mathbf{y}_\delta^N) = \min_{\theta \in \mathcal{T}_{t_\delta^N, T}^N} \mathbb{E}_{t_\delta^N, \mathbf{y}_\delta^N}^{\mathbb{P}^0} [(\phi^N - u^N)(\theta \wedge H_\delta^N, \bar{\mathbf{Y}}_{\theta \wedge H_\delta^N}^{t_\delta^N, \mathbf{y}_\delta^N})]. \quad (\text{V.5.7})$$

Therefore, since $\phi^N \in C^{1,2}(\Lambda_t^N)$ by the previous Step, we have $\phi^N \in \underline{\mathcal{A}}^N u^N(t_\delta^N, \mathbf{y}_\delta^N)$ and the subsolution property provides,

$$\min \left\{ -\partial_t \phi^N(t_\delta^N, \mathbf{y}_\delta^N) - (\mathcal{L}\phi^N + \mathbf{f}^N)(t_\delta^N, \mathbf{y}_\delta^N) \cdot \mathbf{i}^N, \right. \\ \left. u^N(t_\delta^N, \mathbf{y}_\delta^N) - \max_{\mathbf{i}' < \mathbf{i}^N} u^N(t_\delta^N, \mathbf{x}_\delta^N, \mathbf{i}') \right\} \leq 0. \quad (\text{V.5.8})$$

Step 6: We finally derive the subsolution property of \bar{u} . First observe that, since $t_\delta^N = \theta_\delta^N(\omega^{\delta,N}) < H_\delta^N(\omega^{\delta,N})$, we have $(t_\delta^N, \mathbf{y}_\delta^N) \xrightarrow{\delta \rightarrow 0} (t^N, \mathbf{y}^N)$. Since ϕ^N , $\mathcal{L}\phi^N + \mathbf{f}^N$ and u^N are continuous, sending δ to 0 in (V.5.8) provides

$$\min \left\{ -\partial_t \phi^N(t^N, \mathbf{y}^N) - (\mathcal{L}\phi^N + \mathbf{f}^N)(t^N, \mathbf{y}^N) \cdot \mathbf{i}^N, u^N(t^N, \mathbf{y}^N) - \max_{\mathbf{i}' < \mathbf{i}^N} u^N(t^N, \mathbf{x}^N, \mathbf{i}') \right\} \leq 0.$$

Since we assumed w.l.o.g. that $u^N(t^N, \mathbf{y}^N) - \max_{\mathbf{i}' < \mathbf{i}^N} u^N(t^N, \mathbf{x}^N, \mathbf{i}') > 0$, this implies

$$-\partial_t \phi^N(t^N, \mathbf{y}^N) - (\mathcal{L}\phi^N + \mathbf{f}^N)(t^N, \mathbf{y}^N) \cdot \mathbf{i}^N \leq 0 \text{ for all } N \geq 1. \quad (\text{V.5.9})$$

Thus, by (V.5.6),

$$\begin{aligned} -\partial_t \phi^N(t^N, \mathbf{y}^N) - (\mathcal{L}\phi^N + \mathbf{f}^N)(t^N, \mathbf{y}^N) \cdot \mathbf{i}^N \\ = -\mathbb{L}\varphi(t^N, m^N(\mathbf{y}^N)) - \frac{1}{N^2} \sum_{k:i_k^N=1} \sigma \sigma^\top : \partial_{\mu\mu}^2 \varphi_{1,1}(t^N, m^N(\mathbf{y}^N), x_k^N, x_k^N) \end{aligned}$$

where

$$\sigma \sigma^\top : \partial_{\mu\mu}^2 \varphi_{1,1}(t^N, m^N(\mathbf{y}^N), x_k^N, x_k^N) := \sigma \sigma^\top(t^N, m^N(\mathbf{y}^N), x_k^N) : \partial_{\mu\mu}^2 \varphi_{1,1}(t^N, m^N(\mathbf{y}^N), x_k^N, x_k^N).$$

By continuity of $(s, \tilde{m}) \mapsto \int \sigma \sigma^\top : \partial_{\mu\mu}^2 \varphi_{1,1}(s, \tilde{m}, x, x) \tilde{m}(dx, 1)$, we have

$$\frac{1}{N} \sum_{k=1}^N \sigma \sigma^\top : \partial_{\mu\mu}^2 \varphi_{1,1}(t^N, m^N(\mathbf{y}^N), x_k^N, x_k^N) i_k^N \xrightarrow{N \rightarrow \infty} \int_{\mathbb{R}^d} \sigma \sigma^\top : \partial_{\mu\mu}^2 \varphi_{1,1}(t, m, x, x) m(dx, 1),$$

and thus

$$\frac{1}{N} \sum_{k=1}^N \sigma \sigma^\top : \partial_{\mu\mu}^2 \varphi_{1,1}(t^N, m^N(\mathbf{y}^N), x_k^N, x_k^N) i_k^N = \mathcal{O}\left(\frac{1}{N}\right).$$

Therefore, (V.5.9) becomes

$$-(\mathbb{L}\varphi + F)(t^N, m^N(\mathbf{y}^N)) + \mathcal{O}\left(\frac{1}{N}\right) \leq 0 \quad \text{for all } N \geq 1.$$

Since $(t^N, m^N(\mathbf{y}^N)) \xrightarrow{N \rightarrow \infty} (t, m)$ and $\mathbb{L}\varphi + F$ is continuous, we conclude by sending $N \rightarrow \infty$.

(ii) We now prove that \underline{v} is a viscosity supersolution of (V.4.2). Fix $\varphi \in \overline{\mathcal{A}}\underline{v}(t, m)$ with corresponding δ and s.t. (t, m) is a strict maximizer of $\varphi - \underline{v}$ on $\mathcal{N}_\delta(t, m)$. With the same procedure and notations as Steps 2 to 5 of (i) (in particular, we rewrite (V.5.5) with reversed inequalities and switch limsup and liminf), we may find $(t_\delta^N, m^N(\mathbf{y}_\delta^N)) \xrightarrow{(\delta, N) \rightarrow (0, \infty)} (t, m)$ s.t.

$$(\phi^N - v^N)(t_\delta^N, \mathbf{y}_\delta^N) = \max_{\theta \in \mathcal{T}_{t_\delta^N, T}^N} \mathbb{E}_{t_\delta^N, \mathbf{y}_\delta^N}^{\mathbb{P}^0} \left[(\phi^N - v^N)(\theta \wedge H_\delta^N, \bar{\mathbf{Y}}_{\theta \wedge H_\delta^N}) \right],$$

Thus $\phi^N \in \overline{\mathcal{A}}^N v^N(t_\delta^N, \mathbf{y}_\delta^N)$ and the viscosity supersolution property provides

$$\begin{aligned} \min \left\{ -(\mathbb{L}\varphi + F)(t_\delta^N, m^N(\mathbf{y}_\delta^N)) + \mathcal{O}\left(\frac{1}{N}\right), \right. \\ \left. v^N(t_\delta^N, m^N(\mathbf{y}_\delta^N)) - \max_{\mathbf{i}' < \mathbf{i}_\delta^N} v^N(t_\delta^N, m^N(\mathbf{x}_\delta^N, \mathbf{i}')) \right\} \geq 0. \end{aligned} \quad (\text{V.5.10})$$

As $(t_\delta^N, m^N(\mathbf{y}_\delta^N)) \xrightarrow{(\delta, N) \rightarrow (0, \infty)} (t, m)$, sending $(\delta, N) \rightarrow (0, \infty)$ in (V.5.10) provides the first part of the viscosity supersolution property:

$$-(\mathbb{L}\varphi + F)(t, m) \geq 0.$$

We now prove that \underline{v} is nondecreasing for \preceq . Let $A \in \mathcal{B}(\mathbb{R}^d)$ and assume that

$$m(dx, 0) \text{ and } m(dx, 1) \text{ have continuous densities.} \quad (\text{V.5.11})$$

By considering the corresponding cumulative density functions, we easily see that

$$(m^N)^A(\mathbf{y}^N) := \frac{1}{N} \sum_{j=1}^N \delta_{(x_k^N, i_k^N \mathbf{1}_A(x_k^N))} \xrightarrow{N \rightarrow \infty} m^A.$$

By (V.5.10), we have $v^N(t^N, m^N(\mathbf{y}^N)) \geq v^N(t^N, (m^N)^A(\mathbf{y}^N))$ for all $N \geq 1$. Taking the liminf and using the fact that $v^N(t^N, m^N(\mathbf{y}^N)) \xrightarrow{N \rightarrow \infty} \underline{v}(t, m)$, we deduce $\underline{v}(t, m) \geq \underline{v}(t, m^A)$. Given the arbitrariness of A and Lemma V.8.1, we deduce that $\underline{v}(t, m) \geq \underline{v}(t, m')$ for all $m' \preceq m$ satisfying (V.5.11). By (V.4.1), density of measures with continuous density in $\mathcal{P}_2(\mathbf{S})$ and continuity of \underline{v} , we easily extend this inequality to all $m' \preceq m$ with continuous transition functions, and finally by [75, Lemma 3.8] to all $m' \preceq m$. \blacksquare

Proof of Theorem V.3.2 We first prove that the V^N are locally bounded, uniformly in N . Fix $(t, m) \in \mathbf{Q}_0$ and $(s, \mathbf{y}^N) \in [0, T] \times \mathbf{S}^N$, with $m^N(\mathbf{y}^N) \xrightarrow{\mathcal{W}_2} m$. Observing that

$$\begin{aligned} \mathbb{E}_{s, \mathbf{x}}^{\mathbb{P}^0} \left[|g(m^N(\mathbf{X}_T)) - g(m^N(\mathbf{x}^N))|^2 \right] &\leq L \mathbb{E}_{s, \mathbf{x}^N}^{\mathbb{P}^0} \left[\mathcal{W}_2^2(m^N(\mathbf{X}_T), m^N(\mathbf{x}^N)) \right] \\ &\leq \frac{L}{N} \sum_{k=1}^N \mathbb{E}_{s, \mathbf{x}}^{\mathbb{P}^0} \left[|X_T^k - x_k^N|^2 \right] \\ &\leq C(T - s) \end{aligned}$$

where, by Lemma V.9.1, C is a constant independent from N , and L is a Lipschitz constant. We obtain similar estimates with F and then we easily see that

$$\left| V^N(t, \mathbf{y}^N) - \int_s^T F(r, m^N(\mathbf{y}^N)) dr - g(m^N(\mathbf{x}^N)) \right| \leq C' \sqrt{T - s} \quad (\text{V.5.12})$$

for some constant C' independent from N . Since f and g are Lipschitz continuous w.r.t. \mathcal{W}_2 , we deduce that

$$|V^N(s, \mathbf{y}^N)| \leq C_T (\mathcal{W}_2(m^N(\mathbf{y}^N), m) + F(t, m) + g(m)),$$

where the constant C_T depends on T only. This proves that the V^N are locally bounded, uniformly in N . By symmetry of the dynamics of \mathbf{Y} and of f and g , we also clearly have $V^N \in \mathcal{S}^N$, and thus

$$\underline{V}(t, m) := \liminf_{\substack{N \rightarrow \infty, s \rightarrow t \\ m^N(\mathbf{y}^N) \xrightarrow{W_2} m}} V^N(s, \mathbf{y}^N), \quad \text{and} \quad \bar{V}(t, m) := \limsup_{\substack{N \rightarrow \infty, s \rightarrow t \\ m^N(\mathbf{y}^N) \xrightarrow{W_2} m}} V^N(s, \mathbf{y}^N). \quad (\text{V.5.13})$$

are well-defined. The Lipschitz continuity of f and g provides the continuity of V^N (see Theorem V.4.4 and \underline{V}). Then, by Theorem V.5.1, \underline{V} and \bar{V} are respectively viscosity supersolution and subsolution of (V.4.2). Moreover, (V.5.12) implies that $\underline{V}(T, \cdot) = \bar{V}(T, \cdot) = g$. We may thus apply our comparison principle [75, Theorem 3.11] to deduce that $\underline{V} \geq \bar{V}$. By (V.5.13), we also have the converse inequality, and thus $\underline{V} = \bar{V}$, and these functions are viscosity solutions of (V.4.2). By uniqueness, they are equal to V and the two semi-limits (V.5.13) are equal to the limit. \blacksquare

V.6 Propagation of chaos for stopped McKean-Vlasov diffusions

The following results extend the classical propagation of chaos to the setting of stopped diffusions:

Proposition V.6.1 *Let $\mathbf{y}^N \in \mathbf{S}^N$, $\tau^N \in (\mathcal{T}_{t,T}^N)^N$, $N \geq 1$, $m \in \mathcal{P}_2(\mathbf{S})$ such that $m^N(\mathbf{y}^N) \xrightarrow{W_2} m$, and \mathbf{Y}^{τ^N} be the \mathbf{S}^N -valued process defined by (V.2.8) started from $\mathbf{Y}_t^{\tau^N} = \mathbf{y}^N$. The sequence $\{\mathbb{P}^0 \circ (m^N(\mathbf{Y}^{\tau^N}))^{-1}\}_{N \geq 1}$ admits a converging subsequence whose limit is supported on $\mathcal{P}(t, m)$.*

Proof *Step 1:* We first prove the existence of a limit. For all $N \geq 1$, denote $\nu^N := \mathbb{P}^0 \circ (m^N(\mathbf{Y}^{\tau^N}))^{-1} \in \mathcal{P}_2(\mathcal{P}_2(\Omega))$. By Lacker [51, Corollary B.1], we have to prove that (i) $\{\nu^N\}_{N \geq 1}$ is uniformly integrable, i.e., $\lim_{R \rightarrow \infty} \sup_{N \geq 1} \mathbb{E}^{\nu^N} \left[\mathcal{W}_2^2(\lambda, \delta_0) \mathbf{1}_{\mathcal{W}_2(\lambda, \delta_0) \geq R} \right] = 0$, where λ is the identity map on $\mathcal{P}_2(\Omega)$, (ii) the sequence of mean measures $\{\mathbb{E}^{\mathbb{P}^0} [m^N(\mathbf{Y}^{\tau^N})]\}_{N \geq 1}$ is tight, where, for all $\tilde{m} : \Omega^0 \rightarrow \mathcal{P}_2(\Omega)$, the mean measure $\mathbb{E}^{\mathbb{P}^0} [\tilde{m}] \in \mathcal{P}_2(\Omega)$ is defined by

$$\langle \mathbb{E}^{\mathbb{P}^0} [\tilde{m}], \varphi \rangle := \mathbb{E}^{\mathbb{P}^0} [\langle \tilde{m}, \varphi \rangle] \quad \text{for all } \varphi \in C_b^0(\Omega).$$

Let $R > 0$. We have

$$\begin{aligned}
\mathbb{E}^{\nu^N} \left[\mathcal{W}_2(\lambda, \delta_0) \mathbf{1}_{\mathcal{W}_2(\lambda, \delta_0) \geq R} \right] &= \mathbb{E}^{\mathbb{P}^0} \left[\mathcal{W}_2(m^N(\mathbf{Y}^{\tau^N}), \delta_0) \mathbf{1}_{\mathcal{W}_2(m^N(\mathbf{Y}^{\tau^N}), \delta_0) \geq R} \right] \\
&\leq \frac{1}{R} \mathbb{E}^{\mathbb{P}^0} \left[\mathcal{W}_2^2(m^N(\mathbf{Y}^{\tau^N}), \delta_0) \right] \\
&\leq \frac{1}{R} \mathbb{E}^{\mathbb{P}^0} \left[\frac{1}{N} \sum_{i=1}^N |\mathbf{Y}^{i, \tau^N}|_\infty^2 \right] \\
&\leq \frac{2}{R} \left\{ \frac{1}{N} \sum_{i=1}^N |y_i^N|^2 + \mathbb{E}^{\mathbb{P}^0} \left[\frac{1}{N} \sum_{i=1}^N |\mathbf{Y}^{i, \tau^N} - y_i^N|_\infty^2 \right] \right\}.
\end{aligned}$$

Since $\mathcal{W}_2(m^N(\mathbf{y}^N), m) \xrightarrow{N \rightarrow \infty} 0$, $\frac{1}{N} \sum_{i=1}^N |y_i^N|^2$ converges as $N \rightarrow \infty$, and thus is bounded. By Lemma V.9.1, we have $\mathbb{E}^{\mathbb{P}^0} \left[\frac{1}{N} \sum_{i=1}^N |X^{i, \tau^N} - x_i^N|_\infty^2 \right] \leq C_T$, where C_T is independent from N . Finally, there exists a constant \tilde{C}_T independent from N and R s.t.

$$\mathbb{E}^{\nu^N} \left[\mathcal{W}_2(\lambda, \delta_0) \mathbf{1}_{\mathcal{W}_2(\lambda, \delta_0) \geq R} \right] \leq \frac{\tilde{C}_{t,L}}{R} \quad \text{for all } N \geq 1 \text{ and } R \geq 0,$$

and thus $\lim_{R \rightarrow \infty} \sup_{N \geq 1} \mathbb{E}^{\nu^N} \left[\mathcal{W}_2(\lambda, \delta_0) \mathbf{1}_{\mathcal{W}_2(\lambda, \delta_0) \geq R} \right] = 0$ and (i) is proved.

To show that $\{\mathbb{E}^{\mathbb{P}^0} [m^N(\mathbf{Y}^{\tau^N})]\}_{N \geq 1}$ is tight, we show the sequences corresponding to the two marginals are tight. For the first marginal, we prove Aldous' criterion (see Billingsley [10, Theorem 16.10]), i.e.,

$$\sup_{N \geq 1} \sup_{\tau \in \mathcal{T}_{0,T}} \left\langle \mathbb{E}^{\mathbb{P}^0} [m^N(\mathbf{Y}^{\tau^N})], |X_{(\tau+\delta) \wedge T} - X_\tau|^2 \right\rangle \xrightarrow{\delta \rightarrow 0} 0. \quad (\text{V.6.1})$$

Yet, by standard estimates for SDEs, we have, for fixed N, τ and δ ,

$$\left\langle \mathbb{E}^{\mathbb{P}^0} [m^N(\mathbf{Y}^{\tau^N})], |X_{(\tau+\delta) \wedge T} - X_\tau|^2 \right\rangle = \frac{1}{N} \sum_{i=1}^N \mathbb{E}^{\mathbb{P}^0} \left[|X_{(\tau+\delta) \wedge T}^{i, \tau^N} - X_\tau^{i, \tau^N}|^2 \right] \leq C\delta$$

for some $C > 0$ independent from δ and N (see Lemma V.9.1). This implies (V.6.1). The second marginals are measures on $\mathbb{I}^0([-1, T])$, and are in continuous bijection with measures on $[0, T]$, and thus are tight. We consequently obtain (ii), and thus $\{\nu^N\}_{N \geq 1}$ admits a subsequence converging to some $\nu \in \mathcal{P}_2(\mathcal{P}_2(\Omega))$.

Step 2: We now prove that ν is supported on $\mathcal{P}(t, m)$, or equivalently, that $\lambda \in \mathcal{P}(t, m)$, ν -a.s. As $\mathcal{W}_2(m^N(\mathbf{y}^N), m) \xrightarrow{N \rightarrow \infty} 0$, it follows that $\lambda_{t-} = m$, ν -a.s. By compactness of the laws of bounded stopping times, the second marginal of λ is the distribution of a stopping

time. It remains to prove that the first marginal of λ is the law of a stopped McKean-Vlasov diffusion (V.2.5). For all $\psi \in C_b^2(\mathbb{R}^d)$, introduce

$$M_s^{\psi,\lambda}(Y) := \psi(X_s) - \int_t^s \mathcal{L}_x \psi(r, X_r, \lambda_r) I_r dr = \psi(X_t) + \int_t^s \sigma \partial_x \psi(r, X_r, \lambda_r) dW_r,$$

Fix $h_t \in C_b^0(\Omega)$ that is \mathcal{F}_t -measurable. We shall prove that

$$\left\langle \lambda, h_t (M_s^{\psi,\lambda} - M_t^{\psi,\lambda})(Y) \right\rangle = 0, \quad \nu\text{-a.s.}, \quad (\text{V.6.2})$$

which means by the arbitrariness of s and h_t that $M^{\psi,\lambda}(X)$ is a λ -martingale, ν -a.s., and thus, by arbitrariness of ψ , that λ is supported on $\mathcal{P}(t, m)$. Denoting $M_{t,s}^i := M_s^i - M_t^i$, we compute:

$$\begin{aligned} \mathbb{E}^{\mathbb{P}^0} \left[\left(\frac{1}{N} \sum_{k=1}^N h_t M_{t,s}^{\psi,\lambda}(Y^{k,\tau^N}) \right)^2 \right] &= \frac{1}{N^2} \sum_{k=1}^N \mathbb{E}^{\mathbb{P}^0} \left[h_t^2 |M_{t,s}^{\psi,\lambda}|^2(Y^{k,\tau^N}) \right] \\ &\leq \frac{1}{N^2} \sum_{k=1}^N \mathbb{E}^{\mathbb{P}^0} \left[\int_t^s h_t^2 |\sigma \partial_x \psi(X^{k,\tau^N})|^2 dr \right] \\ &\leq C \frac{Th_t^2 |\partial_x \psi|^2 \mathbb{E}^{\mathbb{P}^0} [1 + |X|_\infty^2]}{N}, \end{aligned}$$

for some constant C , where the equality is due to the mutual independence of the $\{W^k\}_{k \in [N]}$ and to the \mathcal{F}_t -measurability of h_t . Then

$$\mathbb{E}^{\nu^N} \left[\left\langle \lambda, h_t M_{t,s}^{\psi,\lambda}(Y) \right\rangle^2 \right] = \mathbb{E}^{\mathbb{P}^0} \left[\left(\frac{1}{N} \sum_{k=1}^N h_t M_{t,s}^{\psi,\lambda}(Y^{k,\tau^N}) \right)^2 \right] \xrightarrow{N \rightarrow 0} 0. \quad (\text{V.6.3})$$

On the other hand, as ν^N converges weakly to ν and $\mathbb{P} \in \mathcal{P}_2(\Omega) \mapsto \langle \mathbb{P}, h M_{t,s}^{\psi,\mathbb{P}}(Y) \rangle^2$ is continuous and bounded from below, we also have

$$0 \leq \mathbb{E}^\nu \left[\left\langle \lambda, h_t M_{t,s}^{\psi,\lambda}(Y) \right\rangle^2 \right] \leq \liminf_{N \rightarrow \infty} \mathbb{E}^{\nu^N} \left[\left\langle \lambda, h_t M_{t,s}^{\psi,\lambda}(Y) \right\rangle^2 \right],$$

which together with (V.6.3) proves (V.6.2), and thus we conclude that $\lambda \in \mathcal{P}(t, \mu)$, ν -a.s.

