Université Paris-Dauphine, PSL Research Univerity

CNRS, CEREMADE

Mémoire d'habilitation à diriger des recherches

Xiaolu TAN

Martingale Optimal Transport, Non Markovian Stochastic Control and Branching Diffusion Processes

Le 1er décembre 2017

Jury :

Coordinateur :	Bruno Bouchard	-	Université Paris-Dauphine, PSL Research University
Rapporteurs :	François Delarue	-	Université Nice-Sophia Antipolis
	Huyên Pham	-	Université Paris-Diderot, Paris VII
	Mete Soner	-	ETH Zurich
Examinateurs :	Nicole El Karoui	-	Université Pierre et Marie Curie
	Gilles Pagès	-	Université Pierre et Marie Curie
	Denis Talay	-	INRIA Sophia Antipolis
	Nizar Touzi	-	Ecole Polytechnique

Remerciement

Je tiens d'abord à remercier Nizar Touzi, mon ancien directeur de thèse, qui m'a conduit sur cette voie de recherche. Merci pour tes aides et ton encouragement constants pendant les dernières années. J'aimerais également dire merci à Frédéric Bonnans, Nicole El Karoui et Bruno Bouchard, qui m'ont consacrés de plein de temps pour discuter les maths (et les non-maths :)). J'en ai bénéficié un encadrement supplémentaire même après ma thèse.

J'aimerais remercier François Delarue, Huyên Pham et Mete Soner, qui m'ont donné l'honneur de rapporter mon mémoire d'HDR, et également à Gilles Pagès et Denis Talay pour accepter de participer au jury.

Un grand merci à Pierre Henry-Labordère, Xavier Warin et Nadia Oudjane, avec qui j'ai eu des collaborations très enrichissantes. Merci pour vos idées originales, et votre efficacité dans le travail. J'aimerais aussi remercier Aymeric Kalife, avec qui les échanges ont commencé depuis mon stage du master, et j'en ai beaucoup profité.

Merci à Dylan, Chao et Nabil, c'était une telle chance pour moi de vous connaître dès le début de la thèse, et de grandir ensemble. Même on ne travaille pas toujours sur le même sujet, on pourrait toujours se poser des questions les plus bêtes. Mais, la recherche avance grâce à ces bêtises. Merci à tous les postdoc (et les anciens postdoc) dans le bureau postdoc du CMAP, Zhenjie, Julien, Yiqing, Junjian, Sigrid, Alexandre, Ankush, Mauro, ainsique l'équipe administrative du CMAP, surtout Nasséra, Alexandra, Wilfried, Vincent. Grâce à vous, j'y trouve une belle ambiance de travail, avec la joie et l'amitié lorsque j'y passe.

Merci à Gaoyue, Shuoqing et Fabrice, qui m'ont fait confiance et me permet d'avoir des premières expériences de (co-)encadrement de thèse. Ca fait un grand plaisir et aussi une grande chance pour moi de travailler avec vous.

Je tiens à remercier beaucoup de chercheurs/chercheuses dans la communauté "mathématique finance", notamment celle en France, d'où j'ai eu beaucoup d'aides, de discussion, ou/et des collaborations. J'aimerais remercier notamment Emmanuel, Mathieu, Thibault, Stéfano, Caroline, Peter, Nicolas, Wissal, Laurent, Anis, Jean-François, Simone, Monique, Sergio, Eduardo, Jan, Ania, etc. (Il manque certainement beaucoup de noms dans la liste ...).

Enfin, j'aimerais remercier les collègues au Ceremade, notamment l'équipe EcoFi. Merci à l'équipe formidable de l'administration et de l'informatique de l'Ecole Doctorale de Dauphine (surtout Mme Karine BEAUVALLET), de MIDO et de Ceremade, Marie, Isabelle, César, Gilles, Thomas, etc. Vous êtes tellement aimables, et efficaces pour répondre à mes demandes.

Pour finir, un grand merci à ma famille: ma femme, mon fils, ainsi que les parents dans les deux familles. Merci pour les soutiens et le moral que vous m'apportez depuis toujours.

Sommaire

Chapitre 1

Introduction

English Version

This is a synthesis of my research works since the defense of my PhD thesis, most of which are related to the stochastic optimal control problems (without or with constraints) and nonlinear (path-dependent) PDEs, and are usually motivated by (but not limited to) the applications in finance. I will try to regroup them into three different themes according to their applications and the techniques used:

(i) Martingale Optimal Transport (MOT) problem.

(ii) Non-Markovian stochastic optimal control, path-dependent HJB equations.

(iii) Branching diffusion representation for nonlinear PDEs.

In the following, I will develop the three themes with more details and discuss how these different themes are related with each other.

Martingale Optimal Transport problem

The Martingale Optimal Transport (MOT) problem is a very recent subject in mathematical finance and turns to be a "hot" tropic during the last years. It was initially formulated by Beiglböck, Henry-Labordère and Penkner [18], and Galichon, Henry-Labordère and Touzi [99], motivated by the problem of finding the minimal super-hedging cost for a path-dependent exotic option using semi-static strategies. By considering more general controlled diffusion dynamics in place of all martingales, my PhD thesis (2009-2011) has been dedicated to a similar optimal transport problem (see Tan and Touzi [192]). Similar problems have also been formulated and studied earlier using the Skorokhod embedding approach, initiated by Hobson [119].

In the classical theory of financial market modelling, one usually assumes a reference probability space, where the (stochastic) dynamic of the underlying risky assets is fixed. Then under the reference model, one analyzes the financial risk, prices the derivative options, etc. When the market is perfect, i.e. every contingent option can be replicated perfectly by an auto-financing trading strategy, then under the no-arbitrage condition, the price/value of the option should be equal to its replication cost. From another point of view, this price can be obtained by taking expectation of the discounted payoff under the risk-neutral probability measure, which is the unique martingale measure equivalent to the reference probability measure. Here the martingale measure refers to the probability measure under which the discounted underlying assets are all martingales. When the market is imperfect, the equivalent martingale measure is no more unique. Nevertheless, it follows by the fundamental theorem of asset pricing (FTAP) that the expectation of the discounted payoff under any equivalent martingale measure provides a no-arbitrage price of the option. At the same time, a general option will not be replicable, and one needs to consider the super-replication strategies to obtain a safe price. In this case, a pricing-hedging duality has been obtained, that is, the minimal super-replication cost of the option equals to the supremum of its no-arbitrage prices.

More recently, and in particular since the financial crisis in 2008, the model risk attracts more and more attentions by the practitioners as well as the academic researchers. To overcome the model risk difficulty, a natural way is to avoid imposing a fixed reference model, but rather to consider a family of models, or equivalently a family of reference probability measures. In particular, a model free "safe" price for an option should be its minimal superhedging cost, where the superhedging is defined in sense of quasi sure, i.e. the superhedging holds almost surely under each probability measure in the family.

The above model free pricing method may lead to very high prices for a "simple" option, or even an unreasonable price for the real market. At the same time, the financial market has been much developed such that the "simple" call/put vanilla options are very liquid so that one can use them in the trading strategies to hedge more complicated options (exotic options). One can then expect to obtain more reasonable superhedging cost of exotic options by considering the semi-static strategies, that is, the dynamic trading strategy on underlying risky assets together with the static strategy on the vanilla (call/put) options. From a pricing point of view, if one uses martingale measure to obtain a no-arbitrage price of an option (according to the fundamental theorem of pricing), one should consider those martingale measures consistent with the market information about the vanilla option prices. It is well known that, in the limit case where the call/put prices are known for all possible strikes at some fixed maturity, one can recover a marginal distribution of the underlying risky asset. To conclude, an upper bound of the no-arbitrage price for an exotic option is given by the supremum of the expected value of the discounted payoff under all martingale measures satisfying some marginal distribution constraint. This leads to the name "martingale optimal transport" (MOT) problem, recalling that the classical optimal transport consists in optimising a cost function among all random vector satisfying some marginal constraints (without any martingale structure).

For this MOT problem, the natural questions could be: does the classical pricing-hedging duality still hold in this context? do optimal solutions exist for both pricing and hedging problems? how to find/characterize the optimal solutions? can we approximate it numerically? etc. These questions are of course not independent but are generally related to each other. Moreover, the applications of this new subject are not limited to the finance, but are related closely to other branches of the probability theory, such as the Skrokohod embedding problem, the martingale inequality, Peacocks ("Processus croissant pour l'ordre convexe" in French), etc.

A large part of my contribution in this topic consists in proving the pricing-hedging duality in different situations: such as the case where the underlying assets have continuous paths or càdlàg paths, the limit case of MOT problem with infinitely many marginals constraints, the discrete time case for American options, and/or under additional transaction cost constraints. Moreover, I also studied the continuous time limit of a family of discrete time optimal martingale transport plans. In particular, it leads to a new class of continuous time martingale Peacocks. Further, I suggested an alternative proof for the monotonicity principle of the optimal Skorokhod embedding problem (SEP), using the duality result and an optional section theorem. The monotonicity principle provides a unified characterization on the solutions to the optimal SEP. Finally, I also worked on the numerical approximation of the MOT problems, based on a dual formulation.

Non-Markovian control, path-dependent HJB equations

Although the MOT problem is named as an extension of the classical optimal transport problem, it can also be viewed as an optimal control problem under marginal constraints. This point of view could be very useful in order to recover results and techniques in the control theory. As the reward/cost function in MOT are generally non-Markovian, a better understanding of the non-Markovian control theory would be necessary. At the same time, the new development of finance, the game theory, and the principal-agent problems, etc. provides new motivations to study the non-Markovian control problems.

Stochastic control theory has been largely studied since 1970s, where both PDE approach and the probabilistic approach are developed. When the value function of the control problem is not smooth enough, it is usually the notion of viscosity solution allows to obtain a HJB equations characterization. Notice that the classical HJB equation approach is limited to the Markovian case, but the recently developed path-dependent PDE theory allows to study non-Markovian control problems. Further, the dynamic programming principle has always been an important tool as it is allows to decompose a global optimization problem into a family of local optimization problems. With the development of financial mathematics in 1990s, the stochastic control theory received a renewed attention and remains an important subject until now. Moreover, the BSDE and more recently the 2nd order BSDE theory have been developed and provide new powerful probabilistic tools to study Markovian or non-Markovian control problems.

From a theoretical point of view, the stochastic control theory provides a probabilistic representation for a class of nonlinear PDEs, which is an extension of the Feynmann-Kac formula for linear PDEs. This point of view is generalized by the path-dependent PDE, which in particular provides a unified approach to study the BSDE, 2BSDE, stochastic optimal control, stochastic differential equations, etc.

My first contribution in this tropic consists in providing a review together with some new development on a rigorous and detailed justification of the dynamic programming principle for different formulations of the non-Markovian control problem. Here I use the word “rigorous” since the measurability and the neglected set issues become very subtle, once different singular probability measures are involved. A key argument used here is the measurable selection theorem. Next, I worked on the dynamic programming principle for a control problem on a family of BSDEs. In particular, it allows to generalize the 2BSDE theory to the case without any continuity conditions on the coefficient functions. In this context, as in the classical approaches, the dynamic programming principle allows to characterize the value function process as a supermartingale under a family of involved probability measures. I then worked on the Doob-Meyer decomposition and estimations of supermartingales. Finally, I am also interested by the numerical approximation of the non-Markovian control problems. To solve a continuous time optimization problem, a first step is to discretise the time to obtain a discrete time optimization problem, which could be solved numerically. To prove the convergence from the discrete-time problems to the continuous time problem in this non-Markovian context, I studied two approaches. A first approach is in spirit of Kushner-Dupuis's weak convergence method. Moreover, using a strong invariance principle result, one can obtain a convergence rate under technical conditions. A second approach to prove the convergence consists in extending the seminal work of Barles and Souganidis on the monotone scheme by using the recent developed path-dependent PDE theory.

Branching diffusion representation for nonlinear PDEs

It is quite natural and classical to use the time discretization approach to solve the continuous time problems numerically. However, in the context of the stochastic control problems, or the nonlinear equations, a major limit of this approach is that the monotonicity condition (which is a necessary condition to obtain the convergence) is hard to ensure in high-dimensional case. Moreover, even if the monotonicity condition holds, there are usually conditional expectation terms in the numerical schemes which require an additional estimation step in practical implementation. Such estimation are usually based on a regression technique which can be seen as a variation of the finite element methods, and hence it is usually limited to the lower dimensional case in practice.

Another numerical method for nonlinear PDEs, which avoids the monotonicity condition as well as the regression techniques, could be obtained by using branching diffusion processes. Indeed, it is well known from the very early work of Watanabe, Skorokhod, Mckean, etc. that the KPP type equation can be represented by branching processes. A nonlinear PDE is called a KPP type equation if its nonlinearity part is a polynomial of the value function u, with positive coefficients whose sum equals to 1. As extension of the Feynmann-Kac formula for linear equation, the solution of the KPP equation can be represented as expected value of a functional of a branching diffusion process. This representation has also been generalized by the super-processes as a (scaled) limit of branching process. Using this probabilistic representation, one can estimate the solution of KPP type PDE by a forward Monte Carlo method, that is, simulating a large number of the branching diffusion processes and then taking the average value of the corresponding functional of the branching process.

My contribution in this topic consists mainly in investigating the use of branching diffusion processes in the numerical resolution of nonlinear PDEs. To achieve this, we tried to extend the classical probabilistic representation to a larger class of nonlinear equations, in place of the class of KPP equations. First, we consider the nonlinear PDE whose nonlinearity is a polynomial of u and its derivative Du, with arbitrary coefficients. Next, we also consider the non-Markovian case, using the recent theory on the path-dependent PDEs. To consider the case with Du in the nonlinearity, a key idea is to introduce a Malliavin type weight function in the representation formula. Moreover, restricting to linear case and using a freezing coefficients technique, one can obtain an unbiased simulation method for the SDEs. Finally, for semilinear PDE with general Lipschitz generator, we consider a polynomial approximation of the generator to apply the branching diffusion method. To achieve this, more techniques and ideas are used to control the variance explosion problems.

Version Française

Ce document est une synthèse de mes travaux de recherche depuis la soutenance de ma thèse. Le plupart de ces travaux concerne la théorie du contrôle stochastique (sans ou avec des contraintes) et les équations aux dérivées partielles (dépendant des chemins), ainsi que leurs applications en finance. Néanmoins, je vais les regrouper en trois thèmes selon leurs applications et les techniques utilisées :

(i) Problème de transport optimal martingale.

(ii) Contrôle stochastique Non-Markovien, équation d'HJB dépendant des chemins.

(iii) Représentation des EDP nonlinéaires par le processus de branchement.

Dans la suite, je vais développer les trois thèmes avec plus de détails et discuter également les liens entre eux.

Problème de Transport Optimal Martingale

Le problème de transport optimal martingale est un thème de recherche très récent dans les mathématiques financières. Pendant les dernières années, il est devenu rapidement un sujet populaire dans cette communauté. Le problème a été initialement formulé dans Beiglböck, Henry-Labordère et Penkner [18], et Galichon, Henry- Labordère et Touzi [99], motivé par ses applications en finance. Plus précisément, il s'agit de la recherche du coût minimal de sur-réplication des options exotiques dépendant de chemin en utilisant la stratégie semi-statique, i.e. la stratégie dynamique en actif sous-jacent et la stratégie statique en options vanilles. En considérant des semi-martingales contrôlées au lieu des martingales, ma thèse (2009-2011) a porté principalement sur un problème similaire (voir Tan et Touzi [192]). Antérieurement, ce problème a été formulé et étudié par Hobson [119] avec l'approche du plongement de Skrokohod.

Dans la modélisation classique de finance de marché, on reste dans un context avec un espace de probabilité fixé, où le modèle ou la distribution des processus sous-jacent est fixé. Ensuite, dans ce modèle fixé, on peut analyser le risque financier, évaluer les produits dérivées, etc. Lorsque le marché est parfait, i.e. tous les produits dérivés peuvent être répliquer parfaitement par une stratégie auto-finançante, et sous la condition de l'absence d'opportunité d'arbitrage (**AOA**), le prix d'un produit dérivée est égale à son coût de réplication. D'un autre point de vue, ce prix pourrait aussi être obtenu comme l'espérance du payoff actualisé sous la probabilité risque-neutre, qui est l'unique mesure martingale équivalente à la mesure de référence. Ici, une mesure martingale représente une mesure de probabilité sous laquelle les actifs sous-jacents actualisés sont des martingales. Lorsque le marché n'est plus parfait, la mesure martingale équivalente n'est pas unique. Néanmoins, par le théorème fondamental d'évaluation des actifs, l'espérance du payoff sous une mesure martingale équivalente donne toujours un prix sans arbitrage. En outre, lorsqu'une option n'est pas réplicable, on peut considérer les stratégies de sur-réplication. Le théorème de **dualité** montre que le coût minimal de sur-réplication est égale au maximum des prix sans arbitrage.