■

Proof of Theorem V.3.3 (i) Let $\hat{\nu}^N := \mathbb{P}^0 \circ (m^N(\mathbf{Y}^{\hat{\tau}^N}))^{-1}$. By Proposition V.6.1, we may extract a subsequence (still denoted the same) that converges weakly to some $\hat{\nu}$ supported

on $\mathcal{P}(0, m)$. Since $\hat{\tau}^N$ is ε_N -optimal for (V.7.2), we have

$$\begin{aligned} V_0^N &\leq \mathbb{E}^{\mathbb{P}^0} \left[\frac{1}{N} \sum_{k=1}^N \int_0^T f(r, X_r^{k, \hat{\tau}^N}, m^N(\mathbf{Y}_r^{\hat{\tau}^N})) I_r^k dr + g(m^N(\mathbf{Y}_T^{\hat{\tau}^N})) \right] + \varepsilon_N \\ &= \mathbb{E}^{\mathbb{P}^0} \left[\int_0^T F(r, m^N(\mathbf{Y}_r^{\hat{\tau}^N})) dr + g(m^N(\mathbf{Y}_T^{\hat{\tau}^N})) \right] + \varepsilon_N \\ &= \mathbb{E}^{\hat{\nu}^N} \left[\int_0^T F(r, \lambda_r) dr + g(\lambda_T) \right] + \varepsilon_N \end{aligned}$$

for all $N \geq 1$, where λ denotes the canonical map on $\mathcal{P}_2(\Omega)$. By Theorem V.3.2, we have $V_0^N \xrightarrow{N \rightarrow \infty} V(0, m)$. Since F and g are bounded, and $\varepsilon_N \searrow 0$, this implies

$$V(0, m) \leq \mathbb{E}^{\hat{\nu}} \left[\int_0^T F(r, \lambda_r) dr + g(\lambda_T) \right].$$

Since $\hat{\nu}$ is supported on $\mathcal{P}(0, m)$, we have

$$V(0, m) \geq \int_0^T F(r, \lambda_r) dr + g(\lambda_T), \quad \hat{\nu}\text{-a.s.},$$

and therefore

$$V(0, m) = \int_0^T F(r, \lambda_r) dr + g(\lambda_T), \quad \hat{\nu}\text{-a.s.}$$

This proves that $\hat{\nu}$ is supported on the set of optimal controls.

(ii) If (V.2.3) has a unique solution \mathbb{P}^* , then $\delta_{\mathbb{P}^*}$ is the unique accumulation point of $\{\mathbb{P}^0 \circ (m^N(\mathbf{Y}^{\hat{\tau}^N}))^{-1}\}$. As this sequence takes its values in a compact set, this implies that it converges weakly to $\delta_{\mathbb{P}^*}$, i.e., $m^N(\mathbf{Y}^{\hat{\tau}^N})$ converges weakly to \mathbb{P}^* . \blacksquare

Proof of Theorem V.3.4 (i) Without loss of generality, we assume that $(\Omega^0, \mathcal{F}_T^0, \mathbb{F}^0, \mathbb{P}^0)$ is sufficiently large so that there exists a sequence of i.i.d. processes Y^k , $k \geq 1$, driven by W^k , such that $\mathbb{P}_{Y^k}^0 = \mathbb{P}_Y$. Denote τ_k the jump time of I^k , and $\tau^N := (\tau_1, \dots, \tau_N)$, $N \geq 1$. For $k \in [N]$, let $\tilde{Y}^k := (\tilde{X}^k, \tilde{I}^k)$ be the process driven by

$$\tilde{X}_t^k = x_k^N + \int_0^t b(s, \tilde{X}_s^k, m_s) \tilde{I}_s^k ds + \int_0^t \sigma(s, \tilde{X}_s^k, m_s) \tilde{I}_s^k dW_s^k, \quad \tilde{I}_t^k := i_k^N \mathbf{1}_{t < \tau_k},$$

with $\mathbf{m} := \{m_t := \mathbb{P}_{Y_t} : t \in [-1, T]\}$. By standard estimates using the Lipschitz-continuity of b and σ , we may find a constant C (depending on T) such that

$$\mathbb{E}^{\mathbb{P}^0} [\|X_{\cdot, \wedge t}^{k, \tau^N} - \tilde{X}_{\cdot, \wedge t}^k\|_\infty^2] \leq C \int_0^t \mathbb{E}^{\mathbb{P}^0} [\|X_{\cdot, \wedge s}^{k, \tau^N} - \tilde{X}_{\cdot, \wedge s}^k\|_\infty^2 + \mathcal{W}_2^2(\bar{\mathbf{m}}_s^N, \mathbf{m}_s)] ds,$$

with $\bar{\mathbf{m}}^N := \{\bar{m}_t^N := m^N(\mathbf{Y}_t^{T^N}) : t \in [-1, T]\}$, and where $\|\cdot\|_{\cdot, \wedge t}$ denotes the norm ∞ of a path up to time t , and $\mathbf{m}_s := m_{\cdot, \wedge s}$ (and similarly for $\bar{\mathbf{m}}^N$ and $\tilde{\mathbf{m}}^N$ below). Thus, by Gronwall's lemma,

$$\mathbb{E}^{\mathbb{P}^0} [|X_{\cdot, \wedge t}^{k, \tau^N} - \tilde{X}_{\cdot, \wedge t}^k|^2] \leq C \int_0^t \mathbb{E}^{\mathbb{P}^0} [\mathcal{W}_2^2(\bar{\mathbf{m}}_s^N, \mathbf{m}_s)] ds. \quad (\text{V.6.4})$$

Introduce $\tilde{\mathbf{m}}^N := \{\tilde{m}_t^N := \frac{1}{N} \sum_{k=1}^N \delta_{\tilde{Y}_t^k} : t \in [-1, T]\}$. Note that

$$\mathcal{W}_2(\bar{\mathbf{m}}_t^N, \mathbf{m}_t) \leq \mathcal{W}_2(\bar{\mathbf{m}}_t^N, \tilde{\mathbf{m}}_t^N) + \mathcal{W}_2(\tilde{\mathbf{m}}_t^N, \mathbf{m}_t).$$

By (V.6.4) we have

$$\mathbb{E}^{\mathbb{P}^0} [\mathcal{W}_2^2(\bar{\mathbf{m}}_t^N, \tilde{\mathbf{m}}_t^N)] \leq \frac{1}{N} \mathbb{E}^{\mathbb{P}^0} \left[\sum_{k=1}^N |X_{\cdot, \wedge t}^{k, \tau^N} - \tilde{X}_{\cdot, \wedge t}^k|^2 \right] \leq C \left(\int_0^t \mathbb{E}^{\mathbb{P}^0} [\mathcal{W}_2^2(\bar{\mathbf{m}}_s^N, \mathbf{m}_s)] ds \right).$$

Then, by the Gronwall inequality again, we deduce

$$\sup_{0 \leq t \leq T} \mathbb{E}^{\mathbb{P}^0} [\mathcal{W}_2^2(\bar{\mathbf{m}}_t^N, \mathbf{m}_t)] \leq C \sup_{0 \leq t \leq T} \mathbb{E} [\mathcal{W}_2^2(\tilde{\mathbf{m}}_t^N, \mathbf{m}_t)]$$

On the other hand, by [61, Theorem 8.3 (ii)], we can easily see that

$$\lim_{N \rightarrow \infty} \sup_{0 \leq t \leq T} \mathbb{E}^{\mathbb{P}^0} [\mathcal{W}_2^2(\tilde{\mathbf{m}}_t^N, \mathbf{m}_t)] = 0, \quad (\text{V.6.5})$$

from which we eventually deduce that

$$\lim_{N \rightarrow \infty} \mathbb{E}^{\mathbb{P}^0} [\mathcal{W}_2^2(\bar{\mathbf{m}}^N, \mathbf{m})] = 0.$$

This implies that $\bar{\mathbf{m}}^N \rightarrow \mathbf{m}$ in probability, and thus, we may extract a subsequence that converges \mathbb{P}^0 -a.s. for \mathcal{W}_2 . This is the desired result.

(ii) By (i), there exists a sequence $\hat{\tau}^N$, $N \geq 1$, such that $m^N(\mathbf{Y}^{\hat{\tau}^N}) \xrightarrow{\mathcal{W}_2} \mathbb{P}_Y^*$, \mathbb{P}^0 -a.s. Thus, introducing

$$\varepsilon_N := V^N(0, m^N(\mathbf{y}^N)) - \mathbb{E}^{\mathbb{P}^0} \left[\int_0^T F(r, m^N(\mathbf{Y}_r^{\hat{\tau}^N})) dr + g(m^N(\mathbf{Y}_T^{\hat{\tau}^N})) \right],$$

we deduce from Theorem V.3.2, the boundedness of F and g and the optimality of \mathbb{P}^* that $\varepsilon_N \searrow 0$ (from N sufficiently large), and that $\hat{\tau}^N$ is ε_N -optimal for (V.2.9). \blacksquare

V.7 Multiple optimal stopping: viscosity solutions with tangency in mean

V.7.1 Formulation of the problem

We study the multiple optimal stopping problem in a more general context (without symmetry assumption on the drift and diffusion coefficients) as a subject of independent interest. More precisely, for $k \in [N]$, let $b_k, \sigma_k : [0, T] \times \mathbb{R}^{d \times N} \rightarrow \mathbb{R}^d$ and $\mathbb{S}_d^+(\mathbb{R})$, respectively, be functions satisfying the standard Lipschitz conditions. Let $\boldsymbol{\tau} := (\tau_1, \dots, \tau_N)$ be a family of \mathbb{F}^N -stopping times, and $\mathbf{X}^\tau, \mathbf{I}^\tau := (X^{1,\tau}, \dots, X^{N,\tau})^\top, (I^{1,\tau}, \dots, I^{N,\tau})^\top$ processes defined as the solution of the stochastic differential equations

$$X_t^{k,\tau} = x_k + \int_0^t b_k(s, \mathbf{X}_s^\tau) I_s^k ds + \int_0^t \sigma_k(s, \mathbf{X}_s^\tau) I_s^k dW_s^k, \quad I_s^k = i_k \mathbf{1}_{\tau_k > t} \text{ for all } k \in [N] \quad (\text{V.7.1})$$

where $\mathbf{x} := (x_1, \dots, x_N)^\top \in \mathbb{R}^{d \times N}$, $\mathbf{i} := (i_1, \dots, i_N)^\top \in \{0, 1\}^N$. We also denote by $\bar{\mathbf{X}} := \mathbf{X}^\mathbf{T}$, where $\mathbf{T} := (T, \dots, T)^\top \in \mathbb{R}^N$, the unstopped version of the dynamics (V.7.1). Denoting $\mathbf{y} := (\mathbf{x}, \mathbf{i}) \in \mathbf{S}^N$, and following Kobylanski, Quenez & Rouy-Mironescu [50], we consider the problem

$$\begin{aligned} V^N(t, \mathbf{y}) &:= \sup_{\boldsymbol{\tau} \in (\mathcal{T}_{t,T})^N} \mathbb{E}_{t,\mathbf{y}} \left[\frac{1}{N} \sum_{k=1}^N \int_t^{\tau_k} f^k(r, \mathbf{X}_r^\tau) dr + g(\mathbf{X}_T^\tau) \right] \\ &= \sup_{\boldsymbol{\tau} \in (\mathcal{T}_{t,T})^N} \mathbb{E}_{t,\mathbf{y}} \left[\frac{1}{N} \sum_{k=1}^N \int_t^T f^k(r, \mathbf{X}_r^\tau) I_r^k dr + g(\mathbf{X}_T^\tau) \right] \end{aligned} \quad (\text{V.7.2})$$

where $f^1, \dots, f^N \in C^0([0, T] \times \mathbb{R}^{d \times N}, \mathbb{R})$, $g \in C^0(\mathbb{R}^{d \times N}, \mathbb{R})$, and $\mathbb{E}_{t,\mathbf{y}}[\cdot] := \mathbb{E}_{t,\mathbf{y}}[\cdot | \mathbf{Y}_{t-}^\tau = \mathbf{y}]$, with $\mathbf{Y}^\tau := (\mathbf{X}^\tau, \mathbf{I}^\tau)$.

V.7.2 The single optimal stopping case

In this paragraph, we set $N = 1$ and $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$. We consider the optimization problem

$$V^1(t, x) := \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}_{t,x} \left[\psi(X_\tau) \right],$$

where X is defined by the dynamics $dX_s = b(r, X_r) dr + \sigma(r, X_r) dW_r$. The corresponding dynamic programming equation is given by the so-called obstacle heat equation

$$\min\{-(\partial_t + \mathcal{L})u, u - \psi\} = 0, \quad u(T, \cdot) = \psi, \quad (\text{V.7.3})$$

Let $\mathcal{T}_{t,T}^+ := \{H \in \mathcal{T}_{t,T} : H > t\}$. We consider the notion of viscosity solutions introduced by Ekren, Keller, Touzi & Zhang [37], where the sets of test functions are defined by

$$\begin{aligned} \bar{\mathcal{A}}u(t, x) &:= \left\{ \phi \in C^{1,2}([t, T] \times \mathbb{R}^d) : \right. \\ &\quad \left. \exists H \in \mathcal{T}_{t,T}^+ \text{ s.t. } (\phi - u)(t, x) = \max_{\theta \in \mathcal{T}_{t,T}} \mathbb{E}_{t,x} \left[(\phi - u)(\theta \wedge H, X_{\theta \wedge H}) \right] \right\}, \\ \underline{\mathcal{A}}u(t, x) &:= \left\{ \phi \in C^{1,2}([t, T] \times \mathbb{R}^d) : \right. \\ &\quad \left. \exists H \in \mathcal{T}_{t,T}^+ \text{ s.t. } (\phi - u)(t, x) = \min_{\theta \in \mathcal{T}_{t,T}} \mathbb{E}_{t,x} \left[(\phi - u)(\theta \wedge H, X_{\theta \wedge H}) \right] \right\}. \end{aligned}$$

Definition V.7.1 *We say $u : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ is a viscosity supersolution (resp. subsolution) of (V.7.3) if, for all $(t, x) \in [0, T] \times \mathbb{R}^d$,*

$$\min\{-(\partial_t + \mathcal{L})\phi(t, x), u(t, x) - \psi(x)\} \geq 0 \text{ (resp. } \leq 0) \text{ for all } \phi \in \bar{\mathcal{A}}u(t, x) \text{ (resp. } \underline{\mathcal{A}}u(t, x)).$$

Theorem V.7.2 *Assume ψ is Lipschitz continuous. Then V^1 is the unique continuous viscosity solution of (V.7.3).*

Proof The continuity of V^1 follows from the Lipschitz-continuity of ψ . Then, the fact that V^1 is a viscosity solution of (V.7.3) follows from Ekren [36], and the uniqueness from (i). ■

V.7.3 Reduction to a sequence of single optimal stopping problems

While our convergence results hold under symmetry assumptions on the dynamics of the N -particles system, we study the multiple optimal stopping problem in the general case, where each particle has its own dynamics.

Introduce $\Lambda_t^N := [t, T] \times \mathbf{S}^N$, and $\bar{\Lambda}_t^N := [t, T] \times \mathbf{S}^N$, $t \in [0, T]$.

Proof of Proposition V.4.3 By Kobylanski, Quenez & Rouy-Mironescu [50, Theorem

3.1], we have

$$\begin{aligned}
V^N(t, \mathbf{y}) &= \sup_{\theta \in \mathcal{T}_{t,T}} \mathbb{E}_{t, \mathbf{y}} \left[\max_{l \in [N]} \sup_{\boldsymbol{\tau} \in \mathcal{T}_{\theta, T}^{N-1}} \mathbb{E}_{\theta, \mathbf{Y}_{\theta}^{-l}} \left[\frac{1}{N} \sum_{k=1}^N \int_t^T f^k(r, \mathbf{X}_r^{\boldsymbol{\tau} \otimes_l \theta}) I_r^k dr + g(\mathbf{X}_T^{\boldsymbol{\tau} \otimes_l \theta}) \right] \right] \\
&= \sup_{\theta \in \mathcal{T}_{t,T}} \mathbb{E}_{t, \mathbf{y}} \left[\frac{1}{N} \sum_{k=1}^N \int_t^{\theta} f^k(r, \bar{\mathbf{X}}_r) I_r^k dr \right. \\
&\quad \left. + \max_{l \in [N]} \sup_{\boldsymbol{\tau} \in \mathcal{T}_{\theta, T}^{N-1}} \mathbb{E}_{\theta, \mathbf{Y}_{\theta}^{-l}} \left[\frac{1}{N} \sum_{k=1}^N \int_{\theta}^T f^k(r, \mathbf{X}_r^{\boldsymbol{\tau} \otimes_l \theta}) I_r^k dr + g(\mathbf{X}_T^{\boldsymbol{\tau} \otimes_l \theta}) \right] \right],
\end{aligned}$$

where

$$\mathbf{Y}^{-l} := (\bar{\mathbf{X}}, \mathbf{i}^{-l}), \mathbf{i}^{-l} := (i_1, \dots, i_{l-1}, 0, i_{l+1}, \dots, i_N)^{\top}, \text{ and } \boldsymbol{\tau} \otimes_l \theta := (\tau_1, \dots, \tau_{l-1}, \theta, \tau_{l+1}, \dots, \tau_N).$$

By the Markov property, we have

$$\sup_{\boldsymbol{\tau} \in \mathcal{T}_{\theta, T}^{N-1}} \mathbb{E}_{\theta, \mathbf{Y}_{\theta}^{-l}} \left[\frac{1}{N} \sum_{k=1}^N \int_{\theta}^T f^k(r, \mathbf{X}_r^{\boldsymbol{\tau} \otimes_l \theta}) I_r^k dr + g(\mathbf{X}_T^{\boldsymbol{\tau} \otimes_l \theta}) \right] = V^N(\theta, \bar{\mathbf{X}}_{\theta}, \mathbf{i}^{-l}).$$

By induction, we see that V^N is nondecreasing in \mathbf{i} for the partial order on $\{0, 1\}^N$, and thus

$$\max_{l \in [N]} V^N(\theta, \bar{\mathbf{X}}_{\theta}, \mathbf{i}^{-l}) = \max_{\mathbf{i}' < \mathbf{i}} V^N(\theta, \bar{\mathbf{X}}_{\theta}, \mathbf{i}'),$$

hence the desired result. ■

V.7.4 Viscosity solutions for the cascade obstacle equation

Our purpose is to show that V^N is the unique viscosity solution of the cascade obstacle equation (V.4.4). Define the sets of test functions

$$\begin{aligned}
\bar{\mathcal{A}}u(t, \mathbf{y}) &:= \left\{ \phi \in C^{1,2}(\Lambda_t^N) : \exists H \in \mathcal{T}_{t,T}^+ \text{ s.t. } (\phi - u)(t, \mathbf{y}) = \max_{\theta \in \mathcal{T}_{t,T}} \mathbb{E}_{t, \mathbf{y}} \left[(\phi - u)(\theta \wedge H, \bar{\mathbf{Y}}_{\theta \wedge H}) \right] \right\}, \\
\underline{\mathcal{A}}u(t, \mathbf{y}) &:= \left\{ \phi \in C^{1,2}(\Lambda_t^N) : \exists H \in \mathcal{T}_{t,T}^+ \text{ s.t. } (\phi - u)(t, \mathbf{y}) = \min_{\theta \in \mathcal{T}_{t,T}} \mathbb{E}_{t, \mathbf{y}} \left[(\phi - u)(\theta \wedge H, \bar{\mathbf{Y}}_{\theta \wedge H}) \right] \right\},
\end{aligned}$$

where $\phi \in C^{1,2}(\Lambda_t^N)$ means that $\partial_{x_k} \phi(s, \mathbf{y}), \partial_{x_k, x_k} \phi(s, \mathbf{y})$ exist and are continuous for all $(s, \mathbf{y}) \in \Lambda_t^N$ s.t. $i_k = 1$.

Definition V.7.3 Let $u : \Lambda_0^N \rightarrow \mathbb{R}$ and $\mathbf{f} := (f^1, \dots, f^N)^\top$.

(i) u is a viscosity supersolution of (V.4.4) if, for all $(t, \mathbf{y}) \in \Lambda_0^N$,

$$\min\{-\partial_t \phi(t, \mathbf{y}) - (\mathcal{L}\phi + \mathbf{f})(t, \mathbf{y}) \cdot \mathbf{i}, u(t, \mathbf{y}) - \max_{\mathbf{i}' < \mathbf{i}} u(t, \mathbf{x}, \mathbf{i}')\} \geq 0 \quad \text{for all } \phi \in \overline{\mathcal{A}}u(t, \mathbf{y}).$$

(ii) u is a viscosity subsolution of (V.4.4) if, for all $(t, \mathbf{y}) \in \Lambda_0^N$,

$$\min\{-\partial_t \phi(t, \mathbf{y}) - (\mathcal{L}\phi + \mathbf{f})(t, \mathbf{y}) \cdot \mathbf{i}, u(t, \mathbf{y}) - \max_{\mathbf{i}' < \mathbf{i}} u(t, \mathbf{x}, \mathbf{i}')\} \leq 0 \quad \text{for all } \phi \in \underline{\mathcal{A}}u(t, \mathbf{y}).$$

(iii) u is a viscosity solution of (V.4.4) if it is a viscosity supersolution and subsolution, and if $u|_{t=T} = u|_{\mathbf{i}=0} = g$.

Proof of Theorem V.4.4. As the f^k and g are Lipschitz-continuous, the continuity of V^N proceeds from standard SDE estimates. By (V.4.3), the multiple optimal stopping problem (V.7.2) may be written as a single optimal stopping problem. More precisely, $(t, \mathbf{x}) \mapsto V^N(t, \mathbf{x}, \mathbf{i})$ is a standard optimal stopping problem for every fixed $\mathbf{i} \in \{0, 1\}^N$, with obstacle $(t, \mathbf{x}) \mapsto \max_{\mathbf{i}' < \mathbf{i}} V^N(t, \mathbf{x}, \mathbf{i}')$. We may then apply the result of Theorem V.7.2 (ii). ■

V.7.5 Construction of an optimal stopping strategy

We start by introducing the family of stopping times which will turn out to be optimal for (V.7.2). Given $\tau \in (\mathcal{T}_{t,T})^N$ and the corresponding process $\mathbf{I} := \mathbf{I}^\tau$, we slightly modify our notations and write $\mathbf{X}^{\mathbf{I}} := \mathbf{X}^\tau$. Let u be continuous viscosity solutions of (V.4.4) with subquadratic growth. Given a solution of (V.4.4) u , we define recursively the stopping times $\hat{\tau} := (\hat{\tau}_1, \dots, \hat{\tau}_N)$ and the random $\{0, 1\}^N$ -valued vectors $\{\mathbf{i}^k\}_{k \in [N-1]}$ as follows:

- $\hat{\tau}_0 := t, \mathbf{I}^{(0)} := \mathbf{i}^0 := \mathbf{i}$,
- $\hat{\tau}_{k+1} := \inf \left\{ s \geq \hat{\tau}_k : u(s, \mathbf{X}_s^{\mathbf{I}^{(k)}}, \mathbf{i}^k) = \max_{\mathbf{i}' < \mathbf{i}^k} u(s, \mathbf{X}_s, \mathbf{i}') \right\}$ (V.7.4)
- \mathbf{i}^{k+1} is the smallest element of $\operatorname{argmax}_{\mathbf{i}' < \mathbf{i}^k} u(\hat{\tau}_{k+1}, \mathbf{X}_{\hat{\tau}_{k+1}}^{\mathbf{I}^{(k)}}, \mathbf{i}')$ for the lexicographic order
- $\mathbf{I}_s^{(k+1)} = \mathbf{I}_s^{(k)} \mathbf{1}_{s < \hat{\tau}_{k+1}} + \mathbf{i}^{k+1} \mathbf{1}_{s \geq \hat{\tau}_{k+1}}$ for all $s \in [-1, T]$

for $k = 0, \dots, N-1$. In that case, the above elements can be interpreted as follows:

- $\hat{\tau}_k$ is the stopping time of the k th particle;

- the (random) vector \mathbf{i}^k is the state of the particles system on $[\hat{\tau}_k, \hat{\tau}_{k+1})$: if $i_j^k = 1$, then the particle j is still diffusing on this interval; otherwise, it is stopped; the use of the lexicographic orders allows its uniqueness, hence its wellposedness;
- the sequence $(\mathbf{I}^{(k)})_{k \in [N]}$, defined recursively, allows to construct an optimal control $\mathbf{I}^* := \mathbf{I}^{(N)}$.

Denote $\boldsymbol{\tau}^* := (\tau_1^*, \dots, \tau_N^*)$ the stopping times whose order statistics correspond to $\hat{\tau}$ defined by (V.7.4), i.e. $\tau_j^* := \inf\{s \geq t : I_s^{*,j} = 0\}$ for all $j \in [N]$. We shall prove it is an optimal stopping rule for (V.7.2), or equivalently that $\mathbf{I}^{(N)}$ is optimal for (V.7.2).

Theorem V.7.4 *Assume the value function V^N is continuous, and is the unique viscosity solution of (V.4.4). Then $\boldsymbol{\tau}^*$ is an optimal stopping policy for the problem (V.7.2).*

Proof Recall from Proposition V.4.3 that

$$V^N(t, \mathbf{y}) = \sup_{\theta \in \mathcal{T}_{t,T}} \mathbb{E}_{t,\mathbf{y}} \left[\frac{1}{N} \sum_{k=1}^N \int_t^\theta f^k(r, \bar{\mathbf{X}}_r) I_r^k dr + \max_{\mathbf{i}' < \mathbf{i}} V^N(\theta, \bar{\mathbf{X}}_\theta, \mathbf{i}') \right]. \quad (\text{V.7.5})$$

As V^N is continuous, it follows from the standard optimal stopping theory that an optimal stopping strategy for (V.7.5) is

$$\tilde{\tau}_{\mathbf{i},t} := \inf \left\{ s \geq t : V^N(s, \bar{\mathbf{X}}_s, \mathbf{i}) = \max_{\mathbf{i}' < \mathbf{i}} V^N(s, \bar{\mathbf{X}}_s, \mathbf{i}') \right\}.$$

Given that $u = V^N$ by uniqueness of the solution of (V.4.4), we see by direct induction that $\tilde{\tau}_{\mathbf{i}^k, \hat{\tau}_{k-1}} = \hat{\tau}_k$ for all $1 \leq k \leq N$, and thus

$$V^N(t, \mathbf{y}) = u(t, \mathbf{y}) = \mathbb{E}_{t,\mathbf{y}} \left[\frac{1}{N} \sum_{k=1}^N \int_t^T f^k(r, \mathbf{X}_r^{\boldsymbol{\tau}^*}) I_r^k dr + g(\mathbf{X}_T^{\boldsymbol{\tau}^*}) \right].$$

■

Structure of the optimal stopping times Introduce the set-valued random flow:

$$A_s := \{j \in [N] : I_s^{*,j} = 1\}, \text{ for all } s \in [t, T],$$

which takes its values in the collection of subsets of $[N]$, and

$$\mathcal{X}_s := \{X_s^{*,j}\}_{j \in A_s}, \text{ for all } s \in [t, T],$$

which is the set of the positions of the particles that are not stopped yet at time s . Then, we may rewrite the optimal stopping times τ^* as

$$\tau_j^* = \inf\{s \geq t : X_s^{\tau^*,j} \notin \mathcal{X}_s\} = \tau^j(X_s^{\tau^*,j}, \mathcal{X}) \text{ for all } j \in [N]. \quad (\text{V.7.6})$$

In the symmetric case (i.e., in the setting of Section V.2.3), we may specify further the structure of the optimal stopping times (V.7.6). Denoting $\mathbf{X}_s^{\tau^*, \mathcal{X}} := (X_s^{\tau^*,j})_{j \in \mathcal{X}_s}^\top$ for all $s \in [t, T]$ (i.e., the vectors of the processes not stopped yet a time s), we observe the optimal stopping times have the following structure:

$$\tau_j^* = \tau^*(X_s^{\tau^*,j}, m^{|\mathcal{X}_s|}(\mathbf{X}_s^{\tau^*, \mathcal{X}})) := \inf\left\{s \geq t : \tilde{Z}_s \notin \text{supp}(m^M(\mathbf{Z}_s))\right\} \text{ for all } j \in [N]. \quad (\text{V.7.7})$$

As we expect the flow $m^{|\mathcal{X}_s|}(\mathbf{X}_s^{\tau^*, \mathcal{X}})_{s \in [t, T]}$ to tend towards a flow of deterministic distribution as $N \rightarrow \infty$, the optimal stopping times in the mean field limit is expected to have the closed-loop structure $\tau^* = \tau^*(\mathbf{X})$.

V.8 Approximation of randomized stopping strategies by pure strategies

In this appendix, we prove that any immediate stopping strategy can be approximated by a pure strategy (see [74, §4.3] for a discussion on randomized and pure stopping policies).

Lemma V.8.1 *Denote (ξ, \mathcal{I}) the canonical map on $\mathbb{R}^d \times \{0, 1\}$, and let $m \in \mathcal{P}_2(\mathbf{S})$ be s.t. $m \circ \xi^{-1}$ has a continuous density. Let $m' \preceq m$. For all $\varepsilon > 0$, there exists $A \in \mathcal{B}(\mathbb{R}^d)$ s.t. $\mathcal{W}_2(m^A, m') \leq \varepsilon$.*

Proof We assume without loss of generality that $m \circ \mathcal{I}^{-1} = \delta_1$. It suffices to show that m' may be rewritten as the limit of a sequence $\{m \circ (\xi, \mathbf{1}_{A_n}(\xi))^{-1}\}_{n \geq 1}$, with $A_n \in \mathcal{B}(\mathbb{R}^d)$. First, we compute, for $(\alpha, \beta) \in \mathbb{R}^d \times \mathbb{R}$, the characteristic function

$$\begin{aligned} \mathbb{E}^m [e^{i\alpha \cdot \xi + i\beta \mathcal{I}}] &= \mathbb{E}^m [(e^{i\alpha \cdot \xi + i\beta} - e^{i\alpha \cdot \xi}) \mathcal{I}] + \mathbb{E}^m [e^{i\alpha \cdot \xi}] \\ &= \int_K \psi(x; \alpha, \beta) f(x) \mathbb{E}^m [\mathcal{I} | \xi = x] dx + \mathbb{E}^m [e^{i\alpha \cdot \xi}], \end{aligned}$$

where K and f are the support and the density of ξ , and $\psi(x; \alpha, \beta) := e^{i\alpha \cdot x + i\beta} - e^{i\alpha \cdot x}$.

Step 1: Fix $n \geq 1$, and assume K is compact. Then, there exists $J_n \geq 1$ and two sequences $\{x_j^n \in K\}_{1 \leq j \leq J_n}$ and $\{r_j^n \geq 0\}_{1 \leq j \leq J_n}$ s.t.

$$\bigcup_{j=1}^{J_n} B(x_j^n, r_j^n) \subset K, \max_{1 \leq j \leq J_n} r_j^n \leq \frac{1}{n}, \bigcup_{j=1}^{J_n} B(x_j^n, r_j^n) \text{ is disjoint and } \text{Leb}(K \setminus \bigcup_{j=1}^{J_n} B(x_j^n, r_j^n)) \leq 2^{-n},$$

where Leb denotes the Lebesgue measure on \mathbb{R}^d . There also exists $\{\tilde{r}_j^n \geq 0\}_{1 \leq j \leq J_n}$ s.t.

$$\text{Leb}(B(x_j^n, \tilde{r}_j^n)) = \int_{B(x_j^n, \tilde{r}_j^n)} \mathbb{E}^m[\mathcal{I}|\xi = x] dx \text{ for all } j \leq J_n.$$

Introduce $A_n := \bigcup_{j=1}^{J_n} B(x_j^n, \tilde{r}_j^n)$. Note that $\tilde{r}_j^n \leq r_j^n$ for all $j \leq J_n$, and thus A_n is also a disjoint union. We then have:

$$\begin{aligned} \left| \mathbb{E}^m[e^{i\alpha \cdot \xi + i\beta \mathcal{I}}] - \mathbb{E}^m[e^{i\alpha \cdot \xi + i\beta \mathbf{1}_{A_n}(\xi)}] \right| &\leq \sum_{j=1}^{J_n} \left| \int_{B(x_j^n, r_j^n)} \psi(x; \alpha, \beta) f(x) \mathbb{E}[\mathcal{I}|\xi = x] dx \right. \\ &\quad \left. - \int_{B(x_j^n, \tilde{r}_j^n)} \psi(x; \alpha, \beta) f(x) dx \right| + |\psi f|_\infty 2^{-n} \\ &\leq \sum_{j=1}^{J_n} \int_{B(x_j^n, r_j^n)} |\psi(x; \alpha, \beta) f(x) - \psi(x_j^n; \alpha, \beta) f(x_j^n)| \\ &\quad + \int_{B(x_j^n, \tilde{r}_j^n)} |\psi(x_j^n; \alpha, \beta) f(x_j^n) - \psi(x; \alpha, \beta) f(x)| dx + |\psi f|_\infty 2^{-n} \\ &\leq \sum_{j=1}^{J_n} r_j^n \rho(r_j^n; \alpha, \beta) + \tilde{r}_j^n \rho(\tilde{r}_j^n; \alpha, \beta) + |\psi f|_\infty 2^{-n} \leq 2 \left(\sum_{j=1}^{J_n} r_j^n \right) \rho\left(\frac{1}{n}; \alpha, \beta\right) + |\psi f|_\infty 2^{-n} \\ &\leq 2\text{Leb}(K) \rho\left(\frac{1}{n}; \alpha, \beta\right) + |\psi f|_\infty 2^{-n} \end{aligned}$$

for some modulus of continuity $\rho(\cdot; \alpha, \beta)$ of $\psi(\cdot; \alpha, \beta) f$, which is continuous hence uniformly continuous on the compact K . As α and β are arbitrary, this proves the desired result.

Step 2: Assume K is not compact and fix $\varepsilon > 0$. Let $K^\varepsilon \subset \mathbb{R}^d$ be a compact set s.t. $\int_{K^\varepsilon} f(x) dx \geq 1 - \varepsilon$. Then, applying the procedure of Case 1 to K^ε , we have, for some modulus of continuity $\rho_\varepsilon(\cdot; \alpha, \beta)$,

$$\begin{aligned} \left| \mathbb{E}^m[e^{i\alpha \cdot \xi + i\beta \mathcal{I}}] - \mathbb{E}^m[e^{i\alpha \cdot \xi + i\beta \mathbf{1}_{A_n}(\xi)}] \right| &\leq \left| \mathbb{E}^m[e^{i\alpha \cdot \xi + i\beta \mathcal{I}} \mathbf{1}_{K^\varepsilon}(\xi)] - \mathbb{E}^m[e^{i\alpha \cdot \xi + i\beta \mathbf{1}_{A_n}(\xi)} \mathbf{1}_{K^\varepsilon}(\xi)] \right| \\ &\quad + \left| \mathbb{E}^m[e^{i\alpha \cdot \xi + i\beta \mathcal{I}} \mathbf{1}_{(K^\varepsilon)^c}(\xi)] - \mathbb{E}^m[e^{i\alpha \cdot \xi + i\beta \mathbf{1}_{A_n}(\xi)} \mathbf{1}_{(K^\varepsilon)^c}(\xi)] \right| \\ &\leq 2\text{Leb}(K^\varepsilon) \rho_\varepsilon\left(\frac{1}{n}; \alpha, \beta\right) + |\psi f|_{K^\varepsilon, \infty} 2^{-n} + 2\varepsilon \end{aligned}$$

Since n is independent from ε , this completes the proof. ■

V.9 A useful estimate

Lemma V.9.1 *Let $\mathbf{x}^N \in \mathbb{R}^{d \times N}$, $N \geq 1$, such that $m^N(\mathbf{x}^N) \xrightarrow{\mathcal{W}_2} m$ for some $m \in \mathcal{P}_2(\mathbf{S})$. Let \mathbf{X}^τ be the process defined by the dynamics (V.2.9) starting from $\mathbf{X}_0^\tau = \mathbf{x}^N$, for some $\tau \in (\mathcal{T}_{0,T}^N)^N$. Then there exists a constant $C \geq 0$, independent from N , such that*

$$\mathbb{E} \left[\frac{1}{N} \sum_{k=1}^N |X_s^{k,\tau} - X_t^{k,\tau}|^2 \right] \leq C |s - t| \quad \text{for all } t, s \in [0, T].$$

Proof For $k \in [N]$, denote $b_k(t, \mathbf{x}), \sigma_k(t, \mathbf{x}) := b(t, x_k, m^N(\mathbf{x})), \sigma_k(t, x_k, m^N(\mathbf{x}))$. We have

$$\begin{aligned} |b_k(r, \mathbf{x})|^2 &\leq L \left(1 + |x_k|^2 + \mathcal{W}_2^2(\delta_0, m^N(\mathbf{x})) \right) \\ &\leq L \left(1 + |x_k|^2 + \frac{1}{N} |\mathbf{x}|^2 \right) \end{aligned}$$

for all $\mathbf{x} \in \mathbb{R}^{d \times N}$, with L a Lipschitz constant, and similarly for σ_k . Then, applying standard estimates to the process $\frac{1}{N} \sum_{k=1}^N X^{k,\tau}$, we have

$$\mathbb{E} \left[\frac{1}{N} \sum_{k=1}^N |X_s^{k,\tau} - X_t^{k,\tau}|^2 \right] \leq \tilde{C} \left(1 + \frac{1}{N} |\mathbf{x}^N|^2 \right) |s - t|$$

for some constant $\tilde{C} \geq 0$ depending on L and T only. As $m^N(\mathbf{x}^N) \xrightarrow{\mathcal{W}_2} m$, the sequence $\{\frac{1}{N} |\mathbf{x}^N|^2\}_{N \geq 1}$ converges, and thus is bounded. This provides the desired result. \blacksquare

Chapter VI

A finite-dimensional approximation for viscosity solutions of partial differential equations on Wasserstein space

VI.1 Introduction

The objective of the chapter is to find a finite-dimensional approximation for a general class of PDEs on Wasserstein space, satisfied in the sense of viscosity. This can be seen as the counterpart of Chapter V for a larger class of equations. We use the notion of viscosity solutions developed by Wu & Zhang [77], which is intrinsic and allows for path-dependent PDEs (i.e., the solutions of the equations depend on the time and on a probability measure on the space of continuous paths of \mathbb{R}^d). This class of equations covers in particular equations arising in mean field stochastic control (including the case of controlled volatility). We intend to find a finite-dimensional PDE whose viscosity solutions converge to viscosity solutions of the equation on Wasserstein space. More precisely, in the spirit of Barles & Souganidis [2], we prove that the semi-relaxed limits of the viscosity super/subsolutions of the finite-dimensional PDE are viscosity super/subsolutions of the PDE on Wasserstein space as the dimension goes to infinity. Assuming that the comparison principle holds for the latter equation, this implies the convergence of viscosity solutions.

An important feature that ensures the convergence relies on the choice of the set of the test functions for the finite dimensional problem. Similarly to the viscosity theory developed for path-dependent PDEs (see e.g. Ekren, Keller, Touzi & Zhang [37], Ekren, Touzi & Zhang [39, 40] and Ren, Touzi & Zhang [67]), we only require test functions to be tangent to the super/subsolution through the mean, whereas the tangency is pointwise in the standard literature. This tangency in expectation can be seen as the finite-dimensional counterpart of our set of test functions on Wasserstein space.

Another key ingredient lies in a propagation of chaos-like result for multidimensional semimartingales with bounded characteristics. In the spirit of Lacker [52], who proves a propagation of chaos result for SDEs with relaxed controls, we combine compactness arguments and a martingale problem approach to ensure that the empirical mean of such semimartingales converges to the law of a semimartingale with bounded characteristics. This result allows us to connect our set of test functions for the finite-dimensional equation to our set of test functions on Wasserstein space.

The chapter is organized as follows. In Section VI.2, we present the class of equations on Wasserstein space and the notion of viscosity solutions. In Section VI.3, we introduce the finite-dimensional approximation and state the main result of the chapter, the convergence theorem. In Section VI.4, we present some examples. Section VI.5 extends our results to the case of path-dependent PDEs. Sections VI.6 and VI.7 are respectively dedicated to the proof of the convergence theorem and the propagation of chaos result.

Notations. We denote $\langle \mu, f \rangle := \int f d\mu$. $\mathbb{S}_{d \times N}^D$ denotes the set of blockwise diagonal matrices of the form $\text{Diag}(A_1, \dots, A_N)$, where each $A_i \in \mathbb{S}_d$. Given $\mathbf{x} := (x_1, \dots, x_N) \in E^N$, we denote by $\mu^N(\mathbf{x}) := \frac{1}{N} \sum_{i=1}^N \delta_{x_i} \in \mathcal{P}(E)$, and $\mathcal{P}^N(E) := \{\mu^N(\mathbf{x}) : \mathbf{x} \in E^N\}$.

VI.2 Viscosity solutions of partial differential equations on Wasserstein space

VI.2.1 Differentiability on Wasserstein space

For $t \in [0, T)$, we denote

$$\mathbf{Q}_t := [t, T) \times \mathcal{P}_2(\mathbb{R}^d) \quad \text{and} \quad \overline{\mathbf{Q}}_t := [t, T] \times \mathcal{P}_2(\mathbb{R}^d).$$

Definition VI.2.1 Fix $t \in [0, T)$.

(i) $u : \overline{\mathbf{Q}}_t \longrightarrow \mathbb{R}$ has a functional linear derivative if there exists $\delta_m u : \overline{\mathbf{Q}}_0 \times \mathbb{R}^d \rightarrow \mathbb{R}$ satisfying, for any $s \in [t, T]$ and $\mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$,

$$u(t, \mu') - u(t, \mu) = \int_0^1 \int_{\mathbb{R}^d} \delta_m u(s, \lambda \mu' + (1 - \lambda)\mu, x)(\mu' - \mu)(dx) d\lambda,$$

and $\delta_m u$ has quadratic growth in $x \in \mathbb{R}^d$, locally uniformly in $(s, m) \in \overline{\mathbf{Q}}_t$.

(ii) We denote by $C_b^{1,2}(\overline{\mathbf{Q}}_t)$ the set of bounded functions $u : \overline{\mathbf{Q}}_t \rightarrow \mathbb{R}$ such that $\partial_t u$, $\delta_m u$, $\partial_x \delta_m u$, $\partial_{xx}^2 \delta_m u$ exist and are continuous and bounded in all their variables.

VI.2.2 Partial differential equation on Wasserstein space

Let $F : [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R} \times \mathbb{L}_2^0(\mathbb{R}^d, \mathbb{R}^d) \times \mathbb{L}_2^0(\mathbb{R}^d, \mathbb{S}_d) \longrightarrow \mathbb{R}$, where $\mathbb{L}_2^0(\mathbb{R}^d, \mathbb{R}^d)$ (resp. $\mathbb{L}_2^0(\mathbb{R}^d, \mathbb{S}_d)$) denotes the set of Borel measurable functions from \mathbb{R}^d to \mathbb{R}^d (resp. \mathbb{S}_d) with quadratic growth, and $g : \mathcal{P}_2(\mathbb{R}^d) \longrightarrow \mathbb{R}$. We consider the following equation:

$$-\partial_t u(t, \mu) - F(t, \mu, u(t, \mu), \partial_x \delta_m u(t, \mu, \cdot), \partial_{xx}^2 \delta_m u(t, \mu, \cdot)) = 0, \quad u|_{t=T} = g, \quad (\text{VI.2.1})$$

for all $(t, \mu) \in \mathbf{Q}_0$.

Assumption VI.2.2 (i) F is continuous in the following sense:

$$F(t', \mu', y', Z', \Gamma') \longrightarrow F(t, \mu, y, Z, \Gamma) \text{ as } (t', \mu', y', Z', \Gamma') \rightarrow (t, \mu, y, Z, \Gamma),$$

where the convergence of μ' to μ is in $(\mathcal{P}_2(\mathbb{R}^d), \mathcal{W}_2)$ and the convergence of (Z', Γ') to (Z, Γ) is pointwise for $Z, Z' \in C_b^0(\mathbb{R}^d, \mathbb{R}^d)$ and $\Gamma, \Gamma' \in C_b^0(\mathbb{R}^d, \mathbb{S}_d)$.

(ii) For all $(t, \mu, y, Z, \Gamma) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R} \times \mathbb{L}_2^0(\mathbb{R}^d, \mathbb{R}^d) \times \mathbb{L}_2^0(\mathbb{R}^d, \mathbb{S}_d)$, we have

$$F(t, \mu, y, Z, \Gamma) = F(t, \mu, y, Z', \Gamma')$$

for all Z', Γ' s.t. $Z'|_{\text{supp}(\mu)}, \Gamma'|_{\text{supp}(\mu)} = Z|_{\text{supp}(\mu)}, \Gamma|_{\text{supp}(\mu)}$.

VI.2.3 Viscosity solutions

Let $L > 0$, $\Omega := C^0([0, T], \mathbb{R}^d)$, X be the canonical process on Ω and $\mathbb{F} := \{\mathcal{F}_t\}_{t \in [0, T]}$ be the corresponding filtration.

Definition VI.2.3 We denote by \mathcal{P}_L the set of measures $\mathbb{P} \in \mathcal{P}_2(\Omega, \mathcal{F}_T)$ s.t. X is a \mathbb{P} -semimartingale with drift and diffusion characteristics uniformly bounded by L . For $(t, \mu) \in \mathbf{Q}_0$, we also define $\mathcal{P}_L(t, \mu) := \{\mathbb{P} \in \mathcal{P}_L : \mathbb{P}_{X_t} = \mu\}$.

CHAPTER VI. A FINITE-DIMENSIONAL APPROXIMATION FOR VISCOSITY SOLUTIONS OF PARTIAL DIFFERENTIAL EQUATIONS ON WASSERSTEIN SPACE

We define the viscosity neighborhood of $(t, \mu) \in \mathbf{Q}_0$ by

$$\mathcal{N}_\delta(t, \mu) := \{(s, \mathbb{P}_{X_s}) : s \in [t, t + \delta], \mathbb{P} \in \mathcal{P}_L(t, \mu)\},$$

which is compact under \mathcal{W}_2 (see Wu & Zhang [77, Lemma 4.1]). We may then introduce the sets of test functions:

$$\begin{aligned} \overline{\mathcal{A}}u(t, \mu) &:= \left\{ \varphi \in C_b^{1,2}(\overline{\mathbf{Q}}_t) : (\varphi - u)(t, \mu) = \max_{\mathcal{N}_\delta(t, \mu)} (\varphi - u) \text{ for some } \delta > 0 \right\}, \\ \underline{\mathcal{A}}u(t, \mu) &:= \left\{ \varphi \in C_b^{1,2}(\overline{\mathbf{Q}}_t) : (\varphi - u)(t, \mu) = \min_{\mathcal{N}_\delta(t, \mu)} (\varphi - u) \text{ for some } \delta > 0 \right\}. \end{aligned}$$

Definition VI.2.4 Let $u : \overline{\mathbf{Q}}_0 \rightarrow \mathbb{R}$.