Plus récemment, surtout depuis la crise financière 2008, le risque de modèle devient un sujet de plus en plus important en finance. Pour dépasser cette difficulté, une solution très populaire est de considérer une famille de modèles au lieu d'un modèle de référence fixé. En particulier, pour une option, on doit considérer un prix indépendant du modèle. Ceci est donné par son coût minimal de sur-réplication, où la sur-réplication est au sense quasi-sûre, i.e. presque-sûre sous chaque probabilité dans la famille.

Cette approche robuste simple pourrait donner un prix très élevé pour une option simple, ou même déraisonnable sur le marché financier. Néanmoins, avec le développement de marché financier, les options simple (ou les “options vanilles”) deviennent très liquides, telles que l'on pourrait les utiliser dans une stratégie de réplication. Alors, on pourrait espérer obtenir un prix de sur-réplication plus raisonnable, en considérant les stratégies semi-statiques, i.e. la stratégie dynamique sur les actifs sous-jacent, plus la stratégie statique sur les options vanilles. D'un autre point de vue, lorsque l'on utilise les mesures martingale pour obtenir le prix sans arbitrage, il faut considérer celles cohérentes avec les informations du marché données par les prix des options vanilles. Il est bien connu que, dans le cas limite où les prix des options vanilles sont donnés pour tous les “strikes” à une certaine date de maturité, on pourrait récupérer la loi marginale du sous-jacent à cette date. En conclusion, la borne supérieure du prix sans arbitrage d'une option exotique est donné par le supremum d'espérance de son payoff actualisé sous toutes les mesures martingales satisfaisantes une contrainte de loi marginale. Cela donne le nom: “problème de transport optimal martingale”, en rappelant que le problème de transport optimal classique consiste à optimiser une fonction de coût sur tous les vecteurs aléatoires satisfaisants des contraintes de loi marginale.

Pour ce nouveau problème de transport optimal, les questions naturelles sont: est-ce que la dualité “évaluation-réplication” reste toujours vraie, est-ce que la solution optimale existe pour le problème d'évaluation et pour le problème de surréplication, comment caractériser les solutions optimales, comment les approcher numériquement lorsqu'une solution optimale n'est pas explicitement disponible. Ces questions ne sont pas indépendantes, mais sont plutôt reliées les unes aux autres. En outre, les applications de ce nouveau thème ne seront pas limitées à la finance, mais aussi aux autres thèmes en probabilité, e.g. le problème de plongement de Skorokhod, les inégalités maximales de martingale, PCOC (processus croissant pour l'ordre convexe), etc.

Une grande partie de ma contribution dans ce sujet consiste à prouver la dualité “évaluation-réplication” dans des contextes différents: le cas avec trajectoire continus/càdlàg des sous-jacent, le cas limite du problème de transport optimal martingale avec un nombre infini de contraintes marginales, le cas en temps discret pour les options de type Américain ou avec le coût de transaction. En outre, j'étudie également la limite des plans de transport martingale optimaux en temps discret lorsque la discrétisation de temps passe à zéro. Ceci donne en particulier une nouvelle classe de solution de PCOC en temps continu. Je propose aussi une preuve alternative du principe de la monotonie pour le problème de plongement de Skorokhod optimal, à partir de la dualité et du théorème de section optionnelle. Enfin, je travaille également sur l'approximation numérique du problème de transport optimal martingale en utilisant le résultat de dualité.

Contrôle Non-Markovien, Equation d'HJB dépendant des chemins

Bien que le problème de transport optimal martingale est nommé comme un problème en formulation du problème de transport optimal classical, il peut également être vu comme un problème de contrôle stochastique sous contrainte de marginale. En particulier, ce point de vue nous permet de récupérer les résultats et les techniques de la théorie de contrôle optimal. Puisque le fonction de coût pour un problème de transport optimal martingale est souvent non-Markovienne, une meilleure compréhension du contrôle non-Markovien est nécessaire. Au même temps, le développement de la finance, de la théorie de jeux, du problème de principal-agent, etc. donnent des motivations supplémentaires pour étudier les problèmes de contrôle non-Markovien.

La théorie du contrôle stochastique est largement développée depuis les années 1970, par l'approche EDP ainsi que l'approche probabiliste. Lorsque la fonction valeur du problème n'est pas assez régulière, c'est souvent la notion de solution de viscosité qui permet de la caractériser comme solution de l'équation HJB. Remarquons que l'approche d'EDP est généralement limitée au le cas Markovien, la notion des EDP dépendant des chemins est une extension qui permet d'étudier les problèmes non-Markoviens. Le principe de la programmation dynamique est toujours un outil important, qui permet de décomposer un problème d'optimisation globale en problèmes d'optimisation locaux. Avec le développement des mathématiques financières en 1990, cette théorie a reçu une nouvelle attention pour ses applications en finance. Plus récemment, la théorie des EDSR (Equations Différentielles Stochastiques Rétrogrades) ainsi que des EDSR du second ordre ont été développés et constituent des outils probabilistes importants pour étudier les problèmes de contrôle Markovien ou non-Markovien.

D'un point de vue théorique, la théorie du contrôle stochastique fourni des representations probabilistiques pour les EDP nonlinéaires, qui peut être considérée comme une extension de la formule de Feynmann-Kac pour les EDP linéaires. Ce point de vue est encore généralisé par les EDP dépendant des chemins, qui constitue une approche unifiée pour l'EDSR, l'EDSR du seconde ordre, le control stochastique (Markovien ou non-Markovien), les jeux différentiels, etc.

Mes travaux en contrôle stochastique se situe principalement dans un contexte non-Markovien. Dans un premier temps, je travaille sur le principe de la programmation dynamique. L'idée principale est de donner une analyse/revue de la littérature, ainsi qu'une justification rigoureuse et détaillée du principe de la programmation dynamique pour des formulations différentes du problème de contrôle. Ici, j'utilise le mot "rigoureuse" car la mesurabilité et les ensembles nuls dans ce contexte sont des sujets très subtiles, lorsque des mesures différentes et singulières interviennent. Un outil essentiel que l'on utilise est le théorème de sélection mesurable. Dans un deuxième travail, on étend le principle de la programmation dynamique pour un problème de contrôle d'une famille d'EDSR. Comme application, nous obtenons une extension de la théorie des EDSR du second ordre. Dans ce travail, comme pour d'autres problèmes de contrôle, le principe de la programmation dynamique permet de characteriser le processus fonction valeur comme une sur-martingale sous une famille de probabilités. On s'intéresse ensuite à la décomposition de Doob-Meyer des sur-martingales. Au final, je m'intéresse également à l'approximation numérique du problème de contrôle non-Markovien. Pour résoudre ces problèmes en temps continu, une première étape consiste à discrétiser le temps pour le transformer en un problème d'optimisation en temps discret, qui peut être résolu numériquement. Pour prouver la convergence des problèmes du temps discret au temps continu dans le contexte non-Markovien, une première approche est d'utiliser la méthode de convergence faible de Kushner-Dupuis. En outre, avec la technique du principe d'invariance forte, on pourrait obtenir un taux de convergence sous des conditions techniques. Une deuxième approche consiste à étendre la méthode de Barles-Souganidis des schémas monotone pour les EDP dans le cas non-Markovien avec les techniques d'EDP dépendent des chemins.

Représentation des EDP nonlinéaires par le processus de branchement

Il est très naturel et classique de discrétiser le temps des problèmes en temps continu, pour obtenir un schéma numérique. Néanmoins, pour les problèmes de contrôle ou les équations nonlinéaires, une grande limite de cette approche est que la condition de monotonie (qui est une condition nécéssaire pour assurer la convergence) est difficile à obtenir en grande dimension. De plus, pour certains schémas de Monte-Carlo nonlinéaires, même si la condition de monotonie est assurée, il existe des termes d'espérance conditionnelle à estimer. Cette estimation est souvent basée sur des techniques de simulation-regression, qui est en fait une variation de la méthode des éléments finis, et donc est également limitée en petite dimension.

Une autre méthode numérique pour les EDP nonlinéaires, qui pourrait dépasser cette limite de condition de monotonie, ainsi que les étapes de regression, consiste à utiliser les processus de branchement diffusion. Effectivement, il est bien connu, dans les anciens travaux de Watanabe, Skorokhod, McKean, etc., que les équations de KPP peuvent être représentées par les processus de branchement. Une EDP nonlinéaire est dite de type KPP lorsque la partie nonlinéaire est un polynôme de la fonction valeur, avec des coefficients positives de somme 1. Comme extension de la formule de Feynmann-Kac pour les équations linéaires, la solution de l'équation KPP peut être représentée comme l'espérance d'une fonctionnelle d'un processus de branchement. Cette représentation a été également généralisée par le sur-processus comme limite des processus de branchement. Avec cette représentation, on peut estimer la solution de l'équation KPP par la méthode de Monte Carlo en simulant un grand nombre de processus de branchement, puis calculer la moyenne de la fonctionnelle correspondante.

Mes travaux dans ce sujet consiste principalement à explorer les applications de processus de branchement pour la résolution numérique des EDP nonlinéaires. Dans un premier travail, on cherche à étendre la représentation classique pour une classe des équations nonlinéaires plus large que les EDP de type KPP. Cette classe plus large inclut des EDP dépendant des chemins, ainsi que les EDP nonlinéaires avec une nonlinéarité en u et ses dérivées Du. Pour considérer le cas avec Du, l'idée essentielle est d'introduire un poids de type Malliavin dans la formule de représentation. Ensuite, on revient au cas des EDP linéaires, en utilisant une technique qui consiste à geler les coefficients, cette représentation donne une nouvelle technique de simulation sans biais pour les EDS. Enfin, pour les EDP semilinéaire avec un générateur Lipschitz, on peut considérer une approximation polynomiale du générateur, et puis utiliser la méthode des processus de branchement. Des nouvelles idées et techniques sont utilisées pour contrôler la variance des estimateurs.

Chapitre 2

Martingale Optimal Transport and its Applications in Finance

2.1 Motivation from finance

The martingale optimal transport (MOT) is a novel subject in mathematical finance, while its classical counterpart, the optimal transport (OT), has been an important and inspiring mathematical topic since 18th century, see Monge [156]. Although initially motivated by its applications in finance, it goes beyond the mathematical finance and provides also new results in other branches of mathematics. The problem has been initially formulated by Beiglböck, Henry-Labordère and Penkner [18] for a discrete time version, and by Galichon, Henry-Labordère and Touzi [99] for a continuous time version. Similar problems have also been formulated and studied using the Skorokhod embedding problem (SEP) approach, as initiated by Hobson [119]. Since then, this topic attracts many researchers with different background and skills and generates a large stream of publications during the last years.

Let us start by formulating the problem with its initial financial motivation. In the classical theory of financial mathematics, the financial market is modelled with a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$, in which one has the underlying assets as stochastic processes, denoted by $X = (X_t)_{t\in\mathbb{T}}$. Here $\mathbb{T}$ refers to the set of indexes of time, which could be an interval $[0, T]$ for a continuous time model, or $\{0, 1, \cdots, T\}$ for a discrete time model. Then under this model, one could analyse the financial risk, evaluate the derivative options, etc. At the same time, the No Arbitrage (NA) condition is an essential rule in the option pricing theory. The first fundamental theorem of asset pricing (FTAP) says that the existence of equivalent martingale measure is a (fundamental) sufficient condition[1] to ensure the NA condition. Here the martingale measure means a probability measure on $(\Omega, \mathcal{F})$ under which the discounted value of each risky asset is a martingale. In other words, if a new asset is priced to be a martingale (after discounting) under an equivalent **martingale** measure $\mathbb{Q}$, then one keeps the NA condition. Consequently, for a derivative option with discounted payoff ξ, the value $\mathbb{E}^{\mathbb{Q}}[\xi]$ provides a no-arbitrage price. Moreover, if one considers all equivalent martingale measures to compute the upper bound of the no-arbitrage prices for an option, it is related to the minimal super-hedging cost of the option. This is the so-called pricing/hedging duality.

In practice, one usually looks for a martingale model, where its coefficients could be easily calibrated in order to fit "the best" to the market data. Here the market data refers generally to the call/put vanilla options' price since these options are the most liquid in the financial market. In practice, the classical models such as Dupire's [78] local volatility

[1] In the discrete time market, it is a necessary and sufficient condition.

model and the stochastic volatility model are both calibrated in this way. Moreover, it is observed that the Dupire's volatility function $\sigma(t,x)$ is in fact the conditional expectation of the stochastic volatility σ_t knowing the underlying asset value $X_t = x$. From a mathematical point of view, Gyöngy's [109] theorem (see also [46]) justifies that Dupire's model is the unique Markovian SDE model that provides the same one-dimensional marginal distribution of the underlying process as that of the stochastic volatility model. From an economic point of view, Breeden and Litzenberger [43] have observed that the prices of the vanilla options allow to recover the one-dimensional marginal distribution of the underlying in the limit case where the option prices are available for all possible strikes K. Therefore, to be consistent with the market information, one should consider those equivalent martingale models which provide the same vanilla option prices on the market, or equivalently those martingale models satisfying some one-dimensional marginal distribution constraint μ.

More recently, especially since the 2008 financial crisis, the risk of model misspecification has been more and more highlighted. A modern approach to address the model risk is the so-called robust approach, that is, instead of a fixed model, one considers at the same time a family of models. In the extreme case, this family could contain all possible models. Mathematically, this means that one considers a family $\mathcal{P}$ of probability measures on $(\Omega,\mathcal{F})$ at the same time. In particular, when $\mathcal{P}$ contains all Dirac measures δ_ω, for each $\omega \in \Omega$, it leads to the so-called model free pricing approach.

In summary, to eliminate the model risk and to stay consistent with the market information, one should consider all possible martingale measures $\mathbb{Q}$ on $(\Omega,\mathcal{F})$ satisfying the marginal distribution constraint $X_T \sim^{\mathbb{Q}} \mu$, where μ is obtained from the market vanilla options prices. Let us denote this set by $\mathcal{M}(\mu)$. By taking the supremum, one obtains an upper bound of the model free no-arbitrage price of a derivative option with discount payoff $\xi : \Omega \to \mathbb{R}$, that is,

$$P(\mu) \;=\; \sup_{\mathbb{Q}\in\mathcal{M}(\mu)} \mathbb{E}^{\mathbb{Q}}[\xi]. \tag{2.1.1}$$

Notice that the above problem could be viewed as a natural extension of Kantorovich's [134] relaxed formulation of Monge's [156] optimal transport (OT) problem, where the major difference is that the new introduced problem has an additional martingale constraint on $(X_t)_{t\in\mathbb{T}}$. Hence it is called the martingale optimal transport (MOT) problem.

From a hedging point of view, the market will not be perfect in the model uncertainty framework, and a perfect replication is no more possible for a general option ξ. To obtain a safe price, one could consider the super-hedging cost. Moreover, as soon as the vanilla options are liquid on the market, one should also consider using the static strategies. By Carr-Madan formula, any European option with payoff $\lambda(X_T)$ could be represented as a basket of vanilla options, whose cost is given by $\mu(\lambda) := \int_{\mathbb{R}} \lambda(x)\mu(dx)$, where μ is the one dimensional marginal distribution recovered from the vanilla options' prices as above. Let $\mathcal{H}$ be the collection of all admissible dynamic trading strategies on the underlying X, and denote by $(H\circ X)_t$ the P&L of the strategy H at time t, then a semi-static super-hedging strategy is given by $(\lambda, H) \in \mathbb{L}^1(\mu)\times\mathcal{H}$ satisfying $\lambda(X_T) + (H\circ X)_T \geq \xi$. The value

$$D(\mu) \;=\; \inf\big\{\mu(\lambda) \;:\lambda(X_T) + (H\circ X)_T \geq \xi,\ \mathbb{P}\text{-a.s. } \forall \mathbb{P}\in\mathcal{P}\big\} \tag{2.1.2}$$

provides the minimum super-hedging cost of option ξ using semi-static strategies and under model uncertainty. As extension of the classical pricing-hedging duality, one would expect to have the following duality result:

$$P(\mu) \;=\; D(\mu). \tag{2.1.3}$$

Moreover, the above duality result can also be viewed as a natural extension of the Kantorovich's [134] duality in the classical OT theory.

Once the problems are formulated, the questions arising could be:

- whether and under which conditions the duality (2.1.3) holds true?
- how to find/characterize the optimal solutions for both optimization problems $P(\mu)$ and $D(\mu)$?
- how to compute numerically the optimal solutions/values?
- etc.