(i) u is a viscosity supersolution of (VI.2.1) if, for all $(t, \mu) \in \mathbf{Q}_0$ and $\varphi \in \overline{\mathcal{A}}u(t, \mu)$,

$$-\partial_t \varphi(t, \mu) - F(t, \mu, u(t, \mu), \partial_x \delta_m \varphi(t, \mu, \cdot), \partial_{xx}^2 \delta_m \varphi(t, \mu, \cdot)) \geq 0.$$

(ii) u is a viscosity subsolution of (VI.2.1) if, for all $(t, \mu) \in \mathbf{Q}_0$ and $\varphi \in \underline{\mathcal{A}}u(t, \mu)$,

$$-\partial_t \varphi(t, \mu) - F(t, \mu, u(t, \mu), \partial_x \delta_m \varphi(t, \mu, \cdot), \partial_{xx}^2 \delta_m \varphi(t, \mu, \cdot)) \leq 0.$$

(iii) u is a viscosity solution of (VI.2.1) if it is a viscosity supersolution and subsolution.

Definition of viscosity solutions via semi-jets. Fix $(t, \mu) \in \mathbf{Q}_0$ and $\delta > 0$. For $(v, a, f) \in \mathbb{R} \times \mathbb{R} \times C_b^2(\mathbb{R}^d)$, introduce

$$\psi^{v,a,f}(s, \nu) := v + a(s - t) + \langle \nu - \mu, f \rangle \quad \text{for all } (s, \nu) \in \mathcal{N}_\delta(t, \mu). \quad (\text{VI.2.2})$$

We have the equivalence result:

Proposition VI.2.5 Let $u : \mathbf{Q}_0 \rightarrow \mathbb{R}$.

(i) u is a viscosity supersolution of (VI.2.1) if and only if it satisfies the viscosity supersolutions properties for all test functions of the form (VI.2.2).

(ii) u is a viscosity subsolution of (VI.2.1) if and only if it satisfies the viscosity subsolutions properties for all test functions of the form (VI.2.2).

Proof We only provide the argument for (i). If u is a viscosity supersolution of (VI.2.1), then it satisfies the supersolution property for all $\varphi \in \overline{\mathcal{A}}u(t, \mu)$, in particular for those of the form (VI.2.2).

CHAPTER VI. A FINITE-DIMENSIONAL APPROXIMATION FOR VISCOSITY SOLUTIONS OF PARTIAL DIFFERENTIAL EQUATIONS ON WASSERSTEIN SPACE

Assume now that the supersolution property is verified for all $\psi^{v,a,f} \in \overline{\mathcal{A}u}(t, m)$, $(v, a, f) \in \mathbb{R} \times \mathbb{R} \times C_b^2(\mathbb{R}^d)$. Let $\varphi \in \underline{\mathcal{A}u}(t, \mu)$ and $\varepsilon > 0$, and fix $(v, a, f) := (\varphi(t, \mu), \partial_t \varphi(t, \mu) - \varepsilon, \delta_m \varphi(t, \mu, \cdot))$. We have, for $(s, \nu) \in \mathcal{N}_\delta(t, \mu)$:

$$\begin{aligned} (\varphi - \psi^{v,a,f})(s, \nu) &= \varphi(s, \nu) - \varphi(t, \mu) - a(s-t) - \langle \nu - \mu, f \rangle \\ &= (\varphi(s, \nu) - \varphi(t, \nu) - a(s-t)) + (\varphi(t, \nu) - \varphi(t, \mu) - \langle \nu - \mu, f \rangle) \\ &= (s-t)(\partial_t \varphi(t, \nu) + \eta(s-t) - \partial_t \varphi(t, \mu) + \varepsilon) \\ &\quad + \int_0^1 \langle \nu - \mu, \delta_m \varphi(t, \lambda\mu + (1-\lambda)\nu, \cdot) - \delta_m \varphi(t, \mu, \cdot) \rangle d\lambda \end{aligned}$$

where $\eta(s-t) \xrightarrow{s \rightarrow t} 0$. As $(s, \nu) \in \mathcal{N}_\delta(t, \mu)$, there exists $\mathbb{P} \in \mathcal{P}(t, \mu)$ s.t. $\nu = \mathbb{P}_{X_s}$. Thus, introducing $h_{t,s}^\lambda := \delta_m \varphi(t, \lambda\mu + (1-\lambda)\nu, \cdot) - \delta_m \varphi(t, \mu, \cdot)$, we have

$$\langle \nu - \mu, \delta_m \varphi(t, \lambda\mu + (1-\lambda)\nu, \cdot) - \delta_m \varphi(t, \mu, \cdot) \rangle = \mathbb{E}^\mathbb{P} [h_{t,s}^\lambda(X_s) - h_{t,s}^\lambda(X_t)].$$

As φ is smooth, we may apply Itô's formula to $h_{t,s}^\lambda$, and thus

$$\mathbb{E}^\mathbb{P} [h_{t,s}^\lambda(X_s) - h_{t,s}^\lambda(X_t)] = \mathbb{E}^\mathbb{P} \left[\int_t^s \partial_x h_{t,s}^\lambda(X_r) \cdot dX_r + \frac{1}{2} \partial_{xx}^2 h_{t,s}^\lambda(X_r) : d\langle X \rangle_r \right] \leq (s-t) \frac{\varepsilon}{2},$$

for all $s \in [t, t + \delta]$ and δ sufficiently small, given the boundedness of the characteristics of X under \mathbb{P} , the boundedness and continuity of the derivatives of φ and the continuity of the flow $s \mapsto \mathbb{P}_{X_s}$. Finally, as we also have $\partial_t \varphi(t, \nu) + \eta(s-t) - \partial_t \varphi(t, \mu) \geq -\frac{\varepsilon}{2}$ for δ sufficiently small, we have

$$(\varphi - \psi^{v,a,f})(s, \nu) \geq (s-t) \left(-\frac{\varepsilon}{2} + \varepsilon - \frac{\varepsilon}{2} \right) = 0 \quad \text{for all } (s, \nu) \in \mathcal{N}_\delta(t, \mu),$$

which implies since $\psi^{v,a,f}(t, \mu) = \varphi(t, \mu)$ that $\psi^{v,a,f} \in \overline{\mathcal{A}u}(t, \mu)$. Then, the supersolution property writes

$$-(\partial_t \varphi(t, \mu) - \varepsilon) - F(t, \mu, u(t, \mu), \partial_x \delta_m \varphi(t, \mu, \cdot), \partial_{xx}^2 \delta_m \varphi(t, \mu, \cdot)) \geq 0,$$

and we obtain the desired result by letting $\varepsilon \rightarrow 0$. ■

Under additional assumptions on F (see Wu & Zhang [77, Assumption 3.1]), viscosity solutions satisfy the usual properties of consistency with the classical solution and stability. However, the uniqueness result has only been proved for some specific cases (see [77, Theorem 4.13]). Therefore, we shall use the comparison principle as a standing assumption:

Assumption VI.2.6 (Comparison principle) *Let u, v be respectively USC viscosity subsolution and LSC supersolution of (VI.2.1) such that $u(T, \cdot) \leq v(T, \cdot)$. Then $u \leq v$.*

VI.3 Finite-dimensional approximation

VI.3.1 The approximating equation

Let $N \geq 1$. We shall write in bold character the elements $\mathbf{x} = (x_1, \dots, x_N) \in \mathbb{R}^{d \times N}$ and $\boldsymbol{\gamma} = \text{Diag}(\gamma_1, \dots, \gamma_N) \in \mathbb{S}_{d \times N}^{\mathbb{D}}$. Introduce, for $(t, \mathbf{x}, y, \mathbf{z}, \boldsymbol{\gamma}) \in [0, T] \times \mathbb{R}^{d \times N} \times \mathbb{R} \times \mathbb{R}^{d \times N} \times \mathbb{S}_{d \times N}^{\mathbb{D}}$,

$$F^N(t, \mathbf{x}, y, \mathbf{z}, \boldsymbol{\gamma}) := F(t, \mu^N(\mathbf{x}), y, N\mathbf{z} \cdot \mathbf{1}_{\mathbf{x}}, N\boldsymbol{\gamma} \cdot \mathbf{1}_{\mathbf{x}}), \quad (\text{VI.3.1})$$

where $\mathbf{z} \cdot \mathbf{1}_{\mathbf{x}}(x) := \sum_{k=1}^N z_k \mathbf{1}_{x_k}(x)$ and $\boldsymbol{\gamma} \cdot \mathbf{1}_{\mathbf{x}}(x) := \sum_{k=1}^N \gamma_k \mathbf{1}_{x_k}(x)$ for all $x \in \mathbb{R}^d$. We then introduce the PDE on $[0, T] \times \mathbb{R}^{d \times N}$:

$$-\partial_t u(t, \mathbf{x}) - F^N(t, \mathbf{x}, u(t, \mathbf{x}), \partial_{\mathbf{x}} u(t, \mathbf{x}), \partial_{\mathbf{x}\mathbf{x}}^2 u(t, \mathbf{x})) = 0, \quad u|_{t=T} = g^N, \quad (\text{VI.3.2})$$

with $g^N(\mathbf{x}) := g(\mu^N(\mathbf{x}))$, $\partial_{\mathbf{x}} u(t, \mathbf{x}) := (\partial_{x_1} u, \dots, \partial_{x_N} u)(t, \mathbf{x}) \in \mathbb{R}^{d \times N}$ and $\partial_{\mathbf{x}\mathbf{x}}^2 u(t, \mathbf{x}) := \text{Diag}(\partial_{x_1 x_1}^2 u, \dots, \partial_{x_N x_N}^2 u)(t, \mathbf{x}) \in \mathbb{S}_{d \times N}$.

VI.3.2 Viscosity solutions

We define viscosity solutions for the equation (VI.3.2), as in the non-Markovian PDEs. We refer to Ren, Touzi & Zhang [67] for a general overview of viscosity solutions for such equations. Let $\mathbf{X} := (X^1, \dots, X^N)^\top$ be the canonical process on Ω^N and $\mathbb{F}^N = \{\mathcal{F}_t^N\}_{t \in [0, T]}$ the corresponding filtration. For $t \in [0, T]$, define

$$\Lambda_t^N := [t, T] \times \mathbb{R}^{d \times N}, \quad \text{and} \quad \bar{\Lambda}_t^N := [t, T] \times \mathbb{R}^{d \times N}.$$

Definition VI.3.1 For $(t, \mathbf{x}) \in \Lambda_0^N$, let $\mathcal{P}_L^N(t, \mathbf{x})$ be the set of $\mathbb{P} \in \mathcal{P}_2(\Omega^N, \mathcal{F}_T^N)$ such that

- $\mathbf{X}_t = \mathbf{x}$, \mathbb{P} -a.s.,
- there exist $b^{\mathbb{P}}, \sigma^{\mathbb{P}} : [0, T] \times \Omega^N \rightarrow \mathbb{R}^{d \times N}, \mathbb{S}_{d \times N}^{\mathbb{D}}$, \mathbb{F}^N -measurable, uniformly bounded by L , s.t.

$$d\mathbf{X}_s = b_s^{\mathbb{P}} ds + \sigma_s^{\mathbb{P}} dW_s^{\mathbb{P}}, \quad (\text{VI.3.3})$$

where $W^{\mathbb{P}}$ is a $d \times N$ -dimensional \mathbb{P} -Brownian motion.

Lemma VI.3.2 The set $\mathcal{P}_L^N(t, \mathbf{x})$ is weakly compact.

Proof Let $\tilde{\mathcal{P}}_L^N(t, \mathbf{x})$ be defined as $\mathcal{P}_L^N(t, \mathbf{x})$, without requiring that $\sigma^\mathbb{P}$ is blockwise diagonal. Clearly $\mathcal{P}_L^N(t, \mathbf{x}) \subset \tilde{\mathcal{P}}_L^N(t, \mathbf{x})$, and we know from Zheng [79, Theorem 3] that $\tilde{\mathcal{P}}_L^N(t, \mathbf{x})$ is weakly compact. Since $\mathbb{S}_{d \times N}^D$ is closed, $\mathcal{P}_L^N(t, \mathbf{x})$ is closed for the weak convergence, and therefore is weakly compact. \blacksquare

Let $\mathcal{T}_{t,T}^N$ denote the set of $[t, T]$ -valued \mathbb{F}^N -stopping times, and $\mathcal{T}_{t,T}^{N,+} := \{H \in \mathcal{T}_{t,T}^N : H > t\}$. We define the sets of test functions:

$$\begin{aligned} \bar{\mathcal{A}}^N u(t, \mathbf{x}) &:= \left\{ \phi \in C_b^{1,2}(\bar{\Lambda}_t^N) : \exists H \in \mathcal{T}_{t,T}^{N,+} \text{ s.t. } (\phi - u)(t, \mathbf{x}) = \max_{\theta \in \mathcal{T}_{t,T}^N} \bar{\mathcal{E}}_{t,\mathbf{x}}^N \left[(\phi - u)(\theta \wedge H, \mathbf{X}_{\theta \wedge H}) \right] \right\}, \\ \underline{\mathcal{A}}^N u(t, \mathbf{x}) &:= \left\{ \phi \in C_b^{1,2}(\bar{\Lambda}_t^N) : \exists H \in \mathcal{T}_{t,T}^{N,+} \text{ s.t. } (\phi - u)(t, \mathbf{x}) = \min_{\theta \in \mathcal{T}_{t,T}^N} \underline{\mathcal{E}}_{t,\mathbf{x}}^N \left[(\phi - u)(\theta \wedge H, \mathbf{X}_{\theta \wedge H}) \right] \right\}, \end{aligned}$$

where $\bar{\mathcal{E}}^N$ and $\underline{\mathcal{E}}^N$ are the nonlinear expectations defined by

$$\bar{\mathcal{E}}_{t,\mathbf{x}}^N[\cdot] := \sup_{\mathbb{P} \in \mathcal{P}_L^N(t, \mathbf{x})} \mathbb{E}^\mathbb{P}[\cdot], \quad \underline{\mathcal{E}}_{t,\mathbf{x}}^N[\cdot] := \inf_{\mathbb{P} \in \mathcal{P}_L^N(t, \mathbf{x})} \mathbb{E}^\mathbb{P}[\cdot], \quad (\text{VI.3.4})$$

and $C_b^{1,2}(\bar{\Lambda}_t^N)$ denotes the set of bounded functions of $C^{1,2}(\bar{\Lambda}_t^N)$ with bounded derivatives.

Definition VI.3.3 Let $u : \bar{\Lambda}_0^N \rightarrow \mathbb{R}$.

(i) u is a viscosity supersolution of (VI.2.1) if, for all $(t, \mathbf{x}) \in \Lambda_0^N$ and $\phi \in \bar{\mathcal{A}}^N u(t, \mathbf{x})$,

$$-\partial_t \phi(t, \mathbf{x}) - F^N(t, \mathbf{x}, \phi(t, \mathbf{x}), \partial_{\mathbf{x}} \phi(t, \mathbf{x}), \partial_{\mathbf{x}\mathbf{x}}^2 \phi(t, \mathbf{x})) \geq 0. \quad (\text{VI.3.5})$$

(ii) u is a viscosity subsolution of (VI.2.1) if, for all $(t, \mathbf{x}) \in \Lambda_0^N$ and $\phi \in \underline{\mathcal{A}}^N u(t, \mathbf{x})$,

$$-\partial_t \phi(t, \mathbf{x}) - F^N(t, \mathbf{x}, \phi(t, \mathbf{x}), \partial_{\mathbf{x}} \phi(t, \mathbf{x}), \partial_{\mathbf{x}\mathbf{x}}^2 \phi(t, \mathbf{x})) \leq 0. \quad (\text{VI.3.6})$$

(iii) u is a viscosity solution of (VI.2.1) if it is a viscosity supersolution and subsolution.

VI.3.3 Main results

Let \mathcal{S}^N be the set of functions $h : \bar{\Lambda}_0^N \rightarrow \mathbb{R}$ such that $h(t, \mathbf{x}) = h^N(t, \mu^N(\mathbf{x}))$ for some $h^N : [0, T] \times \mathcal{P}^N(\mathbb{R}^d) \rightarrow \mathbb{R}$.

Theorem VI.3.4 (i) Let $\{v^N \in \mathcal{S}^N\}_{N \geq 1}$ be a sequence of continuous and locally bounded, uniformly in N , viscosity supersolutions of (VI.3.2). The relaxed semi-limit defined by

$$\underline{v}(t, \mu) := \liminf_{\substack{N \rightarrow \infty, s \rightarrow t \\ \mu^N(\mathbf{x}^N) \xrightarrow{W_3} \mu}} v^N(s, \mu^N(\mathbf{x}^N)) \quad \text{for all } (t, \mu) \in \bar{\mathbf{Q}}_0$$

is finite and is a LSC viscosity supersolution of (VI.2.1).

(ii) Let $\{u^N \in \mathcal{S}^N\}_{N \geq 1}$ be a sequence of continuous and locally bounded, uniformly in N , viscosity subsolutions of (VI.3.2). The relaxed semi-limit defined by

$$\bar{u}(t, \mu) := \limsup_{\substack{N \rightarrow \infty, s \rightarrow t \\ \mu^N(\mathbf{x}^N) \xrightarrow{W_2} \mu}} u^N(s, \mu^N(\mathbf{x}^N)) \quad \text{for all } (t, \mu) \in \bar{\mathbf{Q}}_0$$

is finite and is a USC viscosity subsolution of (VI.2.1).

The proof of this result is relegated to Section VI.6.

Theorem VI.3.5 Let $\{V^N \in \mathcal{S}^N\}_{N \geq 1}$ be a sequence of continuous and locally bounded, uniformly in N , viscosity solutions of (VI.3.2) s.t. $V^N|_{t=T} = g^N$, and introduce for all $(t, \mu) \in \bar{\mathbf{Q}}_0$

$$\underline{V}(t, \mu) := \liminf_{\substack{N \rightarrow \infty, s \rightarrow t \\ \mu^N(\mathbf{x}^N) \xrightarrow{W_2} \mu}} V^N(s, \mu^N(\mathbf{x}^N)), \quad \bar{V}(t, \mu) := \limsup_{\substack{N \rightarrow \infty, s \rightarrow t \\ \mu^N(\mathbf{x}^N) \xrightarrow{W_2} \mu}} V^N(s, \mu^N(\mathbf{x}^N)). \quad (\text{VI.3.7})$$

If Assumption VI.2.6 holds and $\underline{V}|_{t=T} = \bar{V}|_{t=T} = g$, then V^N converges to the unique continuous viscosity solution V of (VI.2.1), i.e., the following limit exists:

$$V(t, \mu) = \lim_{\substack{N \rightarrow \infty, s \rightarrow t \\ \mu^N(\mathbf{x}^N) \xrightarrow{W_2} \mu}} V^N(s, \mu^N(\mathbf{x}^N)) \quad \text{for all } (t, \mu) \in \bar{\mathbf{Q}}_0$$

and is the unique viscosity solution of (VI.2.1).

Proof By Theorem VI.3.4, \underline{V} and \bar{V} are respectively viscosity supersolution and subsolution of (VI.2.1). As $\underline{V}|_{t=T} = \bar{V}|_{t=T} = g$, the comparison principle implies $\underline{V} \geq \bar{V}$. By (VI.3.7), we also have the converse inequality, and thus $\underline{V} = \bar{V}$, and these functions are viscosity solutions of (VI.2.1). Given the comparison principle, (VI.2.1) has a unique viscosity solution, and thus $\underline{V} = \bar{V} = V$ and the two semi-limits (VI.3.7) are equal to the limit. ■

VI.4 Application to stochastic control

VI.4.1 Mean field control

Let $k \geq 1$, $A \subset \mathbb{R}^k$ and $(b, \sigma) : [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbf{S}) \times A \rightarrow \mathbb{R}^d, \mathbb{S}_d$, continuous in $(t, a) \in [0, T] \times A$, Lipschitz-continuous in $(x, \mu) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$ uniformly in (t, a) and uniformly

CHAPTER VI. A FINITE-DIMENSIONAL APPROXIMATION FOR VISCOSITY
SOLUTIONS OF PARTIAL DIFFERENTIAL EQUATIONS ON WASSERSTEIN SPACE

bounded by L . For $(t, \mu) \in \mathbf{Q}_0$ and $\alpha : [0, T] \times \Omega \longrightarrow A$, let $\mathbb{P}^{t, \mu, \alpha}$ be s.t. X is a controlled McKean-Vlasov diffusion with drift and diffusion coefficients b and σ , i.e.

$$X_s = \xi + \int_t^s b(r, X_r, \mathbb{P}_{X_r}^{t, \mu, \alpha}, \alpha_r) dr + \int_t^s \sigma(r, X_r, \mathbb{P}_{X_r}^{t, \mu, \alpha}, \alpha_r) dW_r^\alpha, \quad \mathbb{P}^{t, \mu, \alpha}\text{-a.s.}, \quad (\text{VI.4.1})$$

where W^α is a standard d -dimensional $\mathbb{P}^{t, \mu, \alpha}$ -Brownian motion and $\mathbb{P}_\xi^{t, \mu, \alpha} = \mu$. Let \mathcal{A}_t be the set of processes $\alpha : [t, T] \times \Omega \longrightarrow A$ s.t. (VI.4.1) has a unique weak solution. We consider the mean field control problem

$$V(t, \mu) := \sup_{\alpha \in \mathcal{A}_t} \mathbb{E}^{\mathbb{P}^{t, \mu, \alpha}} \left[\int_t^T f(r, X_r, \mathbb{P}_{X_r}^{t, \mu, \alpha}, \alpha_r) dr + g(\mathbb{P}_{X_T}^{t, \mu, \alpha}) \right],$$

where f is bounded and continuous in all its variables, and g is Lipschitz-continuous w.r.t. \mathcal{W}_2 . We know from Wu & Zhang [77, Theorem 5.8] that, if V is continuous, then it satisfies the following Hamilton-Jacobi-Bellman (HJB) equation on Wasserstein space:

$$-\partial_t u(t, \mu) - F_{\text{HJB}}(t, \mu, \partial_x \delta_m u(t, \mu, \cdot), \partial_{xx}^2 \delta_m u(t, \mu, \cdot)) = 0, \quad u(T, \cdot) = g, \quad (\text{VI.4.2})$$

where

$$F_{\text{HJB}}(t, \mu, Z, \Gamma) := \left\langle \mu, \sup_{a \in A} \left\{ b(t, \cdot, \mu, a) \cdot Z(\cdot) + \frac{1}{2} \sigma^2(t, \cdot, \mu, a) : \Gamma(\cdot) + f(t, \cdot, \mu, a) \right\} \right\rangle. \quad (\text{VI.4.3})$$

Proposition VI.4.1 *Assume A is compact. Then F_{HJB} satisfies Assumption VI.2.2.*

Proof Let $(Z, \Gamma) \in C_b^0(\mathbb{R}^d, \mathbb{R}^d) \times C_b^0(\mathbb{R}^d, \mathbb{S}_d)$. First observe that the sup in (VI.4.3) is finite as all terms are bounded. Moreover, as b, σ and f are continuous in (t, μ, a) and A is compact, it comes that

$$(t, \mu) \mapsto \sup_{a \in A} \left\{ b(t, x, \mu, a) \cdot Z(x) + \frac{1}{2} \sigma^2(t, x, \mu, a) : \Gamma(x) + f(t, x, \mu, a) \right\}$$

is continuous for all $x \in \mathbb{R}^d$. We then deduce from the boundedness of Z and Γ that F_{HJB} is continuous in (t, μ) . Also, if $(Z', \Gamma') \longrightarrow (Z, \Gamma)$ for $(Z', \Gamma') \in C_b^0(\mathbb{R}^d, \mathbb{R}^d) \times C_b^0(\mathbb{R}^d, \mathbb{S}_d)$, we deduce from the dominated convergence theorem that $F_{\text{HJB}}(t, \mu, Z', \Gamma') \longrightarrow F_{\text{HJB}}(t, \mu, Z, \Gamma)$. This proves Assumption VI.2.2 (i). (ii) is clearly satisfied as F_{HJB} is an integral w.r.t. μ . ■

Remark VI.4.2 We observe that, in the case of F^{HJB} , Z and Γ may belong to $\mathbb{L}^1(\mu)$. However, we chose to restrict them to sets of bounded functions when we introduced the operator F in order to have a more general framework and avoid possible integrability issues.

Finite-dimensional approximation For $(t, \mathbf{x}) \in [0, T] \times \mathbb{R}^{d \times N}$ and a given control $\alpha = (\alpha^1, \dots, \alpha^N) : [0, T] \times \Omega^N \longrightarrow A^N$, let $\mathbb{P}^{t, \mathbf{x}, \alpha}$ be such that, for all $i \in [N] := \{1, \dots, N\}$,

$$X_s^i = x_i + \int_t^s b(r, X_r^i, \mu^N(\mathbf{X}_r), \alpha_r^i) dr + \int_t^s \sigma(r, X_r^i, \mu^N(\mathbf{X}_r), \alpha_r^i) dW_r^{i, \alpha}, \quad \mathbb{P}^{t, \mathbf{x}, \alpha}\text{-a.s.} \quad (\text{VI.4.4})$$

where $W^\alpha := (W^{1, \alpha}, \dots, W^{N, \alpha})^\top$ is a standard $d \times N$ -dimensional $\mathbb{P}^{t, \mathbf{x}, \alpha}$ -Brownian motion. Let \mathcal{A}_t^N be the set of processes $\alpha : [t, T] \times \Omega^N \longrightarrow A^N$ s.t. (VI.4.4) has a unique weak solution. We define the control problem

$$V^N(t, \mathbf{x}) := \sup_{\alpha \in \mathcal{A}_t^N} \sum_{i=1}^N \mathbb{E}^{\mathbb{P}^{t, \mathbf{x}, \alpha}} \left[\int_t^T f^{i, N}(r, \mathbf{X}_r, \alpha_r^i) dr + g^N(\mathbf{X}_T) \right],$$

with $f^{i, N}(t, \mathbf{x}, a) := f(t, x_i, \mu^N(\mathbf{x}), a)$. We know from standard stochastic control theory that, if V^N is continuous, then it is a viscosity solution of

$$-\partial_t u(t, \mathbf{x}) - \sup_{\mathbf{a} \in A^N} \left\{ \mathbf{b}(t, \mathbf{x}, \mathbf{a}) \cdot \partial_{\mathbf{x}} u(t, \mathbf{x}) + \frac{1}{2} \boldsymbol{\sigma}^2(t, \mathbf{x}, \mathbf{a}) : \partial_{\mathbf{x}\mathbf{x}}^2 u(t, \mathbf{x}) + \mathbf{f}(t, \mathbf{x}, \mathbf{a}) \cdot \mathbf{e} \right\} = 0 \quad (\text{VI.4.5})$$

where $\mathbf{b}(t, \mathbf{x}, \mathbf{a}) := \left(b(t, x_i, \mu^N(\mathbf{x}), a_i) \right)_{1 \leq i \leq N}^\top$, $\boldsymbol{\sigma}(t, \mathbf{x}, \mathbf{a}) := \text{Diag} \left(\sigma(t, x_i, \mu^N(\mathbf{x}), a_i) \right)_{1 \leq i \leq N}$, $\mathbf{f}(t, \mathbf{x}, \mathbf{a}) := \left(f^{i, N}(t, \mathbf{x}, a_i) \right)_{1 \leq i \leq N}^\top$ and $\mathbf{e} := (1, \dots, 1)^\top \in \mathbb{R}^N$.

Proposition VI.4.3 *If Assumption VI.2.6 holds for (VI.4.2) and the $\{V^N\}_{N \geq 1}$ are continuous, then V^N converges to V , i.e.,*

$$V(t, \mu) = \lim_{\substack{N \rightarrow \infty, s \rightarrow t \\ \mathcal{W}_2(\mu^N(\mathbf{x}^N), \mu) \rightarrow 0}} V^N(s, \mathbf{x}^N) \quad \text{for all } (t, \mu) \in \mathbf{Q}_0.$$

Proof We first show that (VI.4.5) is the finite-dimensional approximation of (VI.4.2), i.e.,

$$F_{\text{HJB}}^N(t, \mathbf{x}, \mathbf{z}, \gamma) = \sup_{\mathbf{a} \in A^N} \left\{ \mathbf{b}(t, \mathbf{x}, \mathbf{a}) \cdot \mathbf{z} + \frac{1}{2} \boldsymbol{\sigma}^2(t, \mathbf{x}, \mathbf{a}) : \gamma + \mathbf{f}(t, \mathbf{x}, \mathbf{a}) \cdot \mathbf{e} \right\},$$

CHAPTER VI. A FINITE-DIMENSIONAL APPROXIMATION FOR VISCOSITY
SOLUTIONS OF PARTIAL DIFFERENTIAL EQUATIONS ON WASSERSTEIN SPACE

where F_{HJB}^N is the finite-dimensional approximation of F_{HJB} defined by (VI.3.1). We compute

$$\begin{aligned}
F_{\text{HJB}}^N(t, \mathbf{x}, \mathbf{z}, \boldsymbol{\gamma}) &= F_{\text{HJB}}(t, \mu^N(\mathbf{x}), N\mathbf{z} \cdot \mathbf{1}_{\mathbf{x}}, N\boldsymbol{\gamma} \cdot \mathbf{1}_{\mathbf{x}}) \\
&= \left\langle \mu^N(\mathbf{x}), N \sup_{a \in A} \left\{ b(t, \cdot, \mu^N(\mathbf{x}), a) \cdot \mathbf{z} \cdot \mathbf{1}_{\mathbf{x}} + \frac{1}{2} \sigma^2(t, \cdot, \mu^N(\mathbf{x}), a) : \boldsymbol{\gamma} \cdot \mathbf{1}_{\mathbf{x}} \right. \right. \\
&\quad \left. \left. + f(t, \cdot, \mu^N(\mathbf{x}), a) \right\} \right\rangle \\
&= \sum_{i=1}^N \sup_{a \in A} \left\{ b(t, x_i, \mu^N(\mathbf{x}), a) \cdot z_i + \frac{1}{2} \sigma^2(t, x_i, \mu^N(\mathbf{x}), a) : \gamma_i \right. \\
&\quad \left. + f(t, x_i, \mu^N(\mathbf{x}), a) \right\} \\
&= \sup_{\mathbf{a} \in A^N} \sum_{i=1}^N \left\{ b(t, x_i, \mu^N(\mathbf{x}), a_i) \cdot z_i + \frac{1}{2} \sigma^2(t, x_i, \mu^N(\mathbf{x}), a_i) : \gamma_i \right. \\
&\quad \left. + f(t, x_i, \mu^N(\mathbf{x}), a_i) \right\} \\
&= \sup_{\mathbf{a} \in A^N} \left\{ \mathbf{b}(t, \mathbf{x}, \mathbf{a}) \cdot \mathbf{z} + \frac{1}{2} \boldsymbol{\sigma}^2(t, \mathbf{x}, \mathbf{a}) : \boldsymbol{\gamma} + \mathbf{f}(t, \mathbf{x}, \mathbf{a}) \cdot \mathbf{e} \right\}.
\end{aligned}$$

Moreover, since A is compact and g is Lipschitz-continuous, V^N is continuous. Let K_g be a Lipschitz constant for g and fix $\boldsymbol{\alpha} \in \mathcal{A}^N$. Observe that

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}^{t, \mathbf{x}, \boldsymbol{\alpha}}} \left[|g(\mu^N(\mathbf{X}_T)) - g(\mu^N(\mathbf{x}))|^2 \right] &\leq K_g^2 \mathbb{E}^{\mathbb{P}^{t, \mathbf{x}, \boldsymbol{\alpha}}} \left[\mathcal{W}_2^2(\mu^N(\mathbf{X}_T), \mu^N(\mathbf{x})) \right] \\
&\leq \frac{K_g^2}{N} \sum_{k=1}^N \mathbb{E}^{\mathbb{P}^{t, \mathbf{x}, \boldsymbol{\alpha}}} \left[|X_T^k - x_k|^2 \right] \\
&\leq C(T-t)
\end{aligned}$$

with C is independent from N and $\boldsymbol{\alpha}$, due to the uniform boundedness of the drift and diffusion coefficients, and thus

$$\left| V^N(t, \mathbf{x}) - g^N(\mathbf{x}) \right| \leq \sqrt{C(T-t)}. \tag{VI.4.6}$$

As g is Lipschitz-continuous, this provides

$$|V^N(t, \mathbf{x})| \leq C'(1 + \mathcal{W}_2(\mu^N(\mathbf{x}), \mu) + |g(\mu)|) \quad \text{for all } \mu \in \mathcal{P}_2(\mathbb{R}^d) \text{ and some } C' \geq 0,$$

which implies that $\{V^N\}_{N \geq 1}$ is locally bounded, uniformly in N . Moreover, (VI.4.6) also implies that \underline{V} and \bar{V} (defined similarly to (VI.3.7)) satisfy $\underline{V}|_{t=T} = \bar{V}|_{t=T} = g$. Note also that, by symmetry of the problem, $V^N \in \mathcal{S}^N$. We may then apply Theorem VI.3.5 to derive the convergence result. ■

VI.4.2 Zero-sum stochastic differential games

We present in a more informal way a second example. We consider the following control problem, arising in zero-sum games:

$$V_+(t, \mu) := \inf_{\alpha^1 \in \mathcal{A}_t^1} \sup_{\alpha^2 \in \mathcal{A}_t^2} \mathbb{E}^{\mathbb{P}^{t, \mu, \alpha^1, \alpha^2}} \left[\int_t^T f(s, X_s, \mathbb{P}_{X_s}^{t, \mu, \alpha^1, \alpha^2}, \alpha_s^1, \alpha_s^2) ds + g(\mathbb{P}_{X_T}^{t, \mu, \alpha^1, \alpha^2}) \right],$$

where the measures $\{\mathbb{P}^{t, \mu, \alpha^1, \alpha^2} : \alpha_1 \in \mathcal{A}_t^1, \alpha_2 \in \mathcal{A}_t^2\}$ are such that X has the dynamics (VI.4.1), substituting (α^1, α^2) to α . By Cosso & Pham [28], V_+ is a viscosity solution of (VI.2.1), with operator

$$F_+(t, \mu, Z, \Gamma) := \left\langle \mu, \inf_{a_2 \in A_2} \sup_{a_1 \in A_1} \left\{ b(t, \cdot, \mu, a_1, a_2) \cdot Z + \frac{1}{2} \sigma^2(t, \cdot, \mu, a_1, a_2) : \Gamma + f(t, \cdot, \mu, a_1, a_2) \right\} \right\rangle.$$

Although [28] uses another notion of viscosity solutions, we may consider in the context of our discussion that V_+ is also a viscosity solution in the sense of Definition VI.2.4. Assuming A_1 and A_2 are compact, the corresponding finite-dimensional approximation is then given by (VI.3.2), with operator

$$F_+^N(t, \mathbf{x}, \mathbf{z}, \gamma) := \inf_{\mathbf{a}_1 \in A_1^N} \sup_{\mathbf{a}_2 \in A_2^N} \left\{ \mathbf{b}(t, \mathbf{x}, \mathbf{a}_1, \mathbf{a}_2) \cdot \mathbf{z} + \frac{1}{2} \boldsymbol{\sigma}^2(t, \mathbf{x}, \mathbf{a}_1, \mathbf{a}_2) : \gamma + \mathbf{f}(t, \mathbf{x}, \mathbf{a}_1, \mathbf{a}_2) \cdot \mathbf{e} \right\}.$$

As in the case of mean field control, we may show that this corresponds to the PDE satisfied by the control problem

$$V_+^N(t, \mathbf{x}) := \inf_{\boldsymbol{\alpha}^1 \in (\mathcal{A}_t^1)^N} \sup_{\boldsymbol{\alpha}^2 \in (\mathcal{A}_t^2)^N} \sum_{i=1}^N \mathbb{E}^{\mathbb{P}^{t, \mathbf{x}, \boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2}} \left[\int_t^T f^{i, N}(r, \mathbf{X}_r, \boldsymbol{\alpha}_r^1, \boldsymbol{\alpha}_r^2) dr + g(\mathbf{X}_T) \right],$$

where the measures $\{\mathbb{P}^{t, \mathbf{x}, \boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2} : \boldsymbol{\alpha}_1 \in \mathcal{A}_t^1, \boldsymbol{\alpha}_2 \in \mathcal{A}_t^2\}$ are such that \mathbf{X} has the dynamics (VI.4.4), substituting $(\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2)$ to α .

Proposition VI.4.4 *If Assumption VI.2.6 holds for (VI.2.1) with $F = F_+$, then V_+^N converges to V_+ , i.e.,*

$$V_+(t, \mu) = \lim_{\substack{N \rightarrow \infty, s \rightarrow t \\ \mathcal{W}_2(\mu^N(\mathbf{x}^N), \mu) \rightarrow 0}} V_+^N(s, \mathbf{x}^N) \quad \text{for all } (t, \mu) \in \mathbf{Q}_0.$$

VI.4.3 The uncontrolled case

The purpose of this paragraph is to recover the classical propagation of chaos result for diffusion processes, whose first instance was given by Snitzman [72] for some special models. b and σ do no longer depend on the variable a . We consider the equation

$$-\partial_t u(t, \mu) - \left\langle \mu, b(t, \cdot, \mu) \cdot \partial_x \delta_m u(t, \mu, \cdot) + \frac{1}{2} \sigma^2(t, \cdot, \mu) : \partial_{xx}^2 \delta_m u(t, \mu, \cdot) \right\rangle = 0, \quad u(T, \cdot) = g \quad (\text{VI.4.7})$$

where $g \in C_b^0(\mathcal{P}_2(\mathbb{R}^d), \mathbb{R})$, the set of continuous and bounded functions from $\mathcal{P}_2(\mathbb{R}^d)$ to \mathbb{R} . For $N \geq 1$, we easily see that the corresponding finite-dimensional approximation writes

$$-\partial_t u^N(t, \mathbf{x}) - \mathbf{b}(t, \mathbf{x}) \cdot \partial_{\mathbf{x}} u^N(t, \mathbf{x}) - \frac{1}{2} \sigma^2(t, \mathbf{x}) : \partial_{\mathbf{xx}}^2 u^N(t, \mathbf{x}) = 0, \quad u^N(T, \cdot) = g^N. \quad (\text{VI.4.8})$$

For $(t, \mu, \mathbf{x}^N) \in \overline{\mathbf{Q}}_0 \times \mathbb{R}^{d \times N}$, let $(\bar{\mathbb{P}}^{t, \mu}, \bar{\mathbb{P}}^{t, \mathbf{x}^N}) \in \mathcal{P}_L(t, \mu) \times \mathcal{P}_L^N(t, \mathbf{x}^N)$ be such that X and \mathbf{X} are the uncontrolled versions of (VI.4.1) and (VI.4.4), respectively $\bar{\mathbb{P}}^{t, \mu}$ -a.s. and $\bar{\mathbb{P}}^{t, \mathbf{x}^N}$ -a.s. As $g \in C_b^0(\mathcal{P}_2(\mathbb{R}^d), \mathbb{R})$, we know that, under some smoothness assumptions on b and σ (see Talbi, Touzi & Zhang [75, Lemma 3.6]), we have

$$u(t, \mu) = g(\bar{\mathbb{P}}_{X_T}^{t, \mu}), \quad u^N(t, \mathbf{x}^N) = \mathbb{E}^{\bar{\mathbb{P}}^{t, \mathbf{x}^N}} [g^N(\mathbf{X}_T)] = \mathbb{E}^{\bar{\mathbb{P}}^{t, \mathbf{x}^N}} [g(\mu^N(\mathbf{X}_T^N))],$$

for all $(t, \mu) \in \overline{\mathbf{Q}}_0$ and $\mathbf{x}^N \in \mathbb{R}^{d \times N}$. Thus, applying Proposition VI.4.3, we have $u^N(0, \mathbf{x}^N) \rightarrow u(0, \mu)$ as $N \rightarrow \infty$ and $\mathcal{W}_2(\mu^N(\mathbf{x}^N), \mu) \rightarrow 0$, hence

$$\mathbb{E}^{\bar{\mathbb{P}}^{0, \mathbf{x}^N}} [g(\mu^N(\mathbf{X}_T^N))] \xrightarrow{N \rightarrow \infty} g(\bar{\mathbb{P}}_{X_T}^{0, \mu}),$$

which exactly means that $\bar{\mathbb{P}}^{0, \mathbf{x}^N} \circ (\mu^N(\mathbf{X}_T^N))^{-1}$ converges weakly to $\bar{\mathbb{P}}_{X_T}^{0, \mu}$, as it is true for all $g \in C_b^0(\mathcal{P}_2(\mathbb{R}^d), \mathbb{R})$. This corresponds to the propagation of chaos result proved by Oelschläger [63].

VI.5 Extension to path-dependent PDEs

VI.5.1 Pathwise derivatives

For $t \in [0, T)$, we adapt our previous notations to the path-dependent case:

$$\mathbf{Q}_t := [t, T) \times \mathcal{P}_2(\Omega) \quad \text{and} \quad \overline{\mathbf{Q}}_t := [t, T) \times \mathcal{P}_2(\Omega).$$

For $(t, \mu) \in \overline{\mathbf{Q}}_0$, we denote by $\mu_{[0, t]}$ the law of the stopped process $X_{\cdot \wedge t}$ under μ .

Definition VI.5.1 (i) Given a metric space E , we denote by $C^0([0, T] \times \Omega, E)$ the set of \mathbb{F} -progressively measurable and continuous functions from $[0, T] \times \Omega$ to E , where Ω is equipped with the norm $|\omega| := \sup_{t \in [0, T]} |\omega_t|$ for all $\omega \in \Omega$.

(ii) We denote by $u \in C^{1,2}([0, T] \times \Omega)$ the set of functions $u : [0, T] \times \Omega \rightarrow \mathbb{R}$ such that there exist $\partial_t u \in C^0([0, T] \times \Omega, \mathbb{R})$, $\partial_\omega u \in C^0([0, T] \times \Omega, \mathbb{R}^d)$ and $\partial_{\omega\omega}^2 u \in C^0([0, T] \times \Omega, \mathbb{S}_d)$ such that, for all $\mathbb{P} \in \bigcup_{L>0} \mathcal{P}_L$, u satisfies

$$du(t, X) = \partial_t u(t, X) dt + \partial_\omega u(t, X) \cdot dX_t + \frac{1}{2} \partial_{\omega\omega}^2 u(t, X) : d\langle X \rangle_t, \quad \mathbb{P}\text{-a.s.}$$

Definition VI.5.2 Fix $t \in [0, T]$.

(i) We denote by $C^0(\overline{\mathbf{Q}}_t)$ the set of functions $u : \overline{\mathbf{Q}}_t \rightarrow \mathbb{R}$ continuous for the pseudo-metric:

$$\widetilde{\mathcal{W}}_2((s, \mu), (r, \nu)) := \left(|s - r|^2 + \mathcal{W}_2^2(\mu_{[0, s]}, \nu_{[0, r]}) \right)^{\frac{1}{2}} \quad \text{for all } (s, \mu), (r, \nu) \in \overline{\mathbf{Q}}_t. \quad (\text{VI.5.1})$$

(ii) We denote by $C_b^{1,2}(\overline{\mathbf{Q}}_t)$ the set of bounded functions $u : \overline{\mathbf{Q}}_t \rightarrow \mathbb{R}$ such that $\partial_t u$, $\delta_m u$, $\partial_\omega \delta_m u$, $\partial_{\omega\omega}^2 \delta_m u$ exist, are bounded in all their variables and continuous in (t, μ) in the sense of (i), where the functional linear derivative takes the form $\delta_m u : [t, T] \times \mathcal{P}_2(\Omega) \times \Omega \rightarrow \mathbb{R}$ satisfying, for any $s \in [t, T]$ and $\mu, \mu' \in \mathcal{P}_2(\Omega)$,

$$u(s, \mu') - u(s, \mu) = \int_0^1 \int_\Omega \delta_m u(s, \lambda \mu' + (1 - \lambda) \mu, \omega) (\mu' - \mu)(d\omega) d\lambda.$$

Note that, if $u \in C^0(\overline{\mathbf{Q}}_t)$, then $u(s, \mu) = u(s, \mu_{[0, s]})$ for all $(s, \mu) \in \overline{\mathbf{Q}}_t$.

VI.5.2 Path-dependent equation on Wasserstein space

Let $F : [0, T] \times \mathcal{P}_2(\Omega) \times \mathbb{R} \times \mathbb{L}_2^0(\Omega, \mathbb{R}^d) \times \mathbb{L}_2^0(\Omega, \mathbb{S}_d) \rightarrow \mathbb{R}$, where $\mathbb{L}_2^0(\Omega, \mathbb{R}^d)$ (resp. $\mathbb{L}_2^0(\Omega, \mathbb{S}_d)$) denotes the set of Borel measurable functions from Ω to \mathbb{R}^d (resp. \mathbb{S}_d) with quadratic growth, and $g : \mathcal{P}_2(\Omega) \rightarrow \mathbb{R}$. We consider the following equation:

$$-\partial_t u(t, \mu) - F(t, \mu, u(t, \mu), \partial_\omega \delta_m u(t, \mu, \cdot), \partial_{\omega\omega}^2 \delta_m u(t, \mu, \cdot)) = 0, \quad u|_{t=T} = g, \quad (\text{VI.5.2})$$

for all $(t, \mu) \in \mathbf{Q}_0$. We define semijets similarly to (VI.2.2) and straightforwardly adapt Proposition VI.2.5 and Assumption VI.2.2 to the path-dependent setting.