To finish this motivation section, let us provide two famous theorems which ensure the existence of martingales with given one-dimensional marginal distribution. Given two probability measures μ and ν on $\mathbb{R}^d$ with finite first order moment, we say μ is smaller than ν in convex ordering if $\mu(\phi) \le \nu(\phi)$ for all convex function $\phi : \mathbb{R}^d \to \mathbb{R}$, and denote $\mu \preceq \nu$. Here $\mu(\phi) := \int_{\mathbb{R}^d} \phi(x)\mu(dx)$.

Theorem 2.1.1. (i). **Discrete time case, Strassen's Theorem [187]**: *Let $(\mu_k)_{0\le k\le n}$ be a family of probability measures on $\mathbb{R}^d$. Then there exists a martingale $(M_k)_{0\le k\le n}$ with marginal distributions $(\mu_k)_{0\le k\le n}$ if and only if each μ_k has finite first order moment and $\mu_i \preceq \mu_k$ for all $i \le k$.*
(ii). **Continuous time case, Kellerer's Theorem [137]**: *Let $(\mu_t)_{t\in[0,1]}$ be a family of probability measures on $\mathbb{R}^d$. Then there exists a martingale $(M_t)_{t\in[0,1]}$ with marginal distributions $(\mu_t)_{t\in[0,1]}$ if and only if each μ_t has finite first order moment and $\mu_s \preceq \mu_t$ for all $s \le t$.*

In the following of the chapter, we always assume the increasing condition of $(\mu_k)_{0\le k\le n}$ or $(\mu_t)_{t\in[0,1]}$ in convex ordering.

Assumption 2.1.1. *The family $(\mu_k)_{0\le k\le n}$ (resp. $(\mu_t)_{t\in[0,1]}$) has finite first order, and satisfies $\mu_i \preceq \mu_j$ for all $0 \le i \le j \le n$ (resp. $0 \le i \le j \le 1$).*

2.2 Different formulations and some literature review

The above MOT problem have several variated formulations. First, the financial market could be in discrete time or in continuous time. Next, for the continuous time model, the underlying asset process $(X_t)_{t\in[0,T]}$ could be of continuous path, or of càdlàg path. Further, in place of the one marginal constraint problem, the case with multi-marginals constraints has also its natural motivation as soon as the vanilla options of different maturities are available.

Discrete time MOT The one period discrete time MOT problem should be the formulation the most closed to the classical OT problem. Let $\Omega := \mathbb{R}^d \times \mathbb{R}^d$ be the canonical space, with canonical process $X = (X_0, X_1)$. Then given two probability measures $\mu = (\mu_0, \mu_1)$ on $\mathbb{R}^d$, we denote by

$$\mathcal{P}(\mu_0, \mu_1) \ := \ \big\{\mathbb{P} \in \mathfrak{B}(\Omega) \ : \mathbb{P} \circ X_k^{-1} = \mu_k, \ k = 0, 1\big\}, \tag{2.2.1}$$

and

$$\mathcal{M}(\mu_0, \mu_1) \ := \ \big\{\mathbb{P} \in \mathcal{P}(\mu_0, \mu_1) \ : \mathbb{E}^{\mathbb{P}}[X_1 | X_0] = X_0\big\}.$$

A basic discrete time MOT is given by, for some $\xi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$,

$$\sup_{\mathbb{P} \in \mathcal{M}(\mu_0, \mu_1)} \mathbb{E}^{\mathbb{P}}[\xi(X_0, X_1)]. \tag{2.2.2}$$

Under the above formulation, the duality result (2.1.3) has been proved initially in Beiglböck, Henry-Labordère and Penkner [18] for upper semi-continuous reward functions ξ. In their dual formulation (2.1.2), $\mathcal{P}$ contains all Dirac measures $\delta_{(x,y)}$ for $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$, so that the super-replication is in fact in the pathwise sense. In the one-dimensional case, by setting $\mathcal{P}$ as $\mathcal{M}(\mu_0, \mu_1)$ in the dual formulation (2.1.2), the duality (2.1.3) is proved for any Borel positive reward function ξ by Beiglböck, Nutz and Touzi [21]. As for the characterization of the optimal martingale transport plan, Beiglböck and Juillet [20] obtained a monotonicity principle for this one-period MOT problem. It provides a variational calculus characterization of the support of an optimal transport plan. In particular, it extends the basic idea of the monotonicity principle of the classical OT problem, that is, one cannot improve an optimal transport plan by switching the connection of pairs, see e.g. Villani [196]. For some reward function, it is further proved that the optimal martingale transport plan should support on a special binomial tree structure, named "left-monotone" transport plan. This structure has then been explicitly computed out by Henry-Labordère and Touzi [114], see also Beiglböck, Henry-Labordère and Touzi [19] for a simplified proof. Effort to extend these results to the high-dimensional case, or multi-periods case has been made recently in [64, 159, 162], etc.

Bouchard-Nutz's discrete time framework Bouchard and Nutz [35] introduced a discrete time framework, which is not exactly as that of the MOT problem, but is still in the same spirit. Let $\Omega_0 = \{\omega_0\}$ be a singleton, Ω_1 be a Polish space, Ω_1^t denote the t-fold Cartesian product of Ω_1, then one sets $\Omega_t := \Omega_0 \times \Omega_1^t$. For each t and $\omega \in \Omega_t$, there is a given non-empty convex probability measures set $\mathcal{P}_t(\omega)$, which represents all possible law on Ω_{t+1} knowing $\omega \in \Omega_t$. Then set

$$\Omega := \Omega_T, \qquad \mathcal{P} := \big\{\mathbb{P} := \mathbb{P}_0 \otimes \mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_{T-1} \ : \mathbb{P}_t(\cdot) \in \mathcal{P}_t(\cdot)\big\}.$$

In above, $\mathbb{P}_t(\cdot)$ is a measurable probability kernel from Ω_t to Ω_{t+1}. As a first main result, they introduced a notion of $\mathcal{P}$-quasi-sure no-arbitrage (NA) condition, and extended the fundamental theorem of asset pricing (FTAP) to this model uncertainty setting, that is, NA is equivalent to the existence of a certain family of martingale measures. Next, we consider the super-hedging problem (2.1.2) in the market with presence of finitely many static options. A pricing-hedging duality in form of (2.1.3) has also been obtained for all options with upper semi-analytic measurable payoff. Thanks to the product structure, one can first establish the results in the one-period case, and then extend them to the whole space using a classical dynamic programming argument. This framework has then been adopted by many other authors to study different problems, such as transaction cost problems [36, 12, 33], utility maximisation problems [10], American option pricing [1], etc.

Continuous time MOT or optimal SEP For the continuous time MOT problem, there are two main cases: the continuous path case and the càdlàg path case, with respectively canonical space $\Omega = \mathbb{C}([0,1],\mathbb{R}^d)$ and $\Omega = \mathbb{D}([0,1],\mathbb{R}^d)$. Let $X = (X_t)_{0\le t\le 1}$ denote the canonical process. Given two marginal distributions $\mu = (\mu_0, \mu_1)$, one defines

$$\mathcal{M}(\mu_0,\mu_1) \;:=\; \big\{\mathbb{P}\;: X \text{ is a } \mathbb{P}\text{-martingale}, \quad X_0 \sim^{\mathbb{P}} \mu_0 \; X_1 \sim^{\mathbb{P}} \mu_1\big\},$$

and obtains a MOT problem

$$\sup_{\mathbb{P}\in\mathcal{M}(\mu_0,\mu_1)} \mathbb{E}^{\mathbb{P}}\big[\xi\big], \quad \text{for some } \xi : \Omega \to \mathbb{R}. \tag{2.2.3}$$

The initial formulation of Galichon, Henry-Labordère and Touzi [99] has been given in this way. Tan and Touzi [192] studied the problem by considering the class of laws of Itô processes with given marginal distribution instead of $\mathcal{M}(\mu_0,\mu_1)$, and obtain a duality result as extension of the initial work in Mikami and Thieullen [155]. Moreover, a numerical scheme has been obtained in [192] based on the duality result. However, a first complete and general duality result in this framework has been obtained by Dolinsky and Soner [72] based on a discretization technique. In their dual formulation (2.1.2), the strategy H are chosen to be of finite variation and hence the integrable $(H\circ X)_1$ is defined path by path and hence the super-replication is also pathwise. Extensions of this work to the multi-marginals case, to the càdlàg paths' case, etc. have been obtained in [73, 127, 105], etc.

For the case with continuous path $\Omega = \mathbb{C}([0,1],\mathbb{R}^d)$ with $d = 1$ and when the reward function is time invariant, the MOT problem can be reformulated to be an optimal Skorokhod embedding problem (SEP) by a time changing argument. A basic SEP is usually formulated as follows: given a probability measures μ and a Brownian motion W, one searches a stopping time τ such that $W_\tau \sim \mu$ and $(W_{\tau\wedge t})_{t\ge 0}$ is uniformly integrable. Such a problem has many explicitly constructive solution, including the most famous Root solution [181], Rost solution [182], Azéma-Yor solution [4], Perkins solution [168], Vallois solution [195], etc. see e.g. Obloj [161], Hobson [120] for a survey. An optimal SEP consists in finding an optimal solution of SEP w.r.t. some reward function. To see how the MOT could be reformulated as an optimal SEP, let us first recall Dubins-Schwarz time change theorem (see e.g. [135] or [178]): given a martingale measure $\mathbb{P}$ on Ω, under which X is a martingale measure, then in a possibly enlarged space, one has a Brownian motion W such that $X_t = W_{\langle X\rangle_t}$, $\mathbb{P}$-a.s. Moreover, the quadratic variation $\langle X\rangle_t$ of X is a stopping time w.r.t. the time changed filtration. Suppose that the payoff function ξ satisfies $\xi = \Phi(\langle X\rangle_t, W_\cdot)$ $\mathbb{P}$-a.s. under each martingale measure $\mathbb{P}$, for some $\Phi : \mathbb{R}_+\times\mathbb{C}(\mathbb{R}_+,\mathbb{R}) \to \mathbb{R}$, then it is easy to use Dubins-Schwarz Theorem to deduce that the MOT problem (2.2.3) is equivalent to the optimal SEP

$$\sup_{\tau \in \mathcal{T}(\mu)} \mathbb{E}\big[\Phi(\tau, W_\cdot)\big], \quad \text{with } \mathcal{T}(\mu) := \{\text{All embedding stopping times}\}. \tag{2.2.4}$$

This link has been observed and used by Hobson [119] to study the upper bound of the no-arbitrage prices for a lookback option, in a market with liquid vanilla options. Since then, many works has been devoted to this problem, using the optimal SEP approach, see e.g. [54, 55, 56, 57, 58, 62, 100, 122, 123, 124, 125], etc. Extension of the SEP to the multi-marginals case has also been obtained by [45, 112], etc. In most of these works, the basic idea is to consider some specific constructive solutions of the SEP and the show their optimality for some specific payoff functions. More recently, Beiglböck, Hussmann and Cox [16] formulated a monotonicity principle for the optimal SEP, a concept borrowed from the OT theory, and provide a geometric description of the support of an solution of the optimal SEP. In particular, it provides a unified explanation to the optimality of different well known embedding solutions.

Multi-marginal constraints towards Peacock Although in most cases, the problems and results are given for the one marginal case. The same problem with multi-marginals constraints is also natural and interesting by its financial motivation. Moreover, it would be theoretically interesting/beautiful to consider the limit case when one has a continuous family of marginals $\mu = (\mu_t)_{t\in[0,1]}$. In this case, the family $\mu = (\mu_t)_{t\in[0,1]}$ should have finite first order moment and be increasing in convex ordering as shown in Kellerer's Theorem. A process with one-dimensional marginal distributions increasing in convex ordering is called a "Processus Convex en Ordre Convex" (PCOC, so Peacock with a little imagination) in French. The question of constructing martingale that has the same one-dimensional marginal distributions as a Peacock has been interested by many peoples, for which let us refer to the book of Hirsch, Profeta, Roynette and Yor [118]. From the MOT problem point of view, it would be more interesting to find/construct the martingale Peacocks that enjoy some optimality. In Madan and Yor [150], an explicitly condition is obtained to ensure that the Azéma-Yor embedding solution τ_t^{AY} w.r.t. to different μ_t are automatically ordered, and hence the process $(W_{\tau_t^{AY}})_{t\in[0,1]}$ becomes a martingale Peacock. It is a Markov process, whose generator could be explicitly computed, and moreover, it enjoys implicitly an optimality property as Azéma-Yor embedding. In Henry-Labordère, Tan and Touzi [114], by taking the limit the binomial tree (left-monotone) martingale that obtained in Beiglböck and Juillet [20] and Henry-Labordère and Touzi [117], a martingale Peacock is also obtained, which is also a Markov process and is optimal w.r.t. a class of reward functions. Hobson [121] obtained an explicit solution of martingale Peacock that minimizing the expectation of the total variation. Källblad, Tan and Touzi [133] studied a general optimal SEP as well as the specific lookback option case, given a continuous time family of marginal constraints.

In the rest of the chapter, I will concentrate mainly on the contributions that I made around this topic. To unify and simplify the presentation, the results provided here may be less general than that given in the papers, and the proofs are usually provided in a heuristic and non-rigorous way.

2.3 Duality results

A large part of my work on MOT consists in establishing the duality result in different frameworks and by different techniques. In the classical optimal transport theory, the Kantorovich duality relies intuitively on a minimax argument, or more precisely is obtained by Fenchel's duality theorem. To illustrate the idea, let us recall it briefly. Given two probability measures μ_0, μ_1 on $\mathbb{R}^d$, recall the definition of $\mathcal{P}(\mu_0, \mu_1)$ in (2.2.1).

Theorem 2.3.1 (Kantorovich's duality of OT). *Assume that $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is upper semicontinuous and bounded from above. Then one has*

$$\begin{aligned} P(\mu_0, \mu_1) \; &:= \sup_{\mathbb{P} \in \mathcal{P}(\mu_0,\mu_1)} \mathbb{E}^{\mathbb{P}}\big[c(X_0, X_1)\big] \\ &= \inf\big\{\mu_0(\lambda_0) + \mu_1(\lambda_1) \;\; : \lambda_0(x) + \lambda_1(y) \geq c(x,y), \; \forall (x,y) \in \mathbb{R}^d \times \mathbb{R}^d\big\}. \end{aligned}$$

Proof. (i) First, it is easy to see that the function $(\mu_0, \mu_1) \mapsto P(\mu_0, \mu_1)$ is concave by its definition. Next, let (μ_0^n, μ_1^n) be a sequence of probability measures converges weakly to (μ_0, μ_1). Notice that $\mathcal{P}(\mu_0^n, \mu_1^n)$ is compact and hence there is some optimal $\mathbb{P}_n$ for the problem $P(\mu_0^n, \mu_1^n)$ so that $P(\mu_0^n, \mu_1^n) = \mathbb{E}^{\mathbb{P}_n}\big[c(X_0, X_1)\big]$. The sequence $(\mathbb{P}_n)_{n\geq 1}$ can be easily shown to be relatively compact and to converges along some subsequence to some $\mathbb{P} \in \mathcal{P}(\mu_0, \mu_1)$. Therefore, one has

$$\limsup_{n\to\infty} P(\mu_0^n, \mu_1^n) = \limsup_{n\to\infty} \mathbb{E}^{\mathbb{P}_n}\big[c(X_0, X_1)\big] \;\leq\; \mathbb{E}^{\mathbb{P}}\big[c(X_0, X_1)\big] \;\leq\; P(\mu_0, \mu_1).$$

This implies that $(\mu_0, \mu_1) \mapsto P(\mu_0, \mu_1)$ is concave and upper semi-continuous.

(ii) Next, by Fenchel's duality theorem, a concave and upper semi-continuous function is equal to its bi-conjugate, see e.g. [28, 69]. Applying this on $P(\mu_0, \mu_1)$, it follows by direct computation that its bi-conjugate function is the dual problem $\inf\big\{\mu_0(\lambda_0) + \mu_1(\lambda_1) \;:\; \lambda_0(x) + \lambda_1(y) \geq c(x,y), \; \forall (x,y) \in \mathbb{R}^d \times \mathbb{R}^d\big\}$, where $\lambda_0, \lambda_1 \in C_b(\mathbb{R}^d)$. □

Remark 2.3.1. (i) *We notice that a key argument used in the above proof is that $\mathcal{P}(\mu_0, \mu_1)$ is compact. More important, we use the fact that a sequence $(\mathbb{P}_n)_{n\geq 1}$ with $\mathbb{P}_n \in \mathcal{P}(\mu_0^n, \mu_1^n)$ is tight and hence relatively compact, when $(\mu_0^n, \mu_1^n) \to (\mu_0, \mu_1)$.*

The argument still works for the discrete time MOT problem, but encounters a major difficulty to study the continuous time MOT problem, where the compactness/tightness of laws on the space of continuous time paths is much more difficult to obtain.

(ii) *In Theorems 2.3.2, 2.3.3 and 2.3.4 below, we will use different approaches to overcome this difficulty. We also emphasis that the results in Theorems 2.3.2, 2.3.3 and 2.3.4 will be presented with two marginals $\mu = (\mu_0, \mu_1)$ constraint. Nevertheless, by the same or similar arguments, they can be easily extended to the multi marginals $\mu = (\mu_0, \mu_1, \cdots, \mu_n)$ case, or the infinitely many marginals $\mu = (\mu_t)_{0\leq t\leq 1}$ case.*

The continuous time MOT (or semi-martingale optimal transport) Let us now consider the continuous time MOT problem formulated in (2.2.3), with canonical space $\Omega = \mathbb{C}([0,1], \mathbb{R}^d)$ or $\Omega = \mathbb{D}([0,1], \mathbb{R}^d)$. Dolinsky and Soner [72, 73] proved the duality result for both cases under different technical conditions, where the basic idea is to discretize the canonical space Ω into a countable space. The tightness of laws on the countable space could be easily obtained, and so is the duality result. Then it is enough to use an approximation argument to obtain the duality on Ω.

We studied the problem with different approaches, by trying to obtain the tightness of martingale transport plans in a different way. In a first work in Tan and Touzi [192], we consider a general semi-martingale optimal transport problem as extension of the work in Mikami and Thieullen [155]. Concretely, let $\Omega = C([0,1],\mathbb{R}^d)$ be the canonical space, X the canonical process, μ_0 and μ_1 two probability measures on $\mathbb{R}^d$, we set $\mathcal{SM}(\mu_0,\mu_1)$ as collection of all probability measures $\mathbb{P}$ on Ω such that $\mathbb{P}\circ X_i^{-1} = \mu_i$, $i = 0,1$ and that X admits a semi-martingale decomposition, under $\mathbb{P}$,

$$dX_t \;=\; b_t^{\mathbb{P}} dt \;+\; \sigma^{\mathbb{P}} dW_t^{\mathbb{P}}, \tag{2.3.1}$$

for some predictable process $(b_t^{\mathbb{P}}, \sigma_t^{\mathbb{P}})_{t\in[0,1]}$ and a Brownian motion $W^{\mathbb{P}}$.

Theorem 2.3.2. *Suppose that $L : [0,1]\times\Omega\times\mathbb{R}^d\times\mathbb{S}^d \to \mathbb{R}$ is bounded from above and satisfies*

$$|(b,a)|^p \le C_p(1+|L(t,\omega,b,a)|) \;\; \textit{for some constant } p>1 \textit{ and } C_p>0, \tag{2.3.2}$$

together with some technical regularity conditions. Then one has the duality:

$$\sup_{\mathbb{P}\in\mathcal{SM}(\mu_0,\mu_1)} \mathbb{E}^{\mathbb{P}}\Big[\int_0^1 L\big(t,\omega,b_t^{\mathbb{P}},(\sigma\sigma^\top)_t^{\mathbb{P}}\big)dt\Big] \;=\; \inf_{\lambda_1\in C_b(\mathbb{R}^d)} \big(\mu_0(\lambda_0)+\mu_1(\lambda_1)\big),$$

where, by denoting $\mathcal{SM}(\mu_0)$ the collection of all semi-martingale measures $\mathbb{P}$ under which X admits the decomposition (2.3.1) and such that $\mathbb{P}\circ X_0^{-1} = \mu_0$,

$$\lambda_0(x) := \sup_{\mathbb{P}\in\mathcal{SM}(\delta_x)} \mathbb{E}^{\mathbb{P}}\Big[\int_0^1 L(t,X,b_t^{\mathbb{P}},(\sigma\sigma^\top)_t^{\mathbb{P}})dt - \lambda_1(X_1)\Big].$$

Proof. Let $(\mu_0^n,\mu_1^n)_{n\ge1}$ be a sequence of probability measures converges weakly to (μ_0,μ_1) and $\mathbb{P}_n \in \mathcal{SM}(\mu_0^n,\mu_1^n)$ be an optimal semi-martingale measures in the transport problems, then thanks to the growth condition (2.3.2), it follows that the sequence $(\mathbb{P}_n)_{n\ge1}$ is tight and any limit lies in $\mathcal{SM}(\mu_0,\mu_1)$. Now, similar to Theorem 2.3.1 for the classical OT, one can deduce that the value of the semi-martingale transport problem is upper semi-continuous (and concave) w.r.t. the marginal distributions, and it follows by Fenchel's duality theorem that one has the above duality. □

Remark 2.3.2. (i) *The main idea in the above proof is to use the growth condition (2.3.2) to push the set of optimal transport plans to be tight. The formulation and main idea of proofs is a direct extension of Mikami and Thieullen [155], where their volatility is a fixed constant.*

(ii) *The above dual problem is not exactly a super-hedging problem in form (2.1.2). To obtain a pricing-hedging duality, one still need to apply the optional decomposition theorem such as in [24] to characterize $\mu_0(\lambda_0)+\mu_1(\lambda_1)$ as a super-hedging cost.*

(iii) *Notice also that we do not consider all martingale measures on Ω, but a class of semi-martingales having an Itô decomposition (2.3.1). A first real and complete pricing-hedging duality in the general MOT setting is proved later by Dolinsky and Soner [72].*

In a second work in Guo, Tan and Touzi [105], we consider a MOT problem with càdlàg underlying processes as in Dolinsky and Soner [73]. Let $\Omega := \mathbb{D}([0,1], \mathbb{R}^d)$ be the canonical space of càdlàg paths, X be the canonical space, $\mathcal{M}$ is the collection of all martingale measures on Ω, and $\mathcal{M}(\mu_0, \mu_1)$ is defined as collection of all martingale measures $\mathbb{P}$ on Ω under which $\mathbb{P} \circ X_t^{-1} = \mu_t$ for $t = 0, 1$, given two marginal distributions $\mu = (\mu_0, \mu_1)$. The most popular topology on Ω is the so-called Skorokhod topology, which is metrizable to make Ω a Polish space. Notice that the tightness of $\mathcal{M}(\mu_0, \mu_1)$ depends on the topology on Ω. Once the topology is defined, the compact subsets of Ω is fixed and then the tightness of laws on Ω is determined. The sparser the topology is, the less the open sets there are and the more the compact sets there are, and hence the easier the tightness of $\mathcal{M}(\mu_0, \mu_1)$ could be obtained. Nevertheless, the price to pay is that there are less continuous functions under a sparser topology. In fact, it is easy to see that $\mathcal{M}(\mu_0, \mu_1)$ is not tight when Ω is equipped with the Skorokhod topology. Our main idea is to consider a sparser topology, the S-topology introduced by Jakubowski [131] (see Appendix for its definition), to recover the tightness and then to save the classical proof as in Theorem 2.3.1. Moreover, one needs to consider the Wasserstein topology on the space of marginal distributions to obtain the upper semi-continuity of the MOT value w.r.t. the marginal distributions. As dual space of the space of marginal distributions on $\mathbb{R}^d$ equipped with the Wasserstein topology, one obtains $\mathcal{C}_1$, the space of all continuous function of linear growth (see Proposition A.1.2 in Appendix).

Theorem 2.3.3. *Assume that $\xi : \Omega \to \mathbb{R}$ is bounded from above and is upper semi-continuous w.r.t. the S-topology (see Appendix for a definition), then one has*

$$\sup_{\mathbb{P} \in \mathcal{M}(\mu_0, \mu_1)} \mathbb{E}^{\mathbb{P}}\big[\xi\big] \;=\; \inf_{\lambda_0, \lambda_1 \in \mathcal{C}_1} \Big\{ \mu_0(\lambda_0) + \mu_1(\lambda_1) + \sup_{\mathbb{P} \in \mathcal{M}} \mathbb{E}^{\mathbb{P}}\big[\xi - \lambda_0(X_0) - \lambda_1(X_1)\big] \Big\}.$$

Assume some further regularity conditions, the value of the above MOT problem equals to

$$\inf \big\{ \mu_0(\lambda_0) + \mu_1(\lambda_1) \;: \lambda_0(X_0) + \lambda_1(X_1) + (H \circ X)_1 \geq \xi, \; \textit{for all } \omega \in \Omega \big\}. \tag{2.3.3}$$

Proof. Again, under the S-topology, one can obtain the tightness of $\mathcal{M}(\mu_0, \mu_1)$. Then following the classical arguments, one can prove that $(\mu_0, \mu_1) \mapsto \sup_{\mathbb{P} \in \mathcal{M}(\mu_0, \mu_1)} \mathbb{E}^{\mathbb{P}}\big[\xi\big]$ is u.s.c., where the space of (μ_0, μ_1) is equipped with the Wasserstein topology. It follows by the Fenchel's duality theorem that the first duality holds. Next, the second duality follows by an optional decomposition theorem for the maximization problem $\sup_{\mathbb{P} \in \mathcal{M}} \mathbb{E}^{\mathbb{P}}\big[\xi - \lambda_0(X_0) - \lambda_1(X_1)\big]$ without marginal constraint, for which we use Dolinsky and Soner's [73] discretization and approximation techniques. □

Remark 2.3.3. (i) *In the above formulation of the dual problem, the dynamic strategy H is of finite variation and the stochastic integral $(H \circ X)_1$ can be defined pathwisely as suggested by Dolinsky and Soner [73]. The condition that ξ is u.s.c. under S-topology in our context is more restrictive comparing to [73], where in the later work, ξ is essentially assumed to be continuous under the Skorokhod topology.*

(ii) *Nevertheless, our dual formulation* (2.3.3) *is more complete than that in [73]. More precisely, our sup-replication in* (2.3.3) *is assumed to hold for every $\omega \in \Omega$, while in [73] it is only assumed to hold for ω that are left continuous at final time* 1.

(iii) *Our approach to prove the duality seems to be more systematic. Without any modification, the same results and proofs hold true for many variated MOT problem: such as multi-marginals case, the case when on replace a marginal distribution μ on $\mathbb{R}^d$ by d marginals $(\mu_1, \cdots, \mu_d)$ on $\mathbb{R}$, and the case with infinitely many marginals, i.e. a Peacock $\mu = (\mu_t)_{0 \le t \le 1}$ is given, which all have natural financial applications.*

The optimal Skorokhod embedding problem (SEP) For a class of time invariant payoff functions, the continuous time MOT with continuous paths can be reformulated as an optimal SEP. The previous problem consists in finding an optimal measure on the space of continuous paths, where the tightness of the measures depends essentially on the regularity property of the paths in the support of the measures. Nevertheless, the SEP consists in finding an optimal stopping time, whose law is a probability measure on $\mathbb{R}$ and the tightness would be much easier to obtain. More precisely, a Skorokhod embedding is a stopping time τ on a Browinian motion W, then by considering the joint law of (W, τ), one can view a Skorokhod as a probability measures on $\mathbb{C}(\mathbb{R}_+, \mathbb{R}) \times \mathbb{R}_+$, where the marginal distribution on the $\mathbb{C}(\mathbb{R}_+, \mathbb{R})$ is the Wiener measure. Although the optimal SEP as well as its applications in finance has been widely studied since Hobson's [119] initial paper, this point to view (to see it as a measure on $\mathbb{C}(\mathbb{R}_+, \mathbb{R}) \times \mathbb{R}_+$) has been initially studied systematically by Beiglböck, Cox and Huesmann [16]. As a first result, they establish a duality for the optimal SEP. Secondly, and more importantly, they introduced a monotonicity principle for the optimal SEP which allows to characterize the support of an optimal embedding solutions, and therefore provides a unified explanation on divers well known optimal embeddings that studied by many authors.

In Guo, Tan and Touzi [106], Källblad, Tan and Touzi [133], we follow this point of view to study the optimal SEP under finitely many or infinitely many marginal constraints. Let $\Omega := C([0,1], \mathbb{R})$ be the canonical space of continuous paths on $[0,1]$ with canonical process X, $\Omega_0 := C(\mathbb{R}_+, \mathbb{R})$ be the canonical space of continuous paths on $\mathbb{R}_+$ with canonical process B, $\mathbb{R}_+$ be the canonical space for stopping times, with canonical element T. Let $\mathcal{T}$ denote the the collection of all probability measures $\mathbb{P}$ on $\Omega_0 \times \mathbb{R}_+$ under which B is Brownian motion and T is a stopping time w.r.t. the same filtration, such that $(B_{T \wedge t})_{t \ge 0}$ is uniformly integrable. Further, given marginal distributions (μ_0, μ_1) on $\mathbb{R}$, $\mathcal{M}(\mu_0, \mu_1)$ is defined as collection of all martingale measures $\mathbb{P}$ on Ω such that $P \circ X_t = \mu_t$, $t = 0, 1$, and $\mathcal{T}(\mu_0, \mu_1)$ denotes the collection of $\mathbb{P} \in \mathcal{T}$ such that $B_0 \sim^{\mathbb{P}} \mu_0$ and $B_T \sim^{\mathbb{P}} \mu_1$. We assume that $\xi := \Omega \to \mathbb{R}$ is a time invariant payoff function in sense that

$$\xi(X) = \Phi(X_{\langle X \rangle_\cdot^{-1}}, \langle X \rangle_1), \tag{2.3.4}$$

for some non-anticipative function $\Phi : \Omega_0 \times \mathbb{R}_+ \to \mathbb{R}$, where the quadratic variation $\langle X \rangle$ as well as its inverse $\langle X \rangle^{-1}$ can be defined pathwisely on Ω, so that $X_{\langle X \rangle_\cdot^{-1}}$ is a Brownian motion under any (local) martingale measure on Ω. As discussed in Section 2.2, we have the equivalence:

$$\sup_{\mathbb{P} \in \mathcal{M}(\mu_0, \mu_1)} \mathbb{E}^{\mathbb{P}}[\xi] = \sup_{\mathbb{P} \in \mathcal{T}(\mu_0, \mu_1)} \mathbb{E}^{P}[\Phi(B, T)].$$

We also notice that for all $\mathbb{P} \in \mathcal{T}$, the marginal distribution on $\mathbb{C}(\mathbb{R}_+, \mathbb{R})$ is fixed, then it is convenient to equip $\mathcal{T}$ with the stable convergence introduced by Jacod and Mémin [129] (see its definition in Appendix).

Theorem 2.3.4. *Assume that $\Phi : \Omega_0 \times \mathbb{R}_+ \to \mathbb{R}$ is non-anticipative, bounded from above and $t \mapsto \Phi(\omega, t)$ is u.s.c. for every $\omega \in \Omega_0$, then*

$$\sup_{\mathbb{P}\in\mathcal{T}(\mu_0,\mu_1)} \mathbb{E}^{\mathbb{P}}[\Phi(B,T)] = \inf\Big\{\mu(\lambda)\ : \lambda_0(B_0) + \lambda_1(B_T) + \int_0^T H_s dB_s \geq \Phi(B,T),\ \forall \mathbb{P} \in \mathcal{T}\Big\}.$$

Consequently, when $\xi : \Omega \to \mathbb{R}$ is given by (2.3.4), one has

$$\sup_{\mathbb{P}\in\mathcal{M}(\mu_0,\mu_1)} \mathbb{E}^{P}[\xi] = \inf\Big\{\mu(\lambda)\ : \lambda_0(X_0) + \lambda_1(X_1) + \int_0^1 H_s dX_s \geq \xi,\ \forall \mathbb{P} \in \mathcal{M}\Big\}.$$

Proof. (i) We equip with $\mathcal{T}$ the stable convergence topology under which $\mathbb{P} \mapsto \mathbb{E}^{\mathbb{P}}[\Phi(B,T)]$ is continuous when $t \mapsto \Phi(\omega,t)$ is continuous. Then using the tightness property of $\mathcal{T}(\mu_0,\mu_1)$, one proves that $(\mu_0,\mu_1) \mapsto \sup_{\mathbb{P}\in\mathcal{T}(\mu_0,\mu_1)} \mathbb{E}^P[\Phi(B,T)]$ is u.s.c. and concave, and it follows by Fenchel's duality theorem that one obtains a first duality:

$$\sup_{\mathbb{P}\in\mathcal{T}(\mu_0,\mu_1)} \mathbb{E}^{P}[\Phi(B,T)] \;=\; \inf_{\lambda_0,\lambda_1\in\mathcal{C}_1} \Big\{\mu(\lambda) + \sup_{\mathbb{P}\in\mathcal{T}} \mathbb{E}^{\mathbb{P}}\big[\Phi(B,T) - \lambda_0(B_0) - \lambda_1(B_T)\big]\Big\}.$$

The supremum problem at r.h.s. is in fact an optimal stopping problem on a functional of the Brownian motion, whose value function can be characterized by its Snell envelop, which is a dominating supermartingale. Using Doob-Meyer decomposition and then the martingale representation theorem, it follows that

$$\mathbb{E}^{\mathbb{P}}\big[\Phi(B,T) - \lambda_0(B_0) - \lambda_1(B_T)\big] = \inf\Big\{x_0\ : x_0 + (H\circ B)_T \geq \Phi(B,T) - \lambda_0(B_0) - \lambda_1(B_T), \forall T\Big\}.$$

Plugging it into the first duality, one obtains the duality result for the optimal SEP.