VI.5.3 Viscosity solutions

We redefine, for all $(t, \mu) \in \mathbf{Q}_0$, $\mathcal{P}_L(t, \mu) := \{\mathbb{P} \in \mathcal{P}_L : \mathbb{P}_{X_{t \wedge \cdot}} = \mu_{[0, t]}\}$, as well as the neighborhood

$$\mathcal{N}_\delta(t, \mu) := [t, t + \delta] \times \mathcal{P}_L(t, \mu),$$

which is compact under $\widetilde{\mathcal{W}}_2$ (see again Wu & Zhang [77, Lemma 4.1]). We then introduce the sets of test functions:

$$\begin{aligned} \overline{\mathcal{A}}u(t, \mu) &:= \left\{ \varphi \in C_b^{1,2}(\overline{\mathbf{Q}}_t) : (\varphi - u)(t, \mu) = \max_{\mathcal{N}_\delta(t, \mu)} (\varphi - u) \text{ for some } \delta > 0 \right\}, \\ \underline{\mathcal{A}}u(t, \mu) &:= \left\{ \varphi \in C_b^{1,2}(\overline{\mathbf{Q}}_t) : (\varphi - u)(t, \mu) = \min_{\mathcal{N}_\delta(t, \mu)} (\varphi - u) \text{ for some } \delta > 0 \right\}. \end{aligned}$$

Definition VI.5.3 *Let $u : \overline{\mathbf{Q}}_0 \rightarrow \mathbb{R}$.*

(i) *u is a viscosity supersolution of (VI.5.2) if, for all $(t, \mu) \in \mathbf{Q}_0$ and $\varphi \in \overline{\mathcal{A}}u(t, m)$,*

$$-\partial_t \varphi(t, \mu) - F(t, \mu, u(t, \mu), \partial_\omega \delta_m \varphi(t, \mu, \cdot), \partial_{\omega\omega}^2 \delta_m \varphi(t, \mu, \cdot)) \geq 0.$$

(ii) *u is a viscosity subsolution of (VI.5.2) if, for all $(t, \mu) \in \mathbf{Q}_0$ and $\varphi \in \underline{\mathcal{A}}u(t, m)$,*

$$-\partial_t \varphi(t, \mu) - F(t, \mu, u(t, \mu), \partial_\omega \delta_m \varphi(t, \mu, \cdot), \partial_{\omega\omega}^2 \delta_m \varphi(t, \mu, \cdot)) \leq 0.$$

(iii) *u is a viscosity solution of (VI.5.2) if it is a viscosity supersolution and subsolution.*

VI.5.4 Finite-dimensional approximation

Let $N \geq 1$. We shall write in bold character the elements $\boldsymbol{\omega} = (\omega^1, \dots, \omega^N) \in \Omega^N$. Introduce, for $(t, \boldsymbol{\omega}, y, \mathbf{z}, \boldsymbol{\gamma}) \in [0, T] \times \Omega^N \times \mathbb{R} \times \mathbb{R}^{d \times N} \times \mathbb{S}_{d \times N}^D$,

$$F^N(t, \boldsymbol{\omega}, y, \mathbf{z}, \boldsymbol{\gamma}) := F(t, \mu^N(\boldsymbol{\omega}), y, N\mathbf{z} \cdot \mathbf{1}_\omega, N\boldsymbol{\gamma} \cdot \mathbf{1}_\omega), \quad (\text{VI.5.3})$$

where $\mathbf{z} \cdot \mathbf{1}_\omega(\boldsymbol{\omega}) := \sum_{k=1}^N z_k \mathbf{1}_{\omega^k}(\boldsymbol{\omega})$ and $\boldsymbol{\gamma} \cdot \mathbf{1}_\omega(\boldsymbol{\omega}) := \sum_{k=1}^N \gamma_k \mathbf{1}_{\omega^k}(\boldsymbol{\omega})$ for all $\boldsymbol{\omega} \in \Omega$. We then introduce the path-dependent PDE on $[0, T] \times \Omega^N$:

$$-\partial_t u(t, \boldsymbol{\omega}) - F^N(t, \boldsymbol{\omega}, u(t, \boldsymbol{\omega}), \partial_\omega u(t, \boldsymbol{\omega}), \partial_{\omega\omega}^2 u(t, \boldsymbol{\omega})) = 0, \quad u|_{t=T} = g^N, \quad (\text{VI.5.4})$$

with $g^N(\boldsymbol{\omega}) := g(\mu^N(\boldsymbol{\omega}))$, $\partial_\omega u(t, \boldsymbol{\omega}) := (\partial_{\omega^1} u, \dots, \partial_{\omega^N} u)(t, \boldsymbol{\omega}) \in \mathbb{R}^{d \times N}$ and $\partial_{\omega\omega}^2 u(t, \boldsymbol{\omega}) := \text{Diag}(\partial_{\omega^1 \omega^1}^2 u, \dots, \partial_{\omega^N \omega^N}^2 u)(t, \boldsymbol{\omega}) \in \mathbb{S}_{d \times N}$.

CHAPTER VI. A FINITE-DIMENSIONAL APPROXIMATION FOR VISCOSITY SOLUTIONS OF PARTIAL DIFFERENTIAL EQUATIONS ON WASSERSTEIN SPACE

We now define viscosity solutions for (VI.5.4). We adapt the notations of Section VI.3.2 to the path-dependent case. For $t \in [0, T)$, define

$$\Lambda_t^N := [t, T) \times \Omega^N, \quad \text{and} \quad \bar{\Lambda}_t^N := [t, T) \times \Omega^N.$$

For $(t, \omega) \in \Lambda_0^N$, we define $\mathcal{P}_L^N(t, \omega)$ similarly to $\mathcal{P}_L^N(t, \mathbf{x})$, with the condition $\mathbf{X}_{t\wedge\cdot} = \omega_{t\wedge\cdot}$, \mathbb{P} -a.s., for all $\mathbb{P} \in \mathcal{P}_L^N(t, \omega)$. Similarly to Lemma VI.3.2, we have the following result:

Lemma VI.5.4 *The set $\mathcal{P}_L^N(t, \omega)$ is weakly compact.*

We define the sets of test functions:

$$\begin{aligned} \bar{\mathcal{A}}^N u(t, \omega) &:= \left\{ \phi \in C_b^{1,2}(\bar{\Lambda}_0^N) : \exists H \in \mathcal{T}_{t,T}^{N,+} \text{ s.t. } (\phi - u)(t, \omega) = \max_{\theta \in \mathcal{T}_{t,T}^{N,+}} \bar{\mathcal{E}}_{t,\omega}^N \left[(\phi - u)(\theta \wedge H, \mathbf{X}_{\cdot \wedge \theta \wedge H}) \right] \right\}, \\ \underline{\mathcal{A}}^N u(t, \omega) &:= \left\{ \phi \in C_b^{1,2}(\bar{\Lambda}_0^N) : \exists H \in \mathcal{T}_{t,T}^{N,+} \text{ s.t. } (\phi - u)(t, \omega) = \min_{\theta \in \mathcal{T}_{t,T}^{N,+}} \underline{\mathcal{E}}_{t,\omega}^N \left[(\phi - u)(\theta \wedge H, \mathbf{X}_{\cdot \wedge \theta \wedge H}) \right] \right\}, \end{aligned}$$

where $\bar{\mathcal{E}}^N$ and $\underline{\mathcal{E}}^N$ are the nonlinear expectations defined by

$$\bar{\mathcal{E}}_{t,\omega}^N[\cdot] := \sup_{\mathbb{P} \in \mathcal{P}_L^N(t,\omega)} \mathbb{E}^{\mathbb{P}}[\cdot], \quad \underline{\mathcal{E}}_{t,\omega}^N[\cdot] := \inf_{\mathbb{P} \in \mathcal{P}_L^N(t,\omega)} \mathbb{E}^{\mathbb{P}}[\cdot], \quad (\text{VI.5.5})$$

and $C_b^{1,2}(\bar{\Lambda}_t^N)$ denotes the bounded elements of $C^{1,2}(\bar{\Lambda}_t^N)$ (defined similarly to $C^{1,2}([0, T] \times \Omega)$) with bounded derivatives.

Definition VI.5.5 *Let $u : \bar{\Lambda}_0^N \rightarrow \mathbb{R}$.*

(i) *u is a viscosity supersolution of (VI.2.1) if, for all $(t, \omega) \in \Lambda_0^N$ and $\phi \in \bar{\mathcal{A}}^N u(t, \omega)$,*

$$-\partial_t \phi(t, \omega) - F^N(t, \omega, \phi(t, \omega), \partial_\omega \phi(t, \omega), \partial_{\omega\omega}^2 \phi(t, \omega)) \geq 0.$$

(ii) *u is a viscosity subsolution of (VI.2.1) if, for all $(t, \omega) \in \Lambda_0^N$ and $\phi \in \underline{\mathcal{A}}^N u(t, \omega)$,*

$$-\partial_t \phi(t, \omega) - F^N(t, \omega, \phi(t, \omega), \partial_\omega \phi(t, \omega), \partial_{\omega\omega}^2 \phi(t, \omega)) \leq 0.$$

(iii) *u is a viscosity solution of (VI.2.1) if it is a viscosity supersolution and subsolution.*

In this paragraph, \mathcal{S}^N denotes the set of functions $h : \bar{\Lambda}_0^N \rightarrow \mathbb{R}$ s.t. $h(t, \omega) = h^N(t, \mu^N(\omega))$ for some $h^N : [0, T] \times \mathcal{P}^N(\Omega) \rightarrow \mathbb{R}$.

Theorem VI.5.6 (i) Let $\{v^N \in \mathcal{S}^N\}_{N \geq 1}$ be a sequence of uniformly continuous for (VI.5.1) and locally bounded, uniformly in N , viscosity supersolutions of (VI.5.4). Then, the relaxed semi-limit defined by

$$\underline{v}(t, \mu) := \liminf_{\substack{N \rightarrow \infty, s \rightarrow t \\ \mu^N(\omega^N) \xrightarrow{\mathcal{W}_2} \mu}} v^N(s, \mu^N(\omega^N)) \quad \text{for all } (t, \mu) \in \mathbf{Q}_0$$

is finite and is a LSC viscosity supersolution of (VI.2.1).

(ii) Let $\{u^N \in \mathcal{S}^N\}_{N \geq 1}$ be a sequence of uniformly continuous for (VI.5.1) and locally bounded, uniformly in N , viscosity subsolutions of (VI.5.4). Then, the relaxed semi-limit defined by

$$\bar{u}(t, \mu) := \limsup_{\substack{N \rightarrow \infty, s \rightarrow t \\ \mu^N(\omega^N) \xrightarrow{\mathcal{W}_2} \mu}} u^N(s, \mu^N(\omega^N)) \quad \text{for all } (t, \mu) \in \mathbf{Q}_0$$

is finite and is a USC viscosity subsolution of (VI.2.1).

The proof of this result is relegated to Section VI.6. Similarly to Theorem VI.3.5, this implies the following result:

Theorem VI.5.7 Let $\{V^N \in \mathcal{S}^N\}_{N \geq 1}$ be a sequence of uniformly continuous for (VI.5.1) and locally bounded, uniformly in N , viscosity solutions of (VI.5.4) s.t. $V^N(T, \cdot) = g^N$, and introduce

$$\underline{V}(t, \mu) := \liminf_{\substack{N \rightarrow \infty, s \rightarrow t \\ \mu^N(\omega^N) \xrightarrow{\mathcal{W}_2} \mu}} V^N(s, \mu^N(\omega^N)), \quad \bar{V}(t, \mu) := \limsup_{\substack{N \rightarrow \infty, s \rightarrow t \\ \mu^N(\omega^N) \xrightarrow{\mathcal{W}_2} \mu}} V^N(s, \mu^N(\omega^N)) \quad (\text{VI.5.6})$$

If Assumption VI.2.6 holds and $\underline{V}|_{t=T} = \bar{V}|_{t=T} = g$, then u^N converges to the unique continuous viscosity solution V of (VI.5.2), i.e.,

$$V(t, \mu) = \lim_{\substack{N \rightarrow \infty, s \rightarrow t \\ \mu^N(\omega^N) \xrightarrow{\mathcal{W}_2} \mu}} V^N(s, \omega^N) \quad \text{for all } (t, \mu) \in \mathbf{Q}_0.$$

VI.6 Proof of Theorems VI.3.4 and VI.5.6

VI.6.1 The Markovian setting

We first prove Theorem VI.3.4. We only prove the convergence of the viscosity supersolutions, as the case of the subsolutions is handled similarly. It is clear that \underline{v} is finite as

CHAPTER VI. A FINITE-DIMENSIONAL APPROXIMATION FOR VISCOSITY
SOLUTIONS OF PARTIAL DIFFERENTIAL EQUATIONS ON WASSERSTEIN SPACE

$\{v^N\}_{N \geq 1}$ is locally bounded, uniformly in N . Fix $(t, \mu) \in \mathbf{Q}_0$ and $\varphi \in \overline{\mathcal{A}}\underline{v}(t, \mu)$ with corresponding $\delta_0 \in (0, T - t)$. By Proposition VI.2.5, we may assume w.l.o.g. that φ is a semijet of the form (VI.2.2), with characteristics $(v, a, f) \in \mathbb{R} \times \mathbb{R} \times C_b^2(\mathbb{R}^d)$. Replacing φ with $\tilde{\varphi}(s, \cdot) := \varphi(s, \cdot) - (s - t)^2$, we may also assume w.l.o.g. that (t, μ) is a strict maximum of $(\varphi - \underline{v})$ on $\mathcal{N}_{\delta_0}(t, \mu)$.

Step 1: Let (t^N, \mathbf{x}^N) be a sequence such that $t^N \rightarrow t$, $\mu^N(\mathbf{x}^N) \xrightarrow{\mathcal{W}_2} \mu$, and $v^N(t^N, \mu^N(\mathbf{x}^N)) \rightarrow \underline{v}(t, \mu)$ as $N \rightarrow \infty$. We also introduce, for all $N \geq 1$, the functions $\phi^N(s, \mathbf{x}) := \varphi(s, \mu^N(\mathbf{x}))$ for all $(s, \mathbf{x}) \in \overline{\Lambda}_0^N$. Fix $\delta \in (0, \delta_0)$, and introduce the stopping time

$$H_\delta^N := \inf \{s \geq t^N : \mathcal{W}_2(\mu^N(\mathbf{X}_s), \mu^N(\mathbf{x}^N)) = \delta\} \wedge (t^N + \delta).$$

By Ekren, Touzi & Zhang [38, Theorem 3.5] and weak compactness of $\mathcal{P}_L^N(t, \mathbf{x}^N)$, there exists $(\theta_\delta^N, \mathbb{P}^{N,*}) \in \mathcal{T}_{t^N, T}^N \times \mathcal{P}_L^N(t, \mathbf{x}^N)$ s.t.

$$\mathbb{E}^{\mathbb{P}^{N,*}} \left[(\phi^N - v^N)(\theta_\delta^N \wedge H_\delta^N, \mathbf{X}_{\theta_\delta^N \wedge H_\delta^N}) \right] = \sup_{\theta \in \mathcal{T}_{t^N, T}^N} \overline{\mathcal{E}}_{t^N, \mathbf{x}^N}^N \left[(\phi^N - v^N)(\theta \wedge H_\delta^N, \mathbf{X}_{\theta \wedge H_\delta^N}) \right] \quad \text{(VI.6.1)}$$

where $\overline{\mathcal{E}}_{t^N, \mathbf{x}^N}^N$ is defined by (VI.3.4). Indeed, by continuity of $\phi^N - v^N$, we easily see that the Markov process $s \mapsto (\phi^N - v^N)(s, \mathbf{X}_s)$ is bounded and uniformly continuous on $\{(t, \omega) : t \leq H_\delta^N(\omega)\}$, and that $\mathcal{P}_L^N(t, \mathbf{x}^N)$ satisfies [38, Assumption 3.4]. Also, note that, since $\{v^N\}_{N \geq 1}$ is locally bounded, uniformly in N , and $\mu^N(\mathbf{x}^N) \xrightarrow{\mathcal{W}_2} \mu$, we may assume w.l.o.g. that, after passing to an appropriate subsequence and for δ small enough, $\{(\phi^N - v^N)(\theta \wedge H_\delta^N, \mathbf{X}_{\theta \wedge H_\delta^N})\}_{N \geq 1}$ is uniformly bounded for all $\theta \in \mathcal{T}_{t^N, T}^N$.

Step 2: We now justify that

$$\limsup_{N \rightarrow \infty} \mathbb{P}^{N,*}(\theta_\delta^N < H_\delta^N) > 0. \quad \text{(VI.6.2)}$$

Indeed, assume to the contrary that $\limsup_{N \rightarrow \infty} \mathbb{P}^{N,*}(\theta_\delta^N < H_\delta^N) = \lim_{N \rightarrow \infty} \mathbb{P}^{N,*}(\theta_\delta^N < H_\delta^N) = 0$.

We have:

$$\begin{aligned}
(\varphi - \underline{v})(t, \mu) &= \lim_{N \rightarrow \infty} (\varphi - v^N)(t^N, \mu^N(\mathbf{x}^N)) \\
&\leq \liminf_{N \rightarrow \infty} \mathbb{E}^{\mathbb{P}^{N,*}} \left[(\varphi - v^N)(\theta_\delta^N \wedge H_\delta^N, \mu^N(\mathbf{X}_{\theta_\delta^N \wedge H_\delta^N})) \right] \\
&\leq \limsup_{N \rightarrow \infty} \mathbb{E}^{\mathbb{P}^{N,*}} \left[(\varphi - v^N)(\theta_\delta^N \wedge H_\delta^N, \mu^N(\mathbf{X}_{\theta_\delta^N \wedge H_\delta^N})) \right] \\
&= \limsup_{N \rightarrow \infty} \mathbb{E}^{\mathbb{P}^{N,*}} \left[\left\{ (\varphi - v^N)(H_\delta^N, \mu^N(\mathbf{X}_{H_\delta^N})) (1 - \mathbf{1}_{\theta_\delta^N < H_\delta^N}) \right. \right. \\
&\quad \left. \left. + (\varphi - v^N)(\theta_\delta^N, \mu^N(\mathbf{X}_{\theta_\delta^N})) \mathbf{1}_{\theta_\delta^N < H_\delta^N} \right\} \right] \\
&= \limsup_{N \rightarrow \infty} \mathbb{E}^{\mathbb{P}^{N,*}} \left[(\varphi - v^N)(H_\delta^N, \mu^N(\mathbf{X}_{H_\delta^N})) \right],
\end{aligned}$$

where we used the fact that $\{(\varphi - v^N)(\theta_\delta^N \wedge H_\delta^N, \mu^N(\mathbf{X}_{\theta_\delta^N \wedge H_\delta^N}))\}_{N \geq 1}$ is uniformly bounded and $\limsup_{N \rightarrow \infty} \mathbb{P}^{N,*}(\theta_\delta^N < H_\delta^N) = \lim_{N \rightarrow \infty} \mathbb{P}^{N,*}(\theta_\delta^N < H_\delta^N) = 0$.

Since $(t^N, \mu^N(\mathbf{x}^N)) \xrightarrow{N \rightarrow \infty} (t, \mu)$, by Proposition VI.7.1 and compactness of $[t, T]$, there exists a subsequence $\nu^N := \mathbb{P}^{N,*} \circ (H_\delta^N, \mu^N(\mathbf{X}))^{-1}$ that converges weakly to some $\nu \in \mathcal{P}_2([t, T] \times \mathcal{P}_2(\Omega))$ supported on $[t, t + \delta] \times \mathcal{P}_L(t, \mu)$. Thus, denoting by (τ, m) the canonical mapping on $[0, T] \times \mathcal{P}_2(\Omega)$, we have by upper semicontinuity of $\varphi - v^N$ and continuity of $(\tau, m) \mapsto m_{X_\tau}$,

$$\begin{aligned}
\limsup_{N \rightarrow \infty} \mathbb{E}^{\mathbb{P}^{N,*}} \left[(\varphi - v^N)(H_\delta^N, \mu^N(\mathbf{X}_{H_\delta^N})) \right] &= \limsup_{N \rightarrow \infty} \mathbb{E}^{\nu^N} \left[(\varphi - v^N)(\tau, m) \right] \\
&\leq \mathbb{E}^\nu \left[(\varphi - \underline{v})(\tau, m_{X_\tau}) \right] \\
&\leq (\varphi - \underline{v})(\tau(\bar{\omega}), m_{X_{\tau(\bar{\omega})}}(\bar{\omega})),
\end{aligned}$$

for some $\bar{\omega} \in [0, T] \times \mathcal{P}_2(\Omega)$. We also observe that $\tau > t$, ν -a.s. Indeed, given that, by definition of H_δ^N , we have

$$\mathcal{W}_2(m_{X_\tau}, \mu^N(\mathbf{x}^N)) = \delta \text{ or } \tau = t^N + \delta, \quad \nu^N\text{-a.s.},$$

for all $N \geq 1$, we have, as \underline{v} is LSC,

$$\mathcal{W}_2(\mu, m_{X_\tau}) = \delta \text{ or } \tau = t + \delta, \quad \nu\text{-a.s.}$$

As $\nu \circ m^{-1}$ is supported on $\mathcal{P}_L(t, \mu)$, we have $m_{X_t} = \mu$, ν -a.s. Thus the above inequalities imply that $\tau > t$, ν -a.s. $\bar{\omega}$ may then be chosen s.t. $\tau(\bar{\omega}) > t$, and we have $(\varphi - \underline{v})(t, \mu) \leq (\varphi - \underline{v})(\tau(\bar{\omega}), m_{X_{\tau(\bar{\omega})}}(\bar{\omega}))$, which contradicts the fact that (t, μ) is a strict maximum on

CHAPTER VI. A FINITE-DIMENSIONAL APPROXIMATION FOR VISCOSITY
SOLUTIONS OF PARTIAL DIFFERENTIAL EQUATIONS ON WASSERSTEIN SPACE

$\mathcal{N}_\delta(t, \mu)$. Thus, (VI.6.2) holds true, and we may find a subsequence $\{\omega^{\delta, N}\}_{N \geq 1}$ s.t. $t_\delta^N := \theta_\delta^N(\omega^{\delta, N}) < H_\delta^N(\omega^{\delta, N})$ for all $N \geq 1$.