(ii) For the MOT problem, it is enough to use the time change argument to transform both the primal and dual problem from Ω_0 to Ω. □

Remark 2.3.4. *Our duality result on the optimal SEP provides an extension for the duality result in Beiglböck, Cox and Huesmann [16]. First, we use a complete different approach to prove the duality, where the basic idea and tools are all from the classical optimal control/stopping theory. Such a proof should be more accessible for people familiar with the optimal control/stopping theory. Secondly, by using the stable convergence topology on $\mathcal{T}$, we only assume that $t \mapsto \Phi(\omega,t)$ is a u.s.c., while [16] assumes that $(t,\omega) \mapsto \Phi(\omega,t)$ is u.s.c. Our condition covers the case where Φ is functional of the local time of the underlying process, which is a classical case studied by Vallois's embedding. Finally, our approach works in the same way for the multi-marginals case as well as the infinitely many marginals case, given a Peacock $\mu = (\mu_t)_{0\leq t\leq 1}$.*

The discrete time case for American options Most of the literature on MOT problems, or more generally on the super-hedging problem using semi-static hedging strategies have been dedicated for European type exotic options. It is nevertheless interesting to study the super-hedging problem for American options. In the classical dominated case, when the market is complete, it is well known that an American option's price is given as the supremum of the expected discounted value of the payoff at all stopping times, i.e. the value of an optimal stopping problem. In the non-dominated context, when the perfect hedging is impossible, the super-hedging cost should be formulated as the minimum initial cost to construct a portfolio which dominates the payoff of the option at any possible exercise time. To formulate the pricing problem, a natural guess on the MOT problem for an American option should be the supremum of optimal stopping problems under all possible martingale measures consistent to the market information.

However, in the discrete time framework of Bouchard and Nutz [35], Bayraktar, Huang and Zhou [11] showed that there may exist a duality gap between the pricing problem and the hedging problem for the American options problem. In an unpublished manuscript, Neuberger [157] has considered a discrete time market with discrete state spaces, and then formulated both problems as linear programming problems and then obtained a duality result. This work has been very recently rewritten and extended in Hobson and Neuberger [126]. A insightful result pointed in this paper is that the class of (strong) stopping times w.r.t. the underlying process is not enough to obtain the duality, and one should consider a weak formulation of the optimal stopping problem. Moreover, the weak formulation turns to be more natural as a model-free pricing and super-replication for the American options.

In Aksamit, Deng, Obłój and Tan [1], we study the MOT problem for American options in a more general framework and in a more systematic way. We try to explain why the duality fails and suggest two different approaches to recover the duality, where the first one is in the same spirit as [126] and the second one provides a new point of view on the dynamic trading of vanilla options for super-hedging exotic options.

Concretely, we consider $(\Omega, \mathcal{F})$ as an abstract space, equipped with a discrete time filtration $\mathbb{F} = (\mathcal{F}_t)_{0\le t\le N}$, a underlying adapted process $X = (X_t)_{0\le t\le N}$, and a family $\mathcal{P}$ of probability measures. Further, we are given a family $(g^\lambda)_{\lambda\in\Lambda}$ of static options whose prices are assumed to be 0 without loss of generality. In practice, this could cover different well studied discrete time frameworks:

- The classical dominated case with abstract space $(\Omega, \mathcal{F}, \mathbb{P})$, i.e. $\mathcal{P} = \{\mathbb{P}\}$.
- The Bouchard-Nutz's [35] framework with finite static options, where $\Omega = \Omega_0 \times \Omega_1 \times \cdots \Omega_1$ (as recalled above).
- The Burzoni, Frittelli, Hou, Maggis and Obłój [48] framework with finite static options, where Ω is a Polish space, and $\mathcal{P}$ is the collection of all Borel probability measures on Ω.
- The MOT framework with infinitely many static options, where $\Omega = \mathbb{R}^d \times \cdots \times \mathbb{R}^d$ and $\mathcal{P}$ is the collection of all Borel probability measures on Ω.

An American option is described by its payoff functions $\Phi_k : \Omega \to \mathbb{R}$, $k = 1, \cdots, N$, where Φ_k is the payoff if it is exercised at time k. Our minimal super-hedging cost problem is given by

$$\pi(\Phi) \ := \ \inf \big\{ x \ : x + (H^k \circ X)_N + hg \ge \Phi_k, \ \forall k = 1, \cdots, N, \ \mathcal{P}\text{-q.s.} \big\}. \qquad (2.3.5)$$

In above, $h = (h^\lambda)_{\lambda\in\Lambda}$ is a family with finitely man nonzero elements so that the sum $hg := \sum_{\lambda\in\Lambda} h^\lambda g^\lambda$ is well defined, and each H^k is a $\mathbb{F}$-predictable process such that $H^k_i = H^j_i$

whenever $i \leq j \leq k$. The last condition is to ensure that the dynamic strategies are determined by the information from the filtration $\mathbb{F}$ as well as the fact whether the option is exercised or not. To obtain the robust pricing problem, we first introduce an enlarged space. Let $\overline{\Omega} := \Omega \times \{1, \cdots, N\}$ and $T(\omega, \theta) := \theta$ for all $(\omega, \theta) \in \overline{\Omega}$ be a canonical variable, $\overline{\mathbb{F}} = (\overline{\mathcal{F}}_t)_{0 \leq t \leq N}$ be defined by $\overline{\mathcal{F}}_t := \mathcal{F}_t \vee \big\{\{T \leq k\},\ k = 0, 1, \cdots, t\big\}$. Notice that by the definition of $\overline{\mathbb{F}}$, the variable $T : \overline{\Omega} \to \{1, \cdots, N\}$ is automatically a $\overline{\mathbb{F}}$-stopping time. Let us denote by $\overline{\mathcal{P}}$ the collection of all probability measures $\overline{\mathbb{P}}$ on $\overline{\Omega}$ such that $\overline{\mathbb{P}}|_\Omega \in \mathcal{P}$. Let us denote by $\overline{\mathcal{M}}_g$ the collection of all martingale measures on $\overline{\Omega}$ consistent with the market information, i.e. all probability measures $\overline{\mathbb{Q}}$ dominated by some probability $\overline{\mathbb{P}} \in \overline{\mathcal{P}}$, such that X is a $(\overline{\mathbb{F}}, \overline{\mathbb{Q}})$-martingale, and $\mathbb{E}^{\overline{\mathbb{Q}}}[g^\lambda] = 0$ for all $\lambda \in \Lambda$. Similarly, one can define $\mathcal{M}_g$ as collection of all martingale measures on Ω equipped with filtration $\mathbb{F}$ consistent with the market information, and define $\widehat{\mathcal{M}}_g$ as collection of martingale measures on Ω equipped with and enlarged filtration $\widehat{\mathbb{F}} \supset \mathbb{F}$ which contains additional information from the dynamics of option prices of g^λ. Denote by $\mathcal{T}(\mathbb{F})$ (resp. $\mathcal{T}(\overline{\mathbb{F}})$, $\mathcal{T}(\widehat{\mathbb{F}})$) the collection of all $\mathbb{F}$ (resp. $\overline{\mathbb{F}}$, $\widehat{\mathbb{F}}$) stopping times.

Theorem 2.3.5. *Under different frameworks for $(\Omega, \mathcal{F})$ and $\mathcal{P}$ as listed above with different technical conditions, one has the pricing-hedging duality (and the weak duality):*

$$\pi(\Phi) \;=\; \sup_{\overline{\mathbb{P}} \in \overline{\mathcal{M}}_g} \mathbb{E}^{\overline{\mathbb{P}}}\big[\Phi_T\big] \;\geq\; \sup_{\mathbb{P} \in \mathcal{M}_g} \sup_{\tau \in \mathcal{T}(\mathbb{F})} \mathbb{E}^{\mathbb{P}}\big[\Phi_\tau\big].$$

Moreover, if one is allowed to dynamically trade the options g^λ and to consider the enlarged filtration $\widehat{\mathbb{F}}$, then the minimum super-hedging cost does not change and equals to

$$\sup_{\mathbb{P} \in \widehat{\mathcal{M}}_g} \sup_{\tau \in \mathcal{T}(\widehat{\mathbb{F}})} \mathbb{E}^{\mathbb{P}}\big[\Phi_\tau\big]. \tag{2.3.6}$$

Proof. (i) For the first duality, we can first reformulate the super-hedging problem (2.3.5) equivalently on the enlarged space to be

$$\inf\big\{x\ : x + (\bar{H} \circ X)_N + hg \geq \Phi_T,\ \overline{\mathcal{P}}\text{-q.s.}\big\},$$

where $\bar{H}$ is a $\overline{\mathbb{F}}$-predictable process. The above reformulation turns out to be a super-hedging problem for some European option with payoff function Φ_T defined on the enlarged space $\overline{\Omega}$. Notice that in all the above frameworks, the duality for European options has been established under different technical conditions. Then one can adapt their results and techniques on our enlarged space $\overline{\Omega}$ to deduce the duality.

(ii) When one allows the dynamic trading of the options, one can recover the dynamic programming principle for problem (2.3.6). It follows by the Snell envelop argument that it equals to the pricing problem formulated on $\overline{\Omega}$. □

Remark 2.3.5. (i) *The pricing problem $\sup_{\overline{\mathbb{P}} \in \overline{\mathcal{M}}_g} \mathbb{E}^{\overline{\mathbb{P}}}\big[\Phi_T\big]$ is in fact a weak formulation of an optimal stopping problem. Indeed, take the MOT framework as example, the above pricing problem is equivalent to*

$$\sup_{\alpha} \mathbb{E}^{\mathbb{P}^\alpha}\big[\Phi_{\tau^\alpha}(X^\alpha_\cdot)\big],$$

where $\alpha = (\Omega^\alpha, \mathcal{F}^\alpha, \mathbb{F}^\alpha, \mathbb{P}^\alpha, X^\alpha, \tau^\alpha)$ *is a stopping term, i.e. a filtered probability space equipped with a stopping time* τ^α *and a martingale* X^α *such that* $\mathbb{E}^{\mathbb{P}^\alpha}[g^\lambda(X^\alpha_\cdot)] = 0$. *To see the equivalence, it is enough to notice that any stopping term* α *induces a probability measure in* $\overline{\mathcal{M}}_g$ *and any probability measure in* $\overline{\mathcal{M}}_g$ *is itself a stopping term.*

(ii) *The inequality in Theorem 2.3.5 is the duality gap when one considers (strong) stopping times* $\tau \in \mathcal{T}(\mathbb{F})$, *as observed in [11]. This duality gap is due to the loss of the dynamic programming when one considers the class* $\mathcal{M}_g$ *of martingale measures with terminal distribution constraint. This observation motivates us also to consider* $\widehat{\mathcal{M}}_g$ *to recover the dynamic programming and hence to obtain the duality by considering (strong) stopping times* $\tau \in \mathcal{T}(\widehat{\mathbb{F}})$. *Notice that the introduction of* $\widehat{\mathcal{M}}_g$ *is closed related to the dynamic programming approach of Cox and Källblad [53] for the MOT problem.*

The discrete time case under proportional transaction cost The MOT problem under market friction has been first studied by Dolinsky and Soner [74, 75] in both discrete and continuous time framework, where the proofs are based on discretization technique similar to that in [72]. As mentioned in the introduction part, the pricing-hedging duality relies essentially on the Fundamental Theorem of Asset Pricing (FTAP), which relates the existence of a martingale measure and the no-arbitrage condition. For a dominated market $(\Omega, \mathcal{F}, \mathbb{P})$ in presence of transaction cost, instead of the martingale measure, the good notion is the so-called consistent price system, which is a couple $(\mathbb{Q}, Z)$ such that $\mathbb{Q} \ll \mathbb{P}$, Z is $\mathbb{Q}$-martingale, and Z is "closed" to the underlying stock process X. In a robust framework similar to [35], Bayraktar and Zhang [12], Bouchard and Nutz [36] provide the FTAP for the market with proportional transaction cost, that is, the no-arbitrage condition is equivalent to the existence of a robust version of consistent price system. Burzoni [47] considers a point wise super-hedging framework, and provides a robust FTAP as well as the pricing-hedging duality result.

In a recent work of Bouchard, Deng and Tan [33], we consider the framework of [36], and prove a pricing-hedging duality under a general transaction cost model. Our main idea is to introduce a randomization technique, which reduces the original market into a frictionless market, and then to apply the classical results and techniques to prove the duality. For ease of presentation, we will stay in a simplified context with the framework of Bouchard-Nutz [36]: let $\Omega = \Omega_0 \times \Omega_1 \times \cdots \times \Omega_1$, equipped with a filtration $\mathbb{F}$, $\mathcal{P}$ be a set of Borel probability measures satisfying some measurability conditions. In particular, the case that $\mathcal{P}$ contains all Borel probability measures is a special example in this context. Let X be an 1-dimensional positive underlying process, an admissible strategy is a $\mathbb{F}$-predictable process $H = (H_k)_{1 \le k \le N}$. We set $H_0 = H_{N+1} = 0$ and denote $\Delta H_k := H_k - H_{k-1}$ for all $k = 1, \cdots, N+1$. Then the P&L (profit and loss) of an admissible strategy is given by

$$(H \circ X)_N - \sum_{k=1}^{N+1} \lambda |\Delta H_k| X_{k-1} \;=\; \sum_{k=1}^{N} H_k \Delta X_k - \sum_{k=1}^{N+1} \lambda |\Delta H_k| X_{k-1},$$

where $\lambda \in (0,1)$ is a ratio constant for the transaction cost whenever there is a 1 transaction of the underlying stock X. Given an exotic derivative option $\xi : \Omega \to \mathbb{R}$ and finitely many liquid options $\zeta_i : \Omega \to \mathbb{R}$, $i = 1, \cdots, e$, with price 0, then the minimum super-hedging cost of ξ using static strategy on ζ and dynamic strategy on X with transaction cost is then given by

$$\pi_e \;:=\; \inf\Big\{x \;: x+\sum_{i=1}^{e} h_i\zeta_i + (H\circ X)_N - \sum_{k=1}^{N+1}\lambda|\Delta H_k|X_{k-1} \geq \xi,\; \mathcal{P}\text{-q.s.}\Big\}. \tag{2.3.7}$$

To formulate the pricing problem, we denote by $\mathcal{S}_e$ the collection of all consistent price system $(\mathbb{Q}, Z)$ such that $\mathbb{Q} \leq \mathbb{P}$ for some $\mathbb{P} \in \mathcal{P}$, $(1-\lambda)X_t < Z_t < (1+\lambda)Z_t$, Z is a $(\mathbb{F}, \mathbb{Q})$-martingale, and $\mathbb{E}^{\mathbb{Q}}[\zeta_i] = 0$, $i = 1, \cdots, e$.