Step 3: We now prove that ϕ^N is a test function for v^N in some well chosen point. Introduce $\mathbf{x}^{\delta, N} := \mathbf{X}_{\theta_\delta^N(\omega^{\delta, N})}(\omega^{\delta, N})$ and \mathbf{Y}^N , the nonlinear Snell envelop of $s \mapsto (\phi^N - v^N)(s, \mathbf{X}_s)$, i.e.,

$$\mathbf{Y}_s^N(\omega) := \sup_{\theta \in \mathcal{T}_{t, T}^N} \bar{\mathcal{E}}_{s, \omega}^N \left[(\phi^N - v^N)(\theta \wedge H_\delta^N, \mathbf{X}_{\theta \wedge H_\delta^N}) \right],$$

which satisfies $\mathbf{Y}_s \geq \bar{\mathcal{E}}_{s, \mathbf{Y}_s}^N[\mathbf{Y}_{\theta \wedge H_\delta^N}]$ for all $\theta \in \mathcal{T}_{s, t}^N$. Then we have, for all $\theta \in \mathcal{T}_{t_\delta^N, T}^N$,

$$\begin{aligned} (\phi^N - v^N)(t_\delta^N, \mathbf{x}^{\delta, N}) &= \mathbf{Y}_{t_\delta^N}^N(\omega^{\delta, N}) \\ &\geq \bar{\mathcal{E}}_{t_\delta^N, \omega^{\delta, N}}^N[\mathbf{Y}_{\theta \wedge H_\delta^N}^N] \\ &\geq \bar{\mathcal{E}}_{t_\delta^N, \omega^{\delta, N}}^N[(\phi^N - v^N)(\theta \wedge H_\delta^N, \mathbf{X}_{\theta \wedge H_\delta^N})], \end{aligned}$$

and therefore, as we are in the Markovian case,

$$(\phi^N - v^N)(t_\delta^N, \mathbf{x}^{\delta, N}) = \max_{\theta \in \mathcal{T}_{t_\delta^N, T}^N} \bar{\mathcal{E}}_{t_\delta^N, \mathbf{x}^{\delta, N}}^N \left[(\phi^N - v^N)(\theta \wedge H_\delta^N, \mathbf{X}_{\theta \wedge H_\delta^N}) \right].$$

Observe that $\phi^N \in C_b^{1,2}(\bar{\Lambda}_{t_\delta^N}^N)$. Indeed, since $\varphi = \psi^{v, a, f}$ is a semijet, we have

$$\begin{aligned} \partial_t \phi^N(s, \mathbf{x}) &= \partial_t \varphi(s, \mu^N(\mathbf{x})) = a, \\ \partial_{x_i} \phi^N(s, \mathbf{x}) &= \frac{1}{N} \partial_x f(x_i) = \frac{1}{N} \partial_x \delta_m \varphi(s, \mu^N(\mathbf{x}), x_i), \\ \partial_{x_i x_i}^2 \phi^N(s, \mathbf{x}) &= \frac{1}{N} \partial_{xx}^2 f(x_i) = \frac{1}{N} \partial_{xx}^2 \delta_m \varphi(s, \mu^N(\mathbf{x}), x_i), \end{aligned} \tag{VI.6.3}$$

for all $i \in [N]$. As $H_\delta^N > t_\delta^N$ on $\{\mathbf{X}_{\cdot \wedge t_\delta^N} = \omega_{\cdot \wedge t_\delta^N}^{\delta, N}\}$, we have $\phi^N \in \bar{\mathcal{A}}^N v^N(t_\delta^N, \mathbf{x}^{\delta, N})$ and the supersolution property provides

$$-\partial_t \phi^N(q_\delta^N) - F^N(q_\delta^N, v^N(q_\delta^N), \partial_{\mathbf{x}} \phi^N(q_\delta^N), \partial_{\mathbf{xx}}^2 \phi^N(q_\delta^N)) \geq 0. \tag{VI.6.4}$$

where $q_\delta^N := (t_\delta^N, \mathbf{x}^{\delta, N})$.

Step 4: We finally derive the viscosity supersolution of \underline{v} . By (VI.3.1) and (VI.6.5), we

have:

$$\begin{aligned}
& F^N(q_\delta^N, v^N(q_\delta^N), \partial_{\mathbf{x}}\phi^N(q_\delta^N), \partial_{\mathbf{xx}}^2\phi^N(q_\delta^N)) \\
&= F\left(t_\delta^N, \mu^N(\mathbf{x}^{\delta,N}), v^N(t_\delta^N, \mu^N(\mathbf{x}^{\delta,N})), \sum_{k=1}^N \partial_x \delta_m \varphi(t_\delta^N, \mu^N(\mathbf{x}^{\delta,N}), x_k^{\delta,N}) \mathbf{1}_{x_k^{\delta,N}}, \right. \\
&\qquad\qquad\qquad \left. \sum_{k=1}^N \partial_{xx}^2 \delta_m \varphi(t_\delta^N, \mu^N(\mathbf{x}^{\delta,N}), x_k^{\delta,N}) \mathbf{1}_{x_k^{\delta,N}}\right) \\
&= F\left(t_\delta^N, \mu^N(\mathbf{x}^{\delta,N}), v^N(t_\delta^N, \mu^N(\mathbf{x}^{\delta,N})), \partial_x \delta_m \varphi(t_\delta^N, \mu^N(\mathbf{x}^{\delta,N}), \cdot), \partial_{xx}^2 \delta_m \varphi(t_\delta^N, \mu^N(\mathbf{x}^{\delta,N}), \cdot)\right).
\end{aligned}$$

The second equality is a consequence of Assumption VI.2.2 (ii). Note that

$$(t_\delta^N, \mu^N(\mathbf{x}^{\delta,N}), v^N(t_\delta^N, \mu^N(\mathbf{x}^{\delta,N}))) \xrightarrow{\delta \rightarrow 0} (t^N, \mu^N(\mathbf{x}^N), v^N(t^N, \mu^N(\mathbf{x}^N)))$$

by definition of H_δ^N and continuity of v^N , and thus

$$(t_\delta^N, \mu^N(\mathbf{x}^{\delta,N}), v^N(t_\delta^N, \mu^N(\mathbf{x}^{\delta,N}))) \xrightarrow{(\delta, N) \rightarrow (0, \infty)} (t, \mu, \underline{v}(t, \mu)).$$

We then deduce from the smoothness of φ and the continuity of F (in the sense of Assumption VI.2.2 (i)) that

$$F^N(q_\delta^N, v^N(q_\delta^N), \partial_{\mathbf{x}}\phi^N(q_\delta^N), \partial_{\mathbf{xx}}^2\phi^N(q_\delta^N)) \xrightarrow[\substack{\delta \rightarrow 0 \\ N \rightarrow \infty}]{} F(t, \mu, \underline{v}(t, \mu), \partial_x \delta_m \varphi(t, \mu, \cdot), \partial_{xx}^2 \delta_m \varphi(t, \mu, \cdot)).$$

Finally, as $\partial_t \phi^N(q_\delta^N) = a = \partial_t \varphi(t, \mu)$, sending $(\delta, N) \rightarrow (0, \infty)$ in (VI.6.5) provides the viscosity supersolution property of \underline{v} .

VI.6.2 The path-dependent setting

The proof of Theorem VI.5.6 follows the same arguments. We use the notations of Section VI.5. Let $(t^N, \boldsymbol{\omega}^N)$ be a sequence such that $(t^N, \mu^N(\boldsymbol{\omega}^N)) \rightarrow (t, \mu)$ and $v^N(t^N, \mu^N(\boldsymbol{\omega}^N)) \rightarrow \underline{v}(t, \mu)$ as $N \rightarrow \infty$. We also introduce, for all $N \geq 1$, the functions $\phi^N(s, \boldsymbol{\omega}) := \varphi(s, \mu^N(\boldsymbol{\omega}))$ for all $(s, \boldsymbol{\omega}) \in \Lambda_0^N$. Fix $\delta \in (0, \delta_0)$, and introduce the stopping time

$$H_\delta^N := \inf \{s \geq t^N : \mathcal{W}_2(\mu^N(\mathbf{X}_{\cdot \wedge s}), \mu^N(\boldsymbol{\omega}_{\cdot \wedge t^N}^N)) \geq \delta\} \wedge (t^N + \delta).$$

Observe that, for all $s \geq t^N$, $\boldsymbol{\omega} \in \Omega^N$ s.t. $s \leq H_\delta^N(\boldsymbol{\omega})$, we have

$$\begin{aligned}
|v^N(s, \boldsymbol{\omega})| &\leq v^N(t^N, \boldsymbol{\omega}^N) + \rho_N \left(|s - t^N| + \mathcal{W}_2(\mu^N(\boldsymbol{\omega}_{\cdot \wedge s}), \mu^N(\boldsymbol{\omega}_{\cdot \wedge t^N}^N)) \right) \\
&\leq v^N(t^N, \boldsymbol{\omega}^N) + \rho_N(2\delta),
\end{aligned}$$

where ρ_N is continuity modulus of v^N . Furthermore, ϕ^N is Lipschitz-continuous as it has bounded derivatives. Thus $\phi^N - v^N$ is bounded and uniformly continuous on $\{(s, \boldsymbol{\omega}) : s \leq H_\delta^N(\boldsymbol{\omega})\}$, and by [38] again, there exists $(\theta_\delta^N, \mathbb{P}^{N,*}) \in \mathcal{T}_{t_\delta^N, T}^N \times \mathcal{P}_L^N(t, \boldsymbol{\omega}^N)$ s.t.

$$\mathbb{E}^{\mathbb{P}^{N,*}} \left[(\phi^N - v^N)(\theta_\delta^N \wedge H_\delta^N, \mathbf{X}_{\cdot \wedge \theta_\delta^N \wedge H_\delta^N}) \right] = \sup_{\theta \in \mathcal{T}_{t_\delta^N, T}^N} \bar{\mathcal{E}}_{t_\delta^N, \boldsymbol{\omega}^N}^N \left[(\phi^N - v^N)(\theta \wedge H_\delta^N, \mathbf{X}_{\cdot \wedge \theta \wedge H_\delta^N}) \right],$$

where $\bar{\mathcal{E}}_{t_\delta^N, \boldsymbol{\omega}^N}^N$ is defined by (VI.5.5). Similarly to the Markovian setting, we may find a subsequence $\{\boldsymbol{\omega}^{\delta, N}\}_{N \geq 1}$ s.t. $t_\delta^N := \theta_\delta^N(\boldsymbol{\omega}^{\delta, N}) < H_\delta^N(\boldsymbol{\omega}^{\delta, N})$ for all $N \geq 1$, and satisfying

$$(\phi^N - v^N)(t_\delta^N, \boldsymbol{\omega}^{\delta, N}) = \max_{\theta \in \mathcal{T}_{t_\delta^N, T}^N} \bar{\mathcal{E}}_{t_\delta^N, \boldsymbol{\omega}^{\delta, N}}^N \left[(\phi^N - v^N)(\theta \wedge H_\delta^N, \mathbf{X}_{\cdot \wedge \theta \wedge H_\delta^N}) \right].$$

Observe that $\phi^N \in C_b^{1,2}(\bar{\Lambda}_{t_\delta^N}^N)$. Indeed, since $\varphi = \psi^{v,a,f}$ is a semijet, we have

$$\begin{aligned} \partial_t \phi^N(s, \boldsymbol{\omega}) &= \partial_t \varphi(s, \mu^N(\boldsymbol{\omega})) = a, \\ \partial_{\omega_i} \phi^N(s, \mathbf{x}) &= \frac{1}{N} \partial_{\omega} f(\omega_i) = \frac{1}{N} \partial_x \delta_m \varphi(s, \mu^N(\boldsymbol{\omega}), \omega_i), \\ \partial_{\omega_i \omega_i}^2 \phi^N(s, \boldsymbol{\omega}) &= \frac{1}{N} \partial_{\omega \omega}^2 f(\omega_i) = \frac{1}{N} \partial_{\omega \omega}^2 \delta_m \varphi(s, \mu^N(\boldsymbol{\omega}), \omega_i), \end{aligned}$$

for all $i \in [N]$. As $H_\delta^N > t_\delta^N$ on $\{\mathbf{X}_{t_\delta^N \wedge \cdot} = \boldsymbol{\omega}_{t_\delta^N \wedge \cdot}^{\delta, N}\}$, we have $\phi^N \in \bar{\mathcal{A}}^N v^N(t_\delta^N, \boldsymbol{\omega}^{\delta, N})$ and the supersolution property provides

$$-\partial_t \phi^N(q_\delta^N) - F^N(q_\delta^N, v^N(q_\delta^N), \partial_{\boldsymbol{\omega}} \phi^N(q_\delta^N), \partial_{\boldsymbol{\omega} \boldsymbol{\omega}}^2 \phi^N(q_\delta^N)) \geq 0. \quad (\text{VI.6.5})$$

where $q_\delta^N := (t_\delta^N, \boldsymbol{\omega}^{\delta, N})$. We conclude similarly to (i). \blacksquare

VI.7 A precompactness result

In this section, we state and prove our propagation of chaos-like result for continuous semi-martingale with bounded characteristics:

Proposition VI.7.1 *Let $\{\boldsymbol{\omega}^N\}_{N \geq 1} \in \prod_{N \geq 1} \Omega^N$ and $\mu \in \mathcal{P}_2(\Omega)$ s.t. $\mu^N(\boldsymbol{\omega}_{\cdot \wedge t}^N) \xrightarrow{\mathcal{W}_2} \mu_{[0,t]}$, and $\{\mathbb{P}^N \in \mathcal{P}_L^N(t, \boldsymbol{\omega}^N)\}_{N \geq 1}$. Then, the sequence $\{\mathbb{P}^N \circ (\mu^N(\mathbf{X}))^{-1}\}_{N \geq 1}$ is tight, and all its accumulation points are supported on $\mathcal{P}_L(t, \mu)$.*

Definition VI.7.2 (i) *Denote $Y := (A, M)$ the canonical process on $\Omega^2 := \Omega \times \Omega$. Let $\tilde{\mathcal{P}}_L$ be the set of probability measures \mathbb{P} on Ω^2 such that:*

CHAPTER VI. A FINITE-DIMENSIONAL APPROXIMATION FOR VISCOSITY
SOLUTIONS OF PARTIAL DIFFERENTIAL EQUATIONS ON WASSERSTEIN SPACE

- A is absolutely continuous w.r.t. to the Lebesgue measure on $[0, T]$, with $|\frac{dA_s}{ds}| \leq L$, \mathbb{P} -a.s.,
 - M is a \mathbb{P} -martingale on $[0, T]$, with $\sqrt{\frac{d\langle M \rangle_s}{ds}} \leq L$, \mathbb{P} -a.s.
- (ii) Denote $\mathbf{Y} := (\mathbf{A}, \mathbf{M}) = \{(A^k, M^k)\}_{k \in [N]}$ the canonical process on $\Omega^{N,2} := \Omega^N \times \Omega^N$. Let $\tilde{\mathcal{P}}_L^N$ be the set of probability measures \mathbb{P} on $\Omega^{N,2}$ s.t.:
- \mathbf{A} is absolutely continuous w.r.t. to the Lebesgue measure on $[0, T]$, with $|\frac{dA_s^k}{ds}| \leq L$, for all $k \in [N]$, \mathbb{P} -a.s.,
 - \mathbf{M} is a \mathbb{P} -martingale on $[0, T]$, with $\sqrt{\frac{d\langle M^k \rangle_s}{ds}} \leq L$ and $\langle M^k, M^l \rangle = 0$, for all $k \neq l \in [N]$, \mathbb{P} -a.s.

Lemma VI.7.3 For all $\{\mathbb{P}^N \in \tilde{\mathcal{P}}_L^N\}_{N \geq 1}$, the sequence $\{\mathbb{P}^N \circ (\mu^N(\mathbf{Y}))^{-1}\}_{N \geq 1}$ is tight, and all its accumulation points are supported on $\tilde{\mathcal{P}}_L$.

Proof Step 1: We first prove the existence of a limit. For all $N \geq 1$, denote $\nu^N := \mathbb{P}^N \circ (\mu^N(\mathbf{Y}))^{-1} \in \mathcal{P}(\mathcal{P}_2(\Omega^2))$. By Lacker [51, Corollary B.1], we have to prove that

(i) $\{\nu^N\}_{N \geq 1}$ is uniformly integrable, i.e., $\lim_{R \rightarrow \infty} \sup_{N \geq 1} \mathbb{E}^{\nu^N} \left[\mathcal{W}_2^2(\lambda, \delta_0) \mathbf{1}_{\mathcal{W}_2(\lambda, \delta_0) \geq R} \right] = 0$, where λ is the identity map on $\mathcal{P}_2(\Omega^2)$.

(ii) the sequence of mean measures $\{\mathbb{E}^{\mathbb{P}^N} [\mu^N(\mathbf{Y})]\}_{N \geq 1}$ is tight,

where, for all $\mathbb{P} \in \mathcal{P}_2(\Omega^{N,2})$ and $\tilde{\mu} : \Omega^{N,2} \rightarrow \mathcal{P}_2(\Omega^2)$, the mean measure $\mathbb{E}^{\mathbb{P}}[\tilde{\mu}] \in \mathcal{P}(\Omega^2)$ is defined by

$$\langle \mathbb{E}^{\mathbb{P}}[\tilde{\mu}], \varphi \rangle := \mathbb{E}^{\mathbb{P}}[\langle \tilde{\mu}, \varphi \rangle] \quad \text{for all } \varphi \in C_b^0(\Omega^2).$$

Let $R > 0$. We have

$$\begin{aligned} \mathbb{E}^{\nu^N} \left[\mathcal{W}_2(\lambda, \delta_0) \mathbf{1}_{\mathcal{W}_2(\lambda, \delta_0) \geq R} \right] &= \mathbb{E}^{\mathbb{P}^N} \left[\mathcal{W}_2(\mu^N(\mathbf{Y}), \delta_0) \mathbf{1}_{\mathcal{W}_2(\mu^N(\mathbf{Y}), \delta_0) \geq R} \right] \\ &\leq \frac{1}{R} \mathbb{E}^{\mathbb{P}^N} \left[\mathcal{W}_2^2(\mu^N(\mathbf{Y}), \delta_0) \right] \\ &\leq \frac{2}{R} \mathbb{E}^{\mathbb{P}^N} \left[\frac{1}{N} \sum_{i=1}^N |A^i|^2 + |M^i|^2 \right]. \end{aligned}$$

For each $i \in [N]$, we have

$$\mathbb{E}^{\mathbb{P}^N} [|A^i|^2] \leq (LT)^2 \quad \text{and} \quad \mathbb{E}^{\mathbb{P}^N} [|M^i|^2] \leq 4L^2T,$$

the latter by Doob's inequality. Thus, there exists a constant $C_{T,L}$ independent from N and R s.t.

$$\mathbb{E}^{\nu^N} \left[\mathcal{W}_2(\lambda, \delta_0) \mathbf{1}_{\mathcal{W}_2(\lambda, \delta_0) \geq R} \right] \leq \frac{C_{T,L}}{R} \quad \text{for all } N \geq 1 \text{ and } R \geq 0,$$

CHAPTER VI. A FINITE-DIMENSIONAL APPROXIMATION FOR VISCOSITY
SOLUTIONS OF PARTIAL DIFFERENTIAL EQUATIONS ON WASSERSTEIN SPACE

and therefore $\lim_{R \rightarrow \infty} \sup_{N \geq 1} \mathbb{E}^{\nu^N} \left[\mathcal{W}_2(\lambda, \delta_0) \mathbf{1}_{\mathcal{W}_2(\lambda, \delta_0) \geq R} \right] = 0$ and (i) is proved.

To show that $\{\mathbb{E}^{\mathbb{P}^N} [\mu^N(\mathbf{Y})]\}_{N \geq 1}$ is tight, we prove Aldous' criterion (see Billingsley [10, Theorem 16.10]), i.e.,

$$\sup_{N \geq 1} \sup_{\tau \in \mathcal{T}_{0,T}} \left\langle \mathbb{E}^{\mathbb{P}^N} [\mu^N(\mathbf{Y})], |A_{(\tau+\delta) \wedge T} - A_\tau|^2 + |M_{(\tau+\delta) \wedge T} - M_\tau|^2 \right\rangle \xrightarrow{\delta \rightarrow 0} 0, \quad (\text{VI.7.1})$$

where $\mathcal{T}_{0,T}$ denotes the set of $[0, T]$ -valued \mathbb{F} -stopping times. Yet, we have, for fixed N, τ and δ ,

$$\left\langle \mathbb{E}^{\mathbb{P}^N} [\mu^N(\mathbf{Y})], |M_{(\tau+\delta) \wedge T} - M_\tau|^2 \right\rangle = \frac{1}{N} \sum_{i=1}^N \mathbb{E}^{\mathbb{P}^N} \left[|M_{(\tau+\delta) \wedge T}^i - M_\tau^i|^2 \right] \leq L^2 \delta$$

by Itô's isometry. We obtain a similar estimate for A , and this implies (VI.7.1) and consequently (ii), and thus $\{\nu^N\}_{N \geq 1}$ admits a subsequence converging to some $\nu \in \mathcal{P}(\mathcal{P}_2(\Omega^2))$.

Step 2: We show that ν is supported on $\tilde{\mathcal{P}}_L$. First observe that, by definition of \mathbb{P}^N , we have

$$|A_s^k - A_r^k| \leq L|s - r|, \quad \mathbb{P}^N\text{-a.s.}, \quad \text{for all } k \in [N] \text{ and } s, r \in [0, T],$$

and thus

$$|A_s - A_r| \leq L|s - r|, \quad \mu^N(\mathbf{Y})\text{-a.s.}, \quad \mathbb{P}^N\text{-a.s.}, \quad \text{for all } k \in [N] \text{ and } s, r \in [0, T],$$

and finally

$$\nu^N \left[\lambda \left(|A_s - A_r| \leq L|s - r| \right) = 1 \right] = 1, \quad \text{for all } N \geq 1 \text{ and } s, r \in [0, T].$$

Since $\{|A_s - A_r| \leq L|s - r|\}$ is closed in Ω^2 , $\left\{ \lambda \left(|A_s - A_r| \leq L|s - r| \right) = 1 \right\}$ is closed in $\mathcal{P}_2(\Omega^2)$, and thus the weak convergence of ν^N to ν implies

$$1 = \limsup_{N \rightarrow \infty} \nu^N \left[\lambda \left(|A_s - A_r| \leq L|s - r| \right) = 1 \right] \leq \nu \left[\lambda \left(|A_s - A_r| \leq L|s - r| \right) = 1 \right] \leq 1,$$

that is, $\nu \left[\lambda \left(|A_s - A_r| \leq L|s - r| \right) = 1 \right] = 1$. Since s and r are arbitrary, this implies that A is absolutely continuous w.r.t. the Lebesgue measure on $[0, T]$ with $|\frac{dA_s}{ds}| \leq L$, λ -a.s., ν -a.s.

We now prove that M is a λ -martingale on $[0, T]$, ν -a.s. Fix $r \leq s$ in $[0, T]$, and

$h_r := h(Y_r)$, where $h \in C_b^0(\Omega^2)$. We compute:

$$\begin{aligned} \mathbb{E}^{\nu^N} \left[\langle \lambda, h_r(M_s - M_r) \rangle^2 \right] &= \mathbb{E}^{\mathbb{P}^N} \left[\left(\frac{1}{N} \sum_{i=1}^N h(Y_r^i)(M_s^i - M_r^i) \right)^2 \right] \\ &\leq \frac{1}{N^2} \sum_{i=1}^N |h|^2 \mathbb{E}^{\mathbb{P}^N} \left[|M_s^i - M_r^i|^2 \right] \\ &\leq \frac{|h|^2 L^2 T}{N} \xrightarrow{N \rightarrow \infty} 0, \end{aligned}$$

where we used the fact that $\langle M^k, M^l \rangle \mathbf{1}_{k \neq l} = 0$, and the $\sigma(Y_r)$ -measurability of h_r to derive the first inequality. Thus, as ν^N converges weakly to ν ,

$$0 \leq \mathbb{E}^\nu \left[\langle \lambda, h_r(M_s - M_r) \rangle^2 \right] \leq \liminf_{N \rightarrow \infty} \mathbb{E}^{\nu^N} \left[\langle \lambda, h_r(M_s - M_r) \rangle^2 \right] = 0,$$

hence $\mathbb{E}^\nu \left[\langle \lambda, h_r(M_s - M_r) \rangle^2 \right] = 0$, which implies that

$$\langle \lambda, h_r(M_s - M_r) \rangle = 0, \quad \nu\text{-a.s.},$$

which by the arbitrariness of s, r and h means that M is a λ -martingale, ν -a.s. We prove similarly to A that $\sqrt{\frac{d(M)_s}{ds}} \leq L$, λ -a.s., ν -a.s. ■

Proof of Propostion VI.7.1 Introduce

$$\mathbf{A}_s^N := \mathbf{X}_t + \mathbf{1}_{s \geq t} \int_t^s b_r^{\mathbb{P}^N} dr, \quad \mathbf{M}_s^N := \mathbf{1}_{s \geq t} \int_t^s \sigma_r^{\mathbb{P}^N} dW_r^{\mathbb{P}^N},$$

where $b^{\mathbb{P}^N}, \sigma^{\mathbb{P}^N}$ and $W^{\mathbb{P}^N}$ are as in (VI.3.3). Then, we clearly have

$$\tilde{\mathbb{P}}^N := \mathbb{P}_{(\mathbf{A}^N, \mathbf{M}^N)}^N \in \tilde{\mathcal{P}}_L^N.$$

Therefore, by Lemma VI.7.3, $\nu^N := \tilde{\mathbb{P}}^N \circ (\mu^N(\mathbf{Y}))^{-1}$ converges weakly to some ν supported on $\tilde{\mathcal{P}}_L(t, \mu)$. Define $\hat{\mu}^N := \mathbb{P}^N \circ (\mu^N(\mathbf{X}))^{-1}$ and fix $\varphi \in C_b^0(\mathcal{P}_2(\Omega))$. We have:

$$\begin{aligned} \langle \hat{\mu}^N, \varphi \rangle &= \mathbb{E}^{\mathbb{P}^N} \left[\varphi(\mu^N(\mathbf{X})) \right] = \mathbb{E}^{\mathbb{P}^N} \left[\varphi(\mu^N(\mathbf{Y}) \circ (A + M)^{-1}) \right] \\ &= \mathbb{E}^{\nu^N} \left[\varphi(\lambda \circ (A + M)^{-1}) \right] \xrightarrow{N \rightarrow \infty} \mathbb{E}^\nu \left[\varphi(\lambda \circ (A + M)^{-1}) \right] \end{aligned}$$

by weak convergence of ν^N to ν , since $\lambda \mapsto \varphi(\lambda \circ (A + M)^{-1}) \in C_b^0(\mathcal{P}_2(\Omega^2))$. Thus we have

$$\langle \hat{\mu}^N, \varphi \rangle \xrightarrow{N \rightarrow \infty} \langle \hat{\mu}, \varphi \rangle,$$

where $\hat{\mu} := \nu \circ (\lambda \circ (A + M)^{-1})^{-1}$. Therefore, by arbitrariness of φ , μ^N converges weakly to $\hat{\mu}$, which is clearly supported on $\mathcal{P}_L(t, \mu)$ since ν is supported on $\tilde{\mathcal{P}}_L$ and $\mu^N(\omega^N) \xrightarrow{\mathcal{W}_2} \mu$. ■

Acknowledgments

First of all, I would like to express my deepest gratitude to Nizar Touzi and Jianfeng Zhang, who decisively contributed to the success of this thesis by their insightful ideas, their constant availability and, above all, their tireless dedication to our project. I know how fortunate I have been to work with such brilliant minds.

It has also been an honor for me to have Erhan Bayraktar and Pierre Cardaliaguet as thesis referees. Their expertise and excellent suggestions have been greatly appreciated. I also thank Charles Bertucci, François Delarue, Astrid Hilbert and Dylan Possamaï for accepting to participate to the thesis jury.

I also want to express my sincere affection for all those I have had to chance to meet over the past few years, at the CMAP or at USC: Thibaut Mastrolia, J.E. Paguyo, Kaitong Hu, Bastien Baldacci, Paul Jusselin, Heythem Farhat, Gaoyue Guo, Leila Bassou, Songbo Wang, Wanqing Wang, Emmanouil Sfendourakis, Aldjia Mazari, Jacob Der, and Daniel Douglas.

Finally, I address a special acknowledgment to my parents.

Bibliography

- [1] Daniel Andersson and Boualem Djehiche. A maximum principle for SDEs of mean-field type. *Applied Mathematics & Optimization*, 63(3):341–356, 2011.
- [2] Guy Barles and Panagiotis E Souganidis. Convergence of approximation schemes for fully nonlinear second order equations. *Asymptotic analysis*, 4(3):271–283, 1991.
- [3] Erhan Bayraktar, Alekos Cecchin, and Prakash Chakraborty. Mean field control and finite dimensional approximation for regime-switching jump diffusions. *Preprint arXiv:2109.09134*, 2021.
- [4] Erhan Bayraktar, Alekos Cecchin, Asaf Cohen, and François Delarue. Finite state mean field games with Wright–Fisher common noise as limits of N-player weighted games. *Mathematics of Operations Research*, 2022.
- [5] Erhan Bayraktar, Alekos Cecchin, Asaf Cohen, and François Delarue. Finite state mean field games with Wright–Fisher common noise. *Journal de Mathématiques Pures et Appliquées*, 147:98–162, 2021.
- [6] Erhan Bayraktar, Andrea Cosso, and Huyên Pham. Randomized dynamic programming principle and Feynman-Kac representation for optimal control of McKean-Vlasov dynamics. *Transactions of the American Mathematical Society*, 370(3):2115–2160, 2018.
- [7] Denis Belomestny and John Schoenmakers. Optimal stopping of McKean–Vlasov diffusions via regression on particle systems. *SIAM Journal on Control and Optimization*, 58(1):529–550, 2020.
- [8] Alain Bensoussan, Jens Frehse, and Phillip Yam. *Mean field games and mean field type control theory*, volume 101. Springer, 2013.

BIBLIOGRAPHY

- [9] Charles Bertucci. Optimal stopping in mean field games, an obstacle problem approach. *Journal de Mathématiques Pures et Appliquées*, 120:165–194, 2018.
- [10] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.
- [11] Géraldine Bouveret, Roxana Dumitrescu, and Peter Tankov. Mean-field games of optimal stopping: a relaxed solution approach. *SIAM Journal on Control and Optimization*, 58(4):1795–1821, 2020.
- [12] Philippe Briand, Romuald Elie, and Ying Hu. BSDEs with mean reflection. *The Annals of Applied Probability*, 28(1):482–510, 2018.
- [13] Rainer Buckdahn, Boualem Djehiche, and Juan Li. A general stochastic maximum principle for SDEs of mean-field type. *Applied Mathematics & Optimization*, 64(2):197–216, 2011.
- [14] Rainer Buckdahn, Juan Li, Shige Peng, and Catherine Rainer. Mean-field stochastic differential equations and associated PDEs. *The Annals of Probability*, 45(2):824–878, 2017.
- [15] Matteo Burzoni, Vincenzo Ignazio, A Max Reppen, and H Mete Soner. Viscosity solutions for controlled McKean–Vlasov jump-diffusions. *SIAM Journal on Control and Optimization*, 58(3):1676–1699, 2020.
- [16] Pierre Cardaliaguet. The convergence problem in mean field games with local coupling. *Applied Mathematics & Optimization*, 76(1):177–215, 2017.
- [17] Pierre Cardaliaguet, François Delarue, Jean-Michel Lasry, and Pierre-Louis Lions. *The Master Equation and the Convergence Problem in Mean Field Games:(AMS-201)*, volume 201. Princeton University Press, 2019.
- [18] Pierre Cardaliaguet and Marc Quincampoix. Deterministic differential games under probability knowledge of initial condition. *International Game Theory Review*, 10(01):1–16, 2008.
- [19] Pierre Cardaliaguet and Panagiotis Souganidis. Regularity of the value function and quantitative propagation of chaos for mean field control problems. *Preprint arXiv:2204.01314*, 2022.

BIBLIOGRAPHY

- [20] René Carmona and Francois Delarue. The master equation for large population equilibriums. In *Stochastic analysis and applications 2014*, pages 77–128. Springer, 2014.
- [21] René Carmona and François Delarue. *Probabilistic Theory of Mean Field Games with Applications I-II*. Springer, 2018.
- [22] René Carmona, François Delarue, and Aimé Lachapelle. Control of McKean–Vlasov dynamics versus mean field games. *Mathematics and Financial Economics*, 7(2):131–166, 2013.
- [23] René Carmona, François Delarue, and Daniel Lacker. Mean field games of timing and models for bank runs. *Applied Mathematics & Optimization*, 76(1):217–260, 2017.
- [24] Alekos Cecchin. Finite state N-agent and mean field control problems. *ESAIM: Control, Optimisation and Calculus of Variations*, 27:31, 2021.
- [25] Alekos Cecchin and Guglielmo Pelino. Convergence, fluctuations and large deviations for finite state mean field games via the master equation. *Stochastic Processes and their Applications*, 129(11):4510–4555, 2019.
- [26] Alekos Cecchin, Paolo Dai Pra, Markus Fischer, and Guglielmo Pelino. On the convergence problem in mean field games: a two state model without uniqueness. *SIAM Journal on Control and Optimization*, 57(4):2443–2466, 2019.
- [27] Jean-François Chassagneux, Dan Crisan, and François Delarue. A probabilistic approach to classical solutions of the master equation for large population equilibria. *Mem. Amer. Math. Soc.*, *accepted*, *arXiv:1411.3009*, 2014.
- [28] Andrea Cosso and Huyên Pham. Zero-sum stochastic differential games of generalized McKean–Vlasov type. *Journal de Mathématiques Pures et Appliquées*, 129:180–212, 2019.
- [29] Michael G Crandall, Hitoshi Ishii, and Pierre-Louis Lions. User’s guide to viscosity solutions of second order partial differential equations. *Bulletin of the American mathematical society*, 27(1):1–67, 1992.
- [30] Michael G Crandall and Pierre-Louis Lions. Viscosity solutions of Hamilton-Jacobi equations. *Transactions of the American mathematical society*, 277(1):1–42, 1983.

BIBLIOGRAPHY

- [31] Boualem Djehiche, Romuald Elie, and Said Hamadène. Mean-field reflected backward stochastic differential equations. *preprint arXiv:1911.06079*, 2019.
- [32] Mao Fabrice Djete. Mean field games of controls: on the convergence of Nash equilibria. *arXiv preprint arXiv:2006.12993*, 2020.
- [33] Mao Fabrice Djete. Extended mean field control problem: a propagation of chaos result. *Electronic Journal of Probability*, 27:1–53, 2022.
- [34] Mao Fabrice Djete, Dylan Possamaï, and Xiaolu Tan. McKean–Vlasov optimal control: the dynamic programming principle. *The Annals of Probability*, 50(2):791–833, 2022.
- [35] Josu Doncel, Nicolas Gast, and Bruno Gaujal. Discrete mean field games: Existence of equilibria and convergence. *arXiv preprint arXiv:1909.01209*, 2019.
- [36] Ibrahim Ekren. Viscosity solutions of obstacle problems for fully nonlinear path-dependent PDEs. *Stochastic Processes and their Applications*, 127(12):3966–3996, 2017.
- [37] Ibrahim Ekren, Christian Keller, Nizar Touzi, Jianfeng Zhang, et al. On viscosity solutions of path dependent PDEs. *Annals of Probability*, 42(1):204–236, 2014.
- [38] Ibrahim Ekren, Nizar Touzi, and Jianfeng Zhang. Optimal stopping under nonlinear expectation. *Stochastic Processes and Their Applications*, 124(10):3277–3311, 2014.
- [39] Ibrahim Ekren, Nizar Touzi, Jianfeng Zhang, et al. Viscosity solutions of fully nonlinear parabolic path dependent PDEs: Part I. *Annals of Probability*, 44(2):1212–1253, 2016.
- [40] Ibrahim Ekren, Nizar Touzi, Jianfeng Zhang, et al. Viscosity solutions of fully nonlinear parabolic path dependent PDEs: Part II. *Annals of Probability*, 44(4):2507–2553, 2016.
- [41] Nicole El Karoui. Les aspects probabilistes du contrôle stochastique. In *Ecole d’Eté de Probabilités de Saint-Flour IX-1979*, pages 73–238. Springer, 1981.
- [42] Markus Fischer and Giulia Livieri. Continuous time mean-variance portfolio optimization through the mean field approach. *ESAIM: Probability & Statistics*, 20, 2016.
- [43] Gerald B Folland. *Real analysis: modern techniques and their applications*, volume 40. John Wiley & Sons, 1999.

BIBLIOGRAPHY

- [44] Wilfrid Gangbo, Sergio Mayorga, and Andrzej Swiech. Finite dimensional approximations of Hamilton–Jacobi–Bellman equations in spaces of probability measures. *SIAM Journal on Mathematical Analysis*, 53(2):1320–1356, 2021.
- [45] Wilfrid Gangbo, Truyen Nguyen, and Adrian Tudorascu. Hamilton-Jacobi equations in the Wasserstein space. *Methods and Applications of Analysis*, 15(2):155–184, 2008.
- [46] Xin Guo, Huy en Pham, and Xiaoli Wei. Itô’s formula for flow of measures on semi-martingales. *preprint arXiv:2010.05288*, 2020.
- [47] Minyi Huang, Roland P Malham e, and Peter Caines. Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the nash certainty equivalence principle. *Communications in Information & Systems*, 6(3):221–252, 2006.
- [48] Ioannis Karatzas and Steven Shreve. *Brownian motion and stochastic calculus*, volume 113. springer, 2014.
- [49] Ioannis Karatzas and Steven E Shreve. *Methods of mathematical finance*, volume 39. Springer, 1998.
- [50] Magdalena Kobylanski, Marie-Claire Quenez, and Elisabeth Rouy-Mironescu. Optimal multiple stopping time problem. *The Annals of Applied Probability*, 21(4):1365–1399, 2011.
- [51] Daniel Lacker. Mean field games via controlled martingale problems: existence of Markovian equilibria. *Stochastic Processes and their Applications*, 125(7):2856–2894, 2015.
- [52] Daniel Lacker. Limit theory for controlled McKean–Vlasov dynamics. *SIAM Journal on Control and Optimization*, 55(3):1641–1672, 2017.
- [53] Jean-Michel Lasry and Pierre-Louis Lions. Mean field games. *Japanese journal of mathematics*, 2(1):229–260, 2007.
- [54] Mathieu Lauri ere and Olivier Pironneau. Dynamic programming for mean-field type control. *Comptes Rendus Math ematiques*, 352(9):707–713, 2014.

BIBLIOGRAPHY

- [55] Mathieu Laurière and Ludovic Tangpi. Convergence of large population games to mean field games with interaction through the controls. *arXiv preprint arXiv:2004.08351*, 2020.
- [56] Juan Li. Reflected mean-field backward stochastic differential equations. Approximation and associated nonlinear PDEs. *Journal of Mathematical Analysis and Applications*, 413(1):47–68, 2014.
- [57] Juan Li. Mean-field forward and backward SDEs with jumps and associated nonlocal quasi-linear integral-PDEs. *Stochastic Processes and their Applications*, 128(9):3118–3180, 2018.
- [58] Pierre-Louis Lions. Cours au Collège de France. Available at www.college-de-france.fr, 2007.
- [59] PA Meyer and WA Zheng. Tightness criteria for laws of semimartingales. In *Annales de l'IHP Probabilités et statistiques*, volume 20, pages 353–372, 1984.
- [60] Thilo Meyer-Brandis, Bernt Oksendal, and Xun Yu Zhou. A mean-field stochastic maximum principle via Malliavin calculus. *Stochastics An International Journal of Probability and Stochastic Processes*, 84(5-6):643–666, 2012.
- [61] Chenchen Mou and Jianfeng Zhang. Wellposedness of second order master equations for mean field games with nonsmooth data. *Memoirs of AMS*, *accepted*.
- [62] Marcel Nutz. A mean field game of optimal stopping. *SIAM Journal on Control and Optimization*, 56(2):1206–1221, 2018.
- [63] Karl Oelschläger. A martingale approach to the law of large numbers for weakly interacting stochastic processes. *The Annals of Probability*, pages 458–479, 1984.
- [64] Jesper Lund Pedersen and Goran Peskir. Optimal mean–variance selling strategies. *Mathematics and Financial Economics*, 10(2):203–220, 2016.
- [65] Huyên Pham and Xiaoli Wei. Bellman equation and viscosity solutions for mean-field stochastic control problem. *ESAIM: Control, Optimisation and Calculus of Variations*, 24(1):437–461, 2018.

BIBLIOGRAPHY

- [66] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.
- [67] Zhenjie Ren, Nizar Touzi, and Jianfeng Zhang. An overview of viscosity solutions of path-dependent PDEs. *Stochastic Analysis and Applications 2014*, pages 397–453, 2014.
- [68] Zhenjie Ren, Nizar Touzi, and Jianfeng Zhang. Comparison of viscosity solutions of fully nonlinear degenerate parabolic path-dependent PDEs. *SIAM Journal on Mathematical Analysis*, 49(5):4093–4116, 2017.
- [69] R Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.
- [70] Albert N Shiryaev. *Optimal stopping rules*, volume 8. Springer Science & Business Media, 2007.
- [71] Justin Sirignano and Konstantinos Spiliopoulos. DGM: A deep learning algorithm for solving partial differential equations. *Journal of computational physics*, 375:1339–1364, 2018.
- [72] Alain-Sol Sznitman. Topics in propagation of chaos. In *Ecole d’été de probabilités de Saint-Flour XIX—1989*, pages 165–251. Springer, 1991.
- [73] Mehdi Talbi. A finite-dimensional approximation for viscosity solutions of partial differential equations on Wasserstein space. In preparation.
- [74] Mehdi Talbi, Nizar Touzi, and Jianfeng Zhang. Dynamic programming equation for the mean field optimal stopping problem. *Preprint arXiv:2103.05736*, 2021.
- [75] Mehdi Talbi, Nizar Touzi, and Jianfeng Zhang. Viscosity solutions for obstacle problems on Wasserstein space. *Preprint arXiv:2203.17162*, 2022.
- [76] Mehdi Talbi, Nizar Touzi, and Jianfeng Zhang. A finite-dimensional approximation for obstacle problems on Wasserstein space. In preparation.

BIBLIOGRAPHY

- [77] Cong Wu and Jianfeng Zhang. Viscosity solutions to parabolic master equations and McKean–Vlasov SDEs with closed-loop controls. *Annals of Applied Probability*, 30(2):936–986, 2020.
- [78] Zuo Quan Xu and Xun Yu Zhou. Optimal stopping under probability distortion. *The Annals of Applied Probability*, 23(1):251–282, 2013.
- [79] Wei An Zheng. Tightness results for laws of diffusion processes application to stochastic mechanics. In *Annales de l’IHP Probabilités et statistiques*, volume 21, pages 103–124, 1985.

Titre : Arrêt optimal champ-moyen et approximations d'équations aux dérivées partielles sur l'espace de Wasserstein

Mots clés : Arrêt optimal, champ-moyen, contrôle stochastique.

Résumé : Cette thèse comporte deux parties. La première porte sur l'étude du problème d'arrêt optimal champ moyen, c'est-à-dire de l'arrêt optimal d'une diffusion du type McKean-Vlasov, lorsque le critère à optimiser est une fonction de la distribution du processus arrêté. Ce problème permet de modéliser la situation où un planificateur central, contrôlant une population infinie d'agents en interaction, doit attribuer à chaque agent un temps d'arrêt en vue d'optimiser un certain critère dépendant de la distribution du système. Nous étudions ce problème via une approche type programmation dynamique, qui permet de caractériser sa fonction valeur par une équation aux dérivées partielles sur l'espace des mesures, que nous appelons problème (ou équation) de l'obstacle sur l'espace de Wasserstein, par analogie avec le problème de l'obstacle classique, rencontré notamment en arrêt optimal standard. Nous montrons en particulier que si cette équation dispose d'une solution classique, alors cette dernière est égale à la fonction valeur du problème d'arrêt optimal champ-moyen, et peut être utilisée pour caractériser les stratégies d'arrêt optimales. Nous étendons ensuite cette étude au cas où la fonction valeur n'est pas nécessairement différentiable. Ainsi, nous introduisons une notion de solution de viscosité pour l'équation de l'obstacle sur l'espace de Wasserstein, pour laquelle nous montrons les propriétés

de consistance avec les solutions classiques, de stabilité et d'unicité. Dans la seconde partie de la thèse, nous nous intéressons au développement d'approximations pour certaines classes d'équations aux dérivées partielles sur l'espace des mesures de probabilité. Plus précisément, nous montrons que les solutions de viscosité de ces équations peuvent s'écrire comme limites de solutions de viscosité d'équations définies sur des espaces de dimension finie. Nous nous intéressons d'abord au cas du problème de l'obstacle sur l'espace de Wasserstein, dont il s'avère que l'équation d'approximation correspond à l'équation de la programmation dynamique associée au problème d'arrêt optimal multiple, qui peut être vu comme une formulation du problème d'arrêt optimal champ-moyen en population finie. Nous traitons enfin le cas d'une classe plus générale d'équations paraboliques sur l'espace de Wasserstein. Cette classe couvre notamment le cas des équations d'Hamilton-Jacobi-Bellman champ-moyen, ou encore celui des équations affleurant dans les problèmes de jeux différentiels champ-moyen. Nous montrons également que nos résultats peuvent être étendus aux équations dépendant de la trajectoire (c'est-à-dire, dont les variables de la solution ne sont plus le temps et une mesure sur \mathbf{R}^d , mais le temps et une mesure sur l'espace des trajectoires continues).

Title : Mean field optimal stopping and approximations of partial differential equations on Wasserstein space

Keywords : Optimal stopping, mean field, stochastic control.

Abstract : This thesis consists in two parts. The first one is concerned with the study of the mean field optimal stopping problem, that is the optimal stopping of a McKean-Vlasov diffusion, when the criterion to optimize is a function of the stopped process. This problem models the situation where a central planner controls a continuous infinity of interacting agents by assigning a stopping time to each of them, in order to optimize some criterion which depends on the distribution of the system. We study this problem via a dynamic programming approach, which allows to characterize its value function by a partial differential equation on the space of probability measures, that we call obstacle problem (or equation) on Wasserstein space by analogy with the classical obstacle problem, which arises in particular in standard optimal stopping. We especially show that, if this equation has a classical solution, then it is equal to the value function of the mean field optimal stopping problem, and that it can be used to characterize optimal stopping policies. We next extend our study to the case where the value function is not necessarily differentiable. Thus, we introduce a notion of viscosity solution for the obstacle problem on Wasserstein space, for which we prove the

properties of consistency with classical solutions, stability and uniqueness. In the second part of the thesis, we are interested in developing approximations for some classes of partial differential equations on the space of probability measures. More precisely, we show that viscosity solutions of these equations can be written as limits of viscosity solutions of equations defined on finite-dimensional spaces. We first focus our study on the case of the obstacle problem on Wasserstein space, for which it turns out that the approximating equation corresponds to the dynamic programming equation associated with the multiple optimal stopping problem, which may be seen as a finite population formulation of the mean field optimal stopping problem. We finally consider a larger class of parabolic equations on Wasserstein space. This class covers in particular the scope of mean field Hamilton-Jacobi-Bellman equations, or the case of equations arising in mean field differential games. We also prove that our results can be extended to the case of path-dependent equations (i.e., when the variables of the solution are not the time and a measure on \mathbf{R}^d , but the time and a measure on the space of paths).