Theorem 2.3.6. *Under a technical robust no-arbitrage condition, the set $\mathcal{S}_e$ is non-empty. Moreover, for any Borel ξ and ζ_i, $i = 1, \cdots, e$, one has the duality*

$$\pi_e \;=\; \sup_{(\mathbb{Q},Z)\in\mathcal{S}_e} \mathbb{E}^{\mathbb{Q}}[\xi].$$

Proof. Let us consider a enlarged space $\overline{\Omega} := \Omega \times (1-\lambda, 1+\lambda)^{N+1}$ with an canonical process $\overline{X}_t(\omega,\theta) := X_t(\omega)\theta_t$, for all $(\omega,\theta) = (\omega,\theta_0,\cdots,\theta_N) \in \overline{\Omega}$. Denote by $\overline{\mathcal{P}}$ the collection of all probability measures $\overline{\mathbb{P}}$ on $\overline{\Omega}$ such that $\overline{\mathbb{P}}|_{\Omega} \in \mathcal{P}$. Then by direction computation, the super-hedging cost π_e defined by (2.3.7) is equivalent to

$$\inf\Big\{x \;: x+\sum_{i=1}^{e} h_i\zeta_i + (H\circ\overline{X})_N \geq \xi,\; \overline{\mathcal{P}}\text{-q.s.}\Big\}.$$

Notice that the above formulation is a robust super-hedging problem in a frictionless market and hence one can apply classical results and techniques in frictionless market to obtain the duality. □

Remark 2.3.6. (i) *One can also consider the case with transaction cost on the static options, it is enough to introduce a variated enlarged space and then apply the same technique to obtain the duality result.*

(ii) *The randomization approach in the proof could be used in a more general framework of transaction cost, using the notion of solvency cone, see e.g. Kabanov and Safarian [132]. The modeling allows essentially the direct exchange between different underlying stocks when $d > 1$, which was not possible in [74, 47].*

2.4 Characterization of the optimal solutions

It is interesting and important to establish the duality result to relate the two different optimization problems, but it should be more important to obtain the optimal solution as well as the optimal value. In some special cases, one can obtain or characterize the optimal solutions, where the duality result could be very useful. In classical OT theory, a general characterization of the optimal transport plan is obtained by the monotonicity principle. More precisely, the monotonicity principle provides a characterization on the support of the optimal transport plan, that is, one cannot improve the transport plan by replacing two couples $\{(x_1, y_1), (x_2, y_2)\}$ by $\{(x_1, y_2), (x_2, y_1)\}$, since the above replacement does not change the marginal distribution of the transport plan, see e.g. Villani [196]. This idea has been adapted to the one-period and one-dimensional discrete time MOT problem by Beiglböck and Juillet [20], and to the optimal SEP by Beiglböck, Cox and Huesmann [16]. See also [17] for some extensions.

Left monotone martingale given infinitely many marginals (Peacock) Let us first consider the one-period and one-dimensional discrete time MOT problem, $\sup_{\mathbb{P}\in\mathcal{M}(\mu_0,\mu_1)} \mathbb{E}^{\mathbb{P}}[c(X_0,X_1)]$ as recalled in Section 2.2. The monotonicity principle of [20, 117] leads to the so-called left-monotone martingale transport plan, which is optimal for a class of payoff functions $c(x,y)$ satisfying the so-called martingale Spence-Mirrlees condition, i.e. $y \mapsto \partial_x c(x,y)$ is strictly convex for all $x \in \mathbb{R}$. An explicit construction of the left-monotone transport plan under quite general conditions has been obtained in Henry-Labordère and Touzi [117]. They show that there are two functions $T_u : \mathbb{R} \to \mathbb{R}$ and $\mathbb{T}_d : \mathbb{R} \to \mathbb{R}$ such that $T_d(x) \le x \le T_u(x)$, and the optimal martingale transport plan in $\mathcal{M}(\mu_0,\mu_1)$ is in form

$$\mathbb{P}(dx,dy) \;=\; \mu_0(dx)\big[q(x)\delta_{T_u(x)}(dy) + (1-q(x))\delta_{T_d(x)}(dy)\big],$$

where $q(x)$ is such that $q(x)T_u(x) + (1-q(x))T_d(x) = x$ so that $\mathbb{P}$ a martingale measure. The two functions T_u, T_d is determinated by an ODE on $\mathbb{R}$. Moreover, for the class of payoff function satisfying the martingale Spence-Mirrlees condition, they also provide explicitly the optimizer for the dual super-hedging problem as well as the optimal value of the problem. An easy extension of the above result is the case of $n+1$-marginals $(\mu_0,\mu_1,\cdots,\mu_n)$ with reward function $c(x_0,x_1,\cdots,x_n) := \sum_{k=1}^n c_k(x_{k-1},x_k)$. The above extension is immediate since one can decompose the problem into n MOT problems with two marginals, i.e. $\sup_{\mathbb{P}\in\mathcal{M}(\mu_{k-1},\mu_k)} \mathbb{E}^{\mathbb{P}}\big[c_k(X_{k-1},X_k)\big]$. In this case, one obtains a n periods left-monotone martingale.

In Henry-Labordère, Tan and Touzi [114], we study the case with infinitely many marginals as limit of the case with n marginals, to obtain an explicit characterization of the limit continuous time process. In particular, it provides a new class of martingale Peacock process, which is in addition optimal for a class of reward functions.

Let $(\mu_t)_{t\in[0,1]}$ be a family of one-dimensional marginal distributions on $\mathbb{R}$, and $F(t,\cdot)$ (resp. $f(t,\cdot)$ be the cumulative distribution (resp. density) function of μ_t. Under some regularity condition on $F(t,x)$ and $f(t,x)$, assume that $x \mapsto \partial_t F(t,x)$ has a unique local maximizer $m(t)$ on the support of μ_t, the following functions T_d, j_d and j_u is well defined on D^c with $D := \{(t,x)\ : t\in[0,1], x \le m(t)\}$: first, let $T_d(t,x) \le x$ be defined by

$$\int_{T_d(t,x)}^{x} (x-y)\partial_t f(t,y)dy;$$

then j_d and j_u be defined by

$$j_d(t,x) := x - T_d(t,x), \quad j_u(t,x) := \frac{\partial_t F(t,T_d(t,x)) - \partial_t F(t,x)}{f(t,x)}.$$

Theorem 2.4.1. (i) *Under some technical conditions on $f(t,x)$, the following SDE, with initial condition $X_0 \sim \mu_0$, has a unique weak solution,*

$$X_t = X_0 - \int_0^t \mathbf{1}_{\{X_{s-}>m(s)\}} j_d(s,X_{s-})(dN_s - \nu_s), \quad \nu_s := \frac{j_u}{j_d}(s,X_{s-})\mathbf{1}_{\{X_{s-}>m(s)\}},$$

where N_t is unit size jump process with predictable compensated process ν.

(ii) *The above solution* X *is a martingale such that* $X_t \sim \mu_t$ *for all* $t \in [0,1]$*, and it is an optimal solution to the MOT under infinitely many marginals:*

$$\sup_{M \textit{ martingale}, \, M_t \sim \mu_t} \mathbb{E}\Big[\frac{1}{2}\int_0^1 \partial^2_{yy} c(M_t, M_t) d[M]^c_t + \sum_{0 \le t \le 1} c(M_{t-}, M_t)\Big].$$

Moreover, one can establish a duality and construct an explicit optimal solution for the dual problem.

Remark 2.4.1. (i) *The proof of the above problem is based on an approximation technique, by considering the discretization* $0 = t_0 < t_1 < \cdots < t_n = 1$ *and the corresponding* $n+1$*-marginals MOT problem that are studied in [117].*

(ii) *When* $(\mu_t)_{0 \le t \le 1}$ *is the family of marginal distribution of a Brownian motion, a martingale with marginal distributions* μ *is called a fake Brownian motion. Our technical conditions are in particular satisfied by the Brownian marginals and hence it provides a new class of fake Brownian motion, which is optimal for a class of reward functions.*

An alternative proof of the monotonicity principle of the optimal SEP Beiglböck, Cox and Huesmann [16] introduced and proved a monotonicity principle for the optimal SEP inspired by that in classical OT theory. An extraordinary contribution of this work is that it provides a unified explanation of all well known Skrokohod embedding solutions that are optimal w.r.t. different reward functions, which have been found and studied since several decades. In Guo, Tan and Touzi [107], we suggested an alternative proof in vein of the classical proof for monotonicity principle of OT problems.

Recall that by considering the joint law of the Brownian motion as well as the stopping time, a Skorokhod embedding on the Brownian motion can be viewed as probability measure $\overline{\mathbb{P}}$ on canonical space $\overline{\Omega} := \Omega_0 \times \mathbb{R}_+$ with canonical element (B, T), where $\Omega_0 := \mathbb{C}(\mathbb{R}_+, \mathbb{R})$ denotes the canonical space of continuous paths on $\mathbb{R}_+$. Moreover, the marginal distribution on Ω_0 is the Wiener measure $\mathbb{P}_0$, i.e. $\overline{\mathbb{P}}|_{\Omega_0} = \mathbb{P}_0$. Given a marginal distribution μ on $\mathbb{R}$, we denote by $\mathcal{T}(\mu) = \mathcal{T}(\delta_0, \mu)$ of all such measures $\overline{\mathbb{P}}$ such that $B_0 = 0$ and $B_T \sim^{\overline{\mathbb{P}}} \mu$. Let $\Phi : \Omega_0 \times \mathbb{R}_+ \to \mathbb{R}$ be a non-anticipative reward function, the optimal SEP is given by

$$P(\mu) \;:=\; \sup_{\overline{\mathbb{P}} \in \mathcal{T}(\mu)} \mathbb{E}^{\overline{\mathbb{P}}}\big[\Phi(B, T)\big].$$

We also recall that its dual problem can be given by

$$D(\mu) \;:=\; \inf\big\{\mu(\lambda) \;: \lambda(B_t) + M_t \ge \Phi(B_\cdot, t), \;\; \forall t \ge 0, \; \mathbb{P}_0\text{-a.s.}\big\}.$$

In the above, $M = (M_t)_{t \ge 0}$ is a $\mathbb{P}_0$-martingale w.r.t. the Brownian filtration.

Definition 2.4.1. *A pair* $(\bar\omega = (\omega, \theta), \bar\omega' = (\omega', \theta')) \in \overline{\Omega} \times \overline{\Omega}$ *is said to be a stop-go pair if* $\omega_\theta = \omega'_{\theta'}$ *and*

$$\Phi(\bar\omega) + \Phi(\bar\omega' \otimes \bar\omega'') \;>\; \Phi(\bar\omega \otimes \bar\omega'') + \xi(\bar\omega') \quad \textit{for all} \quad \bar\omega'' \in \overline{\Omega}^+,$$

where $\overline{\Omega}^+ := \big\{\bar\omega = (\omega, \theta) \in \overline{\Omega} : \theta > 0\big\}$ *and* $\bar\omega \otimes \bar\omega'' := (\omega \otimes_\theta \omega'', \theta + \theta')$ *with*

$$(\omega \otimes_\theta \omega'')_t := \omega_t \mathbf{1}_{[0,\theta)}(t) + (\omega_\theta + \omega''_{t-\theta}) \mathbf{1}_{[\theta,\infty)}(t), \quad \textit{for all } t > 0.$$

Denote by SG *the set of all stop-go pairs.*

Let $\Gamma \subseteq \overline{\Omega}$ be a subset, we define $\Gamma^<$ by

$$\Gamma^< \;:=\; \big\{\bar{\omega} = (\omega, \theta) \in \overline{\Omega} \;: \omega_{\theta\wedge\cdot} = \omega'_{\theta\wedge\cdot} \text{ for some } \bar{\omega}' \in \Gamma \text{ with } \theta' > \theta\big\}.$$

Theorem 2.4.2 (Beiglböck, Cox and Huesmann)**.** *Let $\Phi : \overline{\Omega} \to \mathbb{R}$ be a Borel non-anticipative random variable. Assume that the $P(\mu) = D(\mu)$ holds true, and $\overline{\mathbb{P}}^* \in \mathcal{T}(\mu)$ is an optimal embedding so that $P(\mu) = \mathbb{E}^{\overline{\mathbb{P}}^*}[\xi]$. Then there exists a Borel subset $\Gamma^* \subseteq \overline{\Omega}$ such that*

$$\overline{\mathbb{P}}^*\big[\Gamma^*\big] \;=\; 1 \quad \text{and} \quad \mathrm{SG} \,\cap\, \big(\Gamma^{*<} \times \Gamma^*\big) \;=\; \emptyset.$$

Proof. Suppose that there exists a dual minimizer (λ^*, M^*) of of the dual problem $D(\mu)$, i.e.

$$\lambda^*(\omega_t) + M^*_t(\omega) \;\geq\; \Phi(\omega, t), \text{ for all } t \geq 0,\ \mathbb{P}_0\text{-a.s. and } \mu(\lambda^*) \;=\; \mathbb{E}^{\overline{\mathbb{P}}^*}[\Phi(B, T)]. \quad (2.4.1)$$

The above inequality together with the duality implies that

$$\Gamma := \{(\omega, \theta) : \lambda^*(\omega_\theta) + M^*_\theta(\omega) = \Phi(\omega, \theta)\}$$

has full measure under $\overline{\mathbb{P}}^*$. We claim that $(\Gamma^< \times \Gamma) \cap \mathrm{SG} = \emptyset$. Otherwise, there exists a pair $(\bar{\omega}, \bar{\omega}') \in (\Gamma^< \times \Gamma) \cap \mathrm{SG}$, and hence they satisfy the condition

$$\Phi(\bar{\omega}) + \Phi(\bar{\omega}' \otimes \bar{\omega}'') \;>\; \Phi(\bar{\omega} \otimes \bar{\omega}'') + \Phi(\bar{\omega}') \quad \text{for all} \quad \bar{\omega}'' \in \overline{\Omega}^+. \qquad (2.4.2)$$

Let $\overline{\mathbb{Q}}^*_{\bar{\omega}}$ be the conditional probability of $\overline{\mathbb{P}}^*$ given $\{B_{\theta\wedge\cdot} = \omega_{\theta\wedge\cdot}, T > \theta\}$. Then it follows that

$$\Phi(\bar{\omega}) + \mathbb{E}^{\overline{\mathbb{Q}}^*_{\bar{\omega}}}[\Phi(\bar{\omega}' \otimes \cdot)] \;\;>\;\; \mathbb{E}^{\overline{\mathbb{Q}}^*_{\bar{\omega}}}[\Phi(\bar{\omega} \otimes \cdot)] + \Phi(\bar{\omega}').$$

On the other hand, notice that the marginal distribution of $\overline{\mathbb{Q}}^*_{\bar{\omega}}$ on Ω is still a Wiener measure. Then denoting $(M^* + \lambda^*)(\omega, \theta) := M^*_\theta(\omega) + \lambda^*(\omega_\theta)$, one has from (2.4.1) that

$$\Phi(\bar{\omega}) + \mathbb{E}^{\overline{\mathbb{Q}}^*_{\bar{\omega}}}[\Phi(\bar{\omega}' \otimes \cdot)] \;\leq\; (M^* + \lambda^*)(\bar{\omega}) + \mathbb{E}^{\overline{\mathbb{Q}}^*_{\bar{\omega}}}[(M^* + \lambda^*)(\bar{\omega}' \otimes \cdot)].$$

Notice that M^* is assumed to be a martingale, and one has from the definition of SG that $\omega_\theta = \omega'_{\theta'}$, it follows that

$$(M^* + \lambda^*)(\bar{\omega}) + \mathbb{E}^{\overline{\mathbb{Q}}^*_{\bar{\omega}}}[(M^* + \lambda^*)(\bar{\omega}' \otimes \cdot)] = \mathbb{E}^{\overline{\mathbb{Q}}^*_{\bar{\omega}}}[(M^* + \lambda^*)(\bar{\omega} \otimes \cdot)] + (M^* + \lambda^*)(\bar{\omega}').$$

Finally, notice that from the definition of SG and $\overline{\mathbb{Q}}^*_{\bar{\omega}}$, one knows that $\overline{\mathbb{Q}}^*_{\bar{\omega}}[\bar{\omega} \otimes \cdot \in \Gamma] = 1$ and $\bar{\omega}' \in \Gamma$, then

$$\begin{aligned}\Phi(\bar{\omega}) + \mathbb{E}^{\overline{\mathbb{Q}}^*_{\bar{\omega}}}[\Phi(\bar{\omega}' \otimes \cdot)] \;&\leq\; \mathbb{E}^{\overline{\mathbb{Q}}^*_{\bar{\omega}}}[(M^* + \lambda^*)(\bar{\omega} \otimes \cdot)] + (M^* + \lambda^*)(\bar{\omega}') \\ &=\; \mathbb{E}^{\overline{\mathbb{Q}}^*_{\bar{\omega}}}[\Phi(\bar{\omega} \otimes \cdot)] + \Phi(\bar{\omega}'),\end{aligned}$$

which is a contradiction with (2.4.2) and we hence obtain that $(\Gamma^< \times \Gamma) \cap \mathrm{SG} = \emptyset$. □

Remark 2.4.2. (i) *The above proof is only heuristic. For a rigorous proof, the dual minimiser (λ^*, M^*) may not exists and one needs to consider an optimizing sequence (λ_n, M_n). The rest of the proof can follow the same idea with limit $\lim_{n\to\infty}$ in many places.*

(ii) *The above monotonicity principle has been formulated and proved in [16]. Our main contribution is to provide an alternative proof based on the duality $P(\mu) = D(\mu)$. Such a proof is in line with the classical proof for the monotonicity principle of the classical OT problem, see e.g. [196], with a clear structure as illustrated in the above heuristic proof.*

2.5 Numerical methods

In general cases, an explicit description of the optimal solution for both MOT problem (the pricing problem) and the robust super-hedging problem would be impossible and numerical methods is needed. In the work Tan and Touzi [192], Bonnans and Tan [29], we developed a numerical scheme for the MOT problem and the optimal SEP based on its dual problem. To describe the main steps of the algorithm, let us stay in the context of Theorem 2.3.4 for an optimal SEP, and denote by $\mathcal{T}$ the collection of all Brownian stopping times. Let us consider the problem

$$V \;:=\; \inf_{\lambda_0,\lambda_1\in\mathcal{C}_1} V(\lambda) \quad \text{with } V(\lambda) := \mu(\lambda) + \sup_{\tau\in\mathcal{T}} \mathbb{E}\big[\Phi(B,\tau) - \lambda_1(B_\tau) - \lambda_0(B_0)\big].$$

Using the weak duality "max min $\le$ min max", it is easy to check that the value V is bounded between the l.h.s. and the r.h.s. of the first formula in Theorem 2.3.4, and then it follows from the duality result of Theorem 2.3.4 that the value V defined above equals to the original optimal SEP in Theorem 2.3.4.

Our numerical algorithm consists in 4 steps:

1. Let $L > 0$, we denote by Lip_L the collection of all Lipschitz function on $\mathbb{R}$ with Lipschitz constant L, then let

$$V_L \;:=\; \inf_{\lambda_0,\lambda_1\in\mathrm{Lip}_L} V(\lambda)$$

2. Let $M, N > 0$, we denote by $\mathcal{T}_{M,N}$ the collection of all Brownian stopping times τ bounded by N and such that $\sup_{t\ge 0}|B_{t\wedge\tau}| \le M$. Then let

$$V_L^{M,N} := \inf_{\lambda_0,\lambda_1\in\mathrm{Lip}_L} V^{M,N}(\lambda), \quad V^{M,N}(\lambda) := \mu(\lambda) + \sup_{\tau\in\mathcal{T}_{M,N}} \mathbb{E}\big[\Phi(B,\tau) - \lambda_1(B_\tau) - \lambda_0(B_0)\big].$$

3. Let $\Delta = (\Delta t, \Delta x)$ be the time-space discretization parameter, and (t_i, x_j) be a discrete grid of $[0,N]\times[-M,M]$. We denote by Lip_L^{Δ} the collection of all L-Lipschitz function defined on $\{x_j \in [-M,M],\ j\in\mathbb{Z}\}$, and by $V^{M,N,\Delta}(\lambda)$ the numerical solution of the optimal stopping problem $V^{M,N}(\lambda)$ from the finite difference scheme. Let

$$V_L^{M,N,\Delta} := \inf_{\lambda_0,\lambda_1\in\mathrm{Lip}_L^{\Delta}} V^{M,N,\Delta}(\lambda). \tag{2.5.1}$$

4. Notice that Lip_L^{Δ} is a compact set in a finite dimensional space, and one can show that $\lambda \in \mathrm{Lip}_L^{\Delta} \mapsto V^{M,N,\Delta}(\lambda) \in \mathbb{R}$ is a convex function, then one can apply a classical gradient algorithm to solve the minimization problem (2.5.1), see e.g. [14, 81], etc.

Theorem 2.5.1. *Under technical conditions, the numerical algorithm converges.*

Remark 2.5.1. *We are only able to show the general convergence for the first step $V_L \to V$. For the rest of steps, we have obtained convergence rate.*

2.6 Perspectives

Many advances have been made for the MOT theory during the last years, but there are still many open (and interesting) questions. More generally, if one goes back to the original motivation of the MOT problem, that is, to improve the risk management in a robust context using semi-static strategies, then there would be even more interesting questions.

For example, we may consider the general utility maximization problems and see how it would be improved by considering the semi-static strategies, see e.g. [10, 183]. In a more concrete project, we are interested in solving a utility maximization problem under transaction cost using the randomization approach that we developed in Bouchard, Deng and Tan [33].

As shown above, the MOT duality is related essentially to the fundamental theorem of asset pricing (FTAP), as well as the super-martingale's optional decomposition in a robust context. This problem has been very well solved in a discrete time framework by Bouchard and Nutz [35], and in a continuous time framework with continuous underlying path by Biagini, Bouchard, Kardaras and Nutz [24]. In a recent project with Bruno Bouchard, Kostas Kardaras and Marcel Nutz, we are interested in solving this problem for the continuous time case with càdlàg underlying paths.

Chapitre 3

Non-Markovian Stochastic Control, Path-dependent HJB Equation and their Numerical Approximations

3.1 Introduction

The stochastic optimal stopping/control problem is a very important subject/tool in applied mathematics, and especially in mathematical finance. Since 1970s, there should be thousands of papers and numerous monographs which are dedicated to this subject. Motivated by its applications, there would be many different formulations.

Let us start by a simple stochastic control problem in the one-dimensional case: let W be a standard Brownian motion w.r.t. a filtration $\mathbb{F}$, an admissible control process is a $\mathbb{F}$-predictable process $(\nu_t)_{t\geq 0}$ taking value in some space U. Let $(\mu,\sigma) : \mathbb{R}_+ \times \mathbb{R} \times U \to \mathbb{R} \times \mathbb{R}$ be the given coefficient functions, then given a control process $(\nu_t)_{t\geq 0}$, as well as an initial condition x_0, we define the controlled process X^ν as solution of the SDE

$$X_t^\nu = x_0 + \int_0^t \mu(s, X_s^\nu, \nu_s) ds + \int_0^t \sigma(s, X_s^\nu, \nu_s) dW_s.$$

In general, some technical conditions on (μ,σ) are needed to ensure the wellposedness of the above SDE. Then, given reward functions $f : \mathbb{R}_+ \times \mathbb{R} \times U \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$, one has the following standard stochastic optimal control problem:

$$\sup_\nu \mathbb{E}\Big[\int_0^T f(t, X_t^\nu, \nu_t) dt + g(X_T^\nu)\Big]. \tag{3.1.1}$$

There could be numerous variated version of the optimal control problem:

- In terms of the controlled process, it could be
  - the controlled diffusion process, or diffusion-jump process, which could be defined by controlled SDE, with solutions in strong sense, or in weak sense, or in relaxed sense;
  - a general controlled Markov process, such as the continuous time Markov chain, branching process, etc;
  - nonlinear process, such as process with interaction (in sense of McKean-Vlasov), forward process in a coupled FBSDE, etc;
  - etc.

- In terms of the reward function, it could be
  - finite horizon or infinite horizon reward function;
  - asymptotic value such as $\lim_{T\to\infty} \frac{1}{T}\int_0^T f(t, X_t, \nu_t)dt$;
  - etc.
- In terms of the admissible controls, it could be
  - process adapted to a sub-filtration $\mathbb{G} \subset \mathbb{F}$, which leads to the partial observation control problem, or the so-called filtering problem;
  - control under constraints: delta constraint, gamma constraint, expectation constraint, target problem, etc.
  - etc.

As an extension of the classical deterministic control problem, a first approach to characterize the optimal control could be the stochastic maximum principle in spirit of Pontryagin [170]. Assuming the existence of an optimal control, the maximum principle provides a characterization of the optimal control by a stochastic forward backward system, see e.g. Peng [165], Tang and Li [193], etc. Notice also that an optimal control may not exist for a general control problem.

It seems that the Bellman's dynamic programming approach is much more popular for the stochastic control problem in the literature. The dynamic programming principle allows to decompose the global optimization problem into a family of local optimization problems. From a probabilistic point of view, this local property provides a supermartingale characterization of the value function process. This point of view has been explored in El Karoui [85], especially for the optimal stopping problem and the control problem in the dominated case, where the later problem can be reformulated as a control problem on a family of equivalent probability measures. Notice that in [85], the optimal stopping problem has been solved in a good generality, and under mild conditions, one has existence of the optimal stopping times. As for the optimal control problem, one generally needs some convexity condition to ensure the existence of the optimal controls. In this case, one may use Krylov's Markov selection method to select a Markov feedback control as optimal control, see e.g. Haussmann [110], [87], etc. Otherwise, it is the relaxed formulation which allows to obtain the existence, see e.g. Fleming [94], and El Karoui, Nguyen and Jeanblanc [87], etc. The idea of the relaxed control is to see the class of all controls as a compact convex sets of probability measures on a good canonical space, and hence the existence of the optimal control would follow from a compactness argument.

Another approach in vein of the dynamic programming principle consists in characterizing the value function by the PDE, the Bellman equation, in sense of Soblev solution or viscosity solution. Take the basic control problem in (3.1.1) as example, the associated Bellman equation should be

$$\partial_t v(t,x) + \sup_{u\in U}\Big(\mu(\cdot,u)\partial_x v + \frac{1}{2}\sigma^2(\cdot,u)\partial^2_{xx}v + f(\cdot,u)\Big)(t,x) = 0, \quad v(T,\cdot) = g(\cdot). \quad (3.1.2)$$

Such an approach has been explored by many authors for different problems, we can refer to the books of Bensoussan and Lions [15], Fleming and Rishel [95], Fleming and Soner [96], Krylov [141], Yong and Zhou [198], and more recently the monographs Pham [169], Touzi [194], see also the lecture note of Bouchard [32], etc. Moreover, given a solution to the Bellman equation, the verification theorem allows to obtain the optimal (or ε-optimal) controls. Further, the PDE approach opens a new door for the numerical analysis of the control problem. Convergence analysis methods using the notion of viscosity solution has been provided by Barles and Souganidis [9], Barles, Daher and Romano [7], Krylov [142], Barles and Jakobsen [8], etc.

From another point of view, this relation between the control problem and the PDE provides an extension of the Feynmann-Kac formula to the nonlinear case. Here, the non-linearity means that by writing (3.1.2) in form $F(\cdot, v, \partial_t v, \partial_x v, \partial^2_{xx} v)(t,x) = 0$, the map $(y, \theta, z, \gamma) \mapsto F(\cdot, y, \theta, z, \gamma)$ is nonlinear (or more precisely sublinear). In the semilinear case, the backward stochastic differential equation (BSDE) provides a more general extension of the Feynmann-Kac formula. Indeed, let $f : \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a generator, and we consider the semilinear parabolic PDE

$$\partial_t v(t,x) + \frac{1}{2}\sigma^2 \partial^2_{xx} v(t,x) + f(\cdot, v, \sigma \partial_x v)(t,x) = 0, \quad v(T, \cdot) = g(\cdot).$$

Under some technical conditions, the solution $(u, \sigma\partial_x u)$ can be represented by solution (Y, Z) of the BSDE:

$$Y_t = g(X_T) - \int_t^T f(s, X_s, Y_s, Z_s) ds + \int_t^T Z_s dW_s, \text{ with } X_t := x_0 + \int_0^t \sigma(s, X_s) dW_s. \quad (3.1.3)$$

Let us refer to seminal papers of Pardoux and Peng [163], El Karoui, Peng and Quenez [89] for a precise definition of the solution of the BSDE as well as the standard condition for its wellposedness. More recently, there is a new effort to extend this Feynmann-Kac formula to a more general form such as

$$\partial_t v(t,x) + H(t, x, v, \partial_x v, \partial^2_{xx} v) = 0, \quad v(T, \cdot) = g(\cdot),$$

for some map $H : \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. One can also study the Feyanmann-Kac formula in a non-Markovian context, by considering the path-dependent PDE:

$$\partial_t v(t,\omega) + H(t, \omega, v, \partial_\omega v, \partial^2_{\omega\omega} v) = 0, \quad v(T, \omega) = \xi(\omega),$$

where ω represents a continuous (or càdlàg) paths, and the derivatives $\partial_\omega u$ and $\partial^2_{\omega\omega} u$ are given in an appropriate sense. This effort leads to the so-called 2nd order BSDEs and the path-dependent PDEs (PPDEs), see e.g. the series of papers by Cheridito, Soner, Touzi and Victoir [50], Soner, Touzi and Zhang [186], Ekren, Keller, Touzi and Zhang [82, 83, 84], Ren, Touzi and Zhang [175, 176, 177], etc. These new notions of equations have been motivated and have applications in finance for risk management of the non-Markovian exotic options (see e.g. Dupire [79], Becherer and Kentia [13], etc.), in economics for principal-agent problems (see e.g. Cvitanic, Possamaï and Touzi [61]), etc. We also notice that in a series of works by Kharroubi, Pham et al. [139, 98], etc., another probabilistic representation of the Bellman equations has been obtained by considering a constrained BSDE from a randomization approach to the control problem. Their randomization formulation consists in fact an intermediary formulation between the strong/weak formulation and the relaxed formulation of the control problem in El Karoui et el. [87].

My work in this subject stays mainly in a non-Markovian context, by studying the dynamic programming principle, the wellposedness of the 2nd order BSDEs as well as their numerical approximations.

3.2 The dynamic programming principle

The dynamic programming principle plays an essential role in the stochastic control theory. It decomposes a global optimization problem into a family of local optimization problems, and hence allows to provide the local characterization of the value function process, such as the super-martingale property, or Bellman equation characterization, etc.

Let us consider a non-Markovian controlled diffusion processes problem as a direct extension of the Markovian control problem (3.1.1). We still stay in the one-dimensional case for simplicity. Let $\Omega := \mathbb{C}(\mathbb{R}_+, \mathbb{R})$ denote the canonical space with $\mathbb{R}$-valued continuous paths on $\mathbb{R}_+$, with canonical filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ and canonical process B. Let $\mathbb{P}_0$ be the Wiener measure, under which B is a standard Brownian motion, and denote by $\mathbb{F}^{\mathbb{P}_0}$ the augmented filtration under $\mathbb{P}_0$. Let U be a Polish space, $\mathcal{U}$ denote the collection of all $\mathbb{F}$-progressively measurable U-valued process, let $(\mu, \sigma) : \mathbb{R}_+ \times \Omega \times U \to \mathbb{R} \times \mathbb{R}$ be the coefficient functions which are $\mathbb{F}$-progressively measurable, i.e. $(\mu, \sigma)(t, \omega, u) = (\mu, \sigma)(t, \omega_{t\wedge\cdot}, u)$ for all t, ω, u, and such that for some $K > 0$,

$$|(\mu, \sigma)(t, \omega, u) - (\mu, \sigma)(t, \omega', u)| \leq K|\omega - \omega'| \;\text{ and }\; |(\mu, \sigma)(t, \omega, u)| \leq K(1 + |\omega|).$$

Then given $\nu = (\nu_t)_{0\leq t\leq T} \in \mathcal{U}$, as well as an initial condition $(t, \omega) \in \mathbb{R}_+ \times \Omega$, we define $X^{t,\omega,\nu}_\cdot$ as the unique strong solution of the SDE: $X_{s\wedge t} := \omega_s,\ \forall s \leq t$; and

$$X_s = \omega_t + \int_t^s \mu(r, X_{r\wedge\cdot}, \nu_r)dr + \int_t^s \sigma(r, X_{r\wedge\cdot}, \nu_r)dB_r,\ \forall s \geq t,\ \mathbb{P}_0\text{-a.s.} \tag{3.2.1}$$

Given a bounded measurable variable $\xi : \Omega \to \mathbb{R}$, let us define

$$V(t, \omega) := \sup_{\nu\in\mathcal{U}} J(t, \omega, \nu) \quad\text{with}\quad J(t, \omega, \nu) := \mathbb{E}^{\mathbb{P}_0}\big[\xi\big(X^{t,\omega,\nu}_\cdot\big)\big]. \tag{3.2.2}$$

Notice that the above problem is in fact a strong formulation of the control problem, since the probability space as well as the Brownian motion is fixed. By considering the law of $X^{t,\omega,\nu}_\cdot$, one can also reformulate it equivalently by

$$V(t, \omega) := \sup_{\mathbb{P}\in\mathcal{P}(t,\omega)} \mathbb{E}^{\mathbb{P}}[\xi], \quad\text{with}\;\; \mathcal{P}(t, \omega) := \big\{\mathbb{P}_0 \circ (X^{t,\omega,\nu})^{-1}\ : \nu \in \mathcal{U}\big\}. \tag{3.2.3}$$

Then the dynamic programming principle (DPP) is usually given in the following way: for any $(t, \omega) \in \mathbb{R}_+ \times \Omega$ and any finite $\mathbb{F}$-stopping time τ on Ω,

$$V(t, \omega) \;=\; \sup_{\nu\in\mathcal{U}} \mathbb{E}^{\mathbb{P}_0}\big[V\big(\tau^{t,\omega,\nu}, X^{t,\omega,\nu}_\cdot\big)\big] \;=\; \sup_{\mathbb{P}\in\mathcal{P}(t,\omega)} \mathbb{E}^{\mathbb{P}}\big[V(\tau, B_\cdot)\big], \tag{3.2.4}$$

where $\tau^{t,\omega,\nu} := \tau(X^{t,\omega,\nu})$, which consists a $\mathbb{F}^{\mathbb{P}_0}$-stopping time for each (t,ω,ν). The above DPP has very intuitive meaning: the optimization problem on $[t,\infty)$ can be decomposed into two sequential problems on respectively $[t,\tau]$ and $[\tau,\infty)$. Before providing the proof, it is needed to show that V is measurable to ensure that the expectation $\mathbb{E}[V(\cdot)]$ is well defined. In the literatures, a classical way is to assume some regularity conditions on the reward functions, so that the value function is continuous and hence measurable, see e.g. Fleming and Soner [96], Krylov [141], etc. Besides, in the context of the PDE approach, such a regularity is also needed as viscosity solution of a PDE. We also refer to Bouchard and Touzi [41], Touzi [194] for the so-called weak DPP, by considering the semi-continuity envelop of the value function. As for the proof, the classical way is to decompose the equality in the DPP into two inequalities, where an easy one can be obtained by a simple conditioning argument, and the reverse one can be justified by a concatenation argument.

Conditioning of the controlled SDEs In the work of Claisse, Talay and Tan [52], we revisited the conditioning argument in this classical proof of the DPP and pointed out some subtle measurability issue related to the neglected sets. We then tried to formulate this argument and proved it in a detailed and rigorous way.

Given two paths $\omega,\omega' \in \Omega$ and $t \in \mathbb{R}_+$, we define the concatenated path $\omega \otimes_t \omega'$ by

$$(\omega \otimes_t \omega')_s := \omega_s \mathbf{1}_{\{0\le s\le t\}} + \big(\omega_t + \omega'_s - \omega_t\big)\mathbf{1}_{\{s>t\}}.$$

Then for any $\nu \in \mathcal{U}$, and $(t,\omega) \in \mathbb{R}_+ \times \Omega$, we define a shifted control process

$$\nu^{t,\omega}_s(\omega') := \nu_s(\omega \otimes_t \omega'), \quad \text{for all } (s,\omega') \in \mathbb{R}_+ \times \Omega.$$

Theorem 3.2.1. *For any $(t,\mathbf{x},\nu) \in \mathbb{R}_+ \times \Omega \times \mathcal{U}$ and any $\mathbb{F}^{\mathbb{P}_0}$-stopping time τ taking value in $[t,\infty)$, one has*

$$\mathbb{E}^{\mathbb{P}_0}\Big[\xi\big(X^{t,\mathbf{x},\nu}_\cdot\big) \;\Big|\; \mathcal{F}^{\mathbb{P}_0}_\tau\Big](\omega) = \; J\big(\tau(\omega), X^{t,\mathbf{x},\nu}_\cdot(\omega), \nu^{\tau(\omega),\omega}\big) \quad \textit{for } \mathbb{P}_0\textit{-a.e. } \omega \in \Omega. \tag{3.2.5}$$

Remark 3.2.1. (i) *Notice that for each ω, $\nu^{\tau(\omega),\omega}$ is an admissible control process, then (3.2.5) is dominated by $V\big(\tau(\omega), X^{t,\mathbf{x},\nu}_\cdot(\omega)\big)$ $\mathbb{P}_0$-a.s. Taking the expectation and then the supremum over all $\nu \in \mathcal{U}$, then it follows the easy inequality "$\le$" for the DPP (3.2.4).*

(ii) *This conditioning or pseudo-Markov property has been very often used as a trivial fact, without any further justification. In the case where τ equals to a deterministic time t, a sketch of proof has been provided in Fleming and Souganidis [97] by considering the canonical space $\mathbb{C}([0,\infty),\mathbb{R})$ as the product space $\mathbb{C}([0,t],\mathbb{R}) \times \mathbb{C}([t,\infty),\mathbb{R})$.*

(iii) *In our context, one needs to show that the process $X^{t,\mathbf{x},\nu}_\cdot$ is still the solution of the controlled SDEs associated with the shifted control $\nu^{\tau(\omega),om}$, which is a kind of flow property of the controlled SDE. Nevertheless, there could be some quite subtle issues related to the $\mathbb{P}$-null sets. For every control ν, the process $X^{t,\mathbf{x},\nu}_\cdot$ is defined by a controlled SDE w.r.t. the completed filtration $\mathbb{F}^{\mathbb{P}_0}$, up to some $\mathbb{P}_0$-null sets. When taking conditional expectation $\mathbb{E}^{\mathbb{P}_0}[\cdot|\mathcal{F}^{\mathbb{P}_0}_\tau]$, one needs to consider a family conditional probability measures of $\mathbb{P}_0$ w.r.t. $\mathcal{F}^{\mathbb{P}_0}_\tau$. Then a first problem is that the conditional probability measure of $\mathbb{P}_0$ w.r.t $\mathcal{F}^{\mathbb{P}_0}_\tau$ does not exist, a second problem is that even it exists, one needs to check that the solution of the* *controlled SDE w.r.t. the completed filtration under the conditional probability measure should equal to its solution w.r.t.* $\mathbb{F}^{\mathbb{P}_0}$ *under* $\mathbb{P}_0$*, up to some "null sets". Therefore, too many null sets under different probability measures are involved.*

(iv) *The main contribution of this work is to point out that a rigorous proof may not be so trivial even in this basic situation. We provide two proofs, a first one is based on the sketch of the proof in Fleming and Souganidis [97] (which was given in a context* $\tau \equiv t$*), and a second one is based on the associated martingale problem satisfying by* $(X^{t,\mathbf{x},\nu}, B)$ *under* $\mathbb{P}_0$*. Such a detailed and rigorous proof in this basic context would help to avoid possible gaps/mistakes when one studies more general and more sophisticated control problems.*

Dynamic programming principle by measurable selection Another classical approach to prove the DPP consists in using the measurable selection theorem, which avoids assuming the regularity conditions. Let $f : E \times F \to \mathbb{R}$ be a measurable function in a product space, and define $g : F \to \mathbb{R} \cup \{\infty\}$ by $g(x) := \sup_{e \in E} f(e, x)$. Then under some topological structure, the measurable selection theorem confirms that g is also (universally) measurable, and there is some $h : F \to E$ such that $h(x) \in E$ is a ε-optimizer of $\sup_{e \in E} f(e, x)$ for every $x \in F$. Let us refer to e.g. Parthasarathy [164] for a review of the different versions of the selection theorems, and in Bertsekas and Shreve [22] for a detailed presentation of the analytic selection theorem, which is used a lot in the context of the optimal control theory.

For the discrete time optimal control problems, this approach has been explored by Bertsekas and Shreve [22], Dellacherie [66], etc. For the continuous time controlled diffusion processes problem, if the volatility coefficient function is not controlled, the law of the controlled diffusion process are all absolutely continuous w.r.t. a reference measure, then the control problem can be transformed into an optimization problem over a class of equivalent measures, see e.g. El Karoui [85]. This situation is in fact covered by the later developed BSDE theory: the value function process is the Y-part of the solution to the BSDE and the DPP for the control problem is a simple conditioning argument on the solution of the BSDEs. For more general situation where the volatility is controlled, El Karoui, Huu Nguyen and Jeanblanc [87] suggested a martingale problem formulation by considering the law of the controlled processes (see (3.2.3)), and then proved a the DPP using the measurable selection techniques. The same technique is more recently used in Nutz and van Handel [160] to prove the DPP for the time consistency of the so-called sub-linear expectation. Based on these work, in El Karoui and Tan [91, 92], we try to give a detailed review on the measurable selection theorems, as well as how it could be used to deduce the DPP for a control problem. Moreover, we obtained the DPP for a large class of controlled/stopped martingale problems, which covers in particular various formulations of the controlled diffusion processes problems.

To illustrate the idea, let us stay in the classical situation with canonical space $\Omega := C(\mathbb{R}_+, \mathbb{R})$, canonical process B and canonical filtration $\mathbb{F}$, equipped with a class $(\mathcal{P}(t, \omega))_{(t,\omega) \in \mathbb{R}_+ \times \Omega}$ of non-empty probability measure sets on Ω. Namely, $\mathcal{P}(t, \omega)$ denotes the collection of all possible distributions of the controlled process, given initial condition (t, ω), on the canonical space. In particular, one can have (3.2.3) in mind as example of the family $\mathcal{P}(t, \omega)$. Then, for some $\xi : \Omega \to \mathbb{R}$, we naturally formulate the control problem by

$$V(t,\omega) := \sup_{\mathbb{P}\in\mathcal{P}(t,\omega)} \mathbb{E}^{\mathbb{P}}[\xi]. \tag{3.2.6}$$

To formulate the assumptions, let us recall some notions on the probability measures on Ω from Stroock and Varadhan [188]. Let $\mathbb{P}$ be a Borel probability measure on Ω, and τ a $\mathbb{F}$-stopping time, then there is a family $(\mathbb{P}_\omega)_{\omega\in\Omega}$ of probability measures on Ω, called a regular conditional probability distribution (r.c.p.d.) of $\mathbb{P}$ knowing $\mathcal{F}_\tau$, such that

- $\omega \mapsto \mathbb{P}_\omega$ is $\mathcal{F}_\tau$-measurable,
- $\mathbb{P}[A|\mathcal{F}_\tau](\omega) = \mathbb{E}^{\mathbb{P}_\omega}[A]$ for $\mathbb{P}$-a.e. $\omega\in\Omega$,
- $\mathbb{P}_\omega[B_{t\wedge\cdot} = \omega_{t\wedge\cdot}] = 1$ for every $\omega\in\Omega$.

Next, given a family $(\mathbb{Q}_\omega)_{\omega\in\Omega}$ of probability measures such that $\omega\mapsto\mathbb{Q}_\omega$ is $\mathcal{F}_\tau$-measurable and $\mathbb{P}_\omega[B_{\tau(\omega)\wedge\cdot} = \omega_{\tau(\omega)\wedge\cdot}] = 1$ for every $\omega\in\Omega$, one can define a concatenated probability measure $\mathbb{P}\otimes_\tau\mathbb{Q}_\cdot$ by

$$\mathbb{P}\otimes_\tau\mathbb{Q}_\cdot[A] := \int_\Omega \mathbb{Q}_\omega[A]d\mathbb{P}(\omega).$$

In particular, we know that $\mathbb{P} = \mathbb{P}\otimes_\tau\mathbb{Q}_\cdot$ on $\mathcal{F}_\tau$, and $(\mathbb{Q}_\omega)_{\omega\in\Omega}$ consists a r.c.p.d. of $\mathbb{P}\otimes_\tau\mathbb{Q}_\cdot$ knowing $\mathcal{F}_\tau$.

Assumption 3.2.1. (i) *Let $(t,\omega)\in\mathbb{R}_+\times\Omega$, one has $\mathcal{P}(t,\omega) = \mathcal{P}(t,\omega_{t\wedge\cdot})$ and $\mathbb{P}[X_{t\wedge\cdot} = \omega_{t\wedge\cdot}] = 1$ for every $\mathbb{P}\in\mathcal{P}(t,\omega)$; moreover, the graph set $[[\mathcal{P}]] := \{(t,\omega,\mathbb{P}) \ : \mathbb{P}\in\mathcal{P}(t,\omega)\}$ is Borel measurable.*

(ii) *Let $(t,\omega)\in\mathbb{R}_+\times\Omega$, τ be a $\mathbb{F}$-stopping time taking value in $[t,\infty)$, and $\mathbb{P}\in\mathcal{P}(t,\omega)$, for a r.c.p.d. $(\mathbb{P}_{\omega'})_{\omega'\in\Omega}$ of $\mathbb{P}$ knowing $\mathcal{F}_\tau$, one has $\mathbb{P}_{\omega'}\in\mathcal{P}(\tau(\omega'),\omega')$ for $\mathbb{P}$-a.e. $\omega'\in\Omega$.*

(iii) *Let $(t,\omega)\in\mathbb{R}_+\times\Omega$, τ be a $\mathbb{F}$-stopping time taking value in $[t,\infty)$, and $\mathbb{P}\in\mathcal{P}(t,\omega)$, given a family $(\mathbb{Q}_{\omega'})_{\omega'\in\Omega}$ such that $\mathbb{Q}_{\omega'}\in\mathcal{P}(\tau(\omega'),\omega')$, one has $\mathbb{P}\otimes_\tau\mathbb{Q}_\cdot\in\mathcal{P}(t,\omega)$.*

Theorem 3.2.2. *Let Assumption 3.2.1 hold true, and assume that $\xi:\Omega\to\mathbb{R}$ is Borel measurable. Then for every $(t,\omega)\in\mathbb{R}_+\times\Omega$ and every $\mathbb{F}$-stopping time τ taking value in $[t,\infty)$, the value function V defined by (3.2.6) is universally measurable, and it satisfies the DPP*

$$V(t,\omega) \;=\; \sup_{\mathbb{P}\in\mathcal{P}(t,\omega)} \mathbb{E}^{\mathbb{P}}\big[V(\tau,B_\cdot)\big]. \tag{3.2.7}$$

Proof. First, the measurability of V is a direct consequence of the measurable selection theorem, knowing that $(t,\omega,\mathbb{P})\mapsto\mathbb{E}^{\mathbb{P}}[\xi]$ is Borel measurable. Then a first inequality "$\le$" can be obtained by considering an arbitrary $\mathbb{P}\in\mathbb{P}(t,\omega)$ and a r.c.p.d. of $\mathbb{P}$ knowing $\mathcal{F}_\tau$, and then using Assumption 3.2.1 (ii). Finally, for the reverse inequality, one can choose a ε-optimal control $\mathbb{Q}^\varepsilon_\omega$ for every initial condition $(\tau(\omega),\omega)$, in a measurable way. Then it is enough to consider an arbitrary $\mathbb{P}$ and its concatenated measure $\mathbb{P}\otimes_\tau\mathbb{Q}^\varepsilon_\cdot$, and then to use Assumption 3.2.1 (ii). □

Remark 3.2.2. *One can consider $(t,\omega) \mapsto V(t,\omega)$ as a process defined on $\mathbb{R}_+$. Then from the DPP (3.2.7), one obtains that for all $0 \le s \le t$,*

$$V(s,\omega) \ \ge \ \mathbb{E}^{\mathbb{P}}\big[V(t,B)\big|\mathcal{F}_s\big], \ \mathbb{P}\text{-a.s. for all } \mathbb{P} \in \mathcal{P}(0,\omega).$$

Assume that V is $\mathbb{F}$-optional, then it is a $\mathbb{P}$-supermartingale. In practice, it may be difficult to check directly that V is $\mathbb{F}$-optional. Under a fixed probability, one can do a modification of V to obtain a $\mathbb{F}^{\mathbb{P}}$-optional supermartingale. Another way to obtain the supermartingale is to consider its right-continuous modification.

Our next result confirms that the above framework with Assumption 3.2.1 is convenient to study optimal control problem. Let us consider a general controlled martingale problem. Following the language of Ethier and Kurtz [90], we say a generator $\mathbb{G}$ for a control problem is a set of couples of functions (f,g), where $f : \mathbb{R} \to \mathbb{R}$, $g : \mathbb{R}_+ \times \Omega \times U \times \mathbb{R} \to \mathbb{R}$. As examples, on can have in mind the couples $(\varphi, \mathcal{L}\varphi)$ for an infinitesimal generator $\mathcal{L}$ of a Markov process with φ in the domain of the generator. Given a generator $\mathbb{G}$ and initial condition $(t,\omega) \in \mathbb{R}_+ \times \Omega$, a control term (resp. relaxed control term) is a term

$$\alpha \ = \ \big(\Omega^\alpha, \mathcal{F}^\alpha, \mathbb{F}^\alpha, \mathbb{P}^\alpha, X^\alpha, \nu^\alpha(\text{resp.} m^\alpha)\big),$$

where $\big(\Omega^\alpha, \mathcal{F}^\alpha, \mathbb{F}^\alpha, \mathbb{P}^\alpha\big)$ is a filtered probability space, equipped with a continuous process X^α such that $X^\alpha_{t\wedge\cdot} = \omega_{t\wedge\cdot}$, $\mathbb{P}^\alpha$-a.s., ν^α is a U-valued predictable process (resp. m^α is a $\mathcal{P}(U)$-valued predictable process with $\mathcal{P}(U)$ denotes the collection of all Borel probability measures on U). Moreover, the process $(C^\alpha_s(f,g))_{s\ge t}$ defined below is a local martingale for every $(f,g) \in \mathbb{G}$,

$$C^\alpha_s(f,g) := f(X^\alpha_s) - \int_t^s g(r, X^\alpha_{r\wedge\cdot}, \nu^\alpha_s \ (\text{resp.} m^\alpha_r), X^\alpha_s) ds,$$

with $g(r,\omega,m^\alpha_r,x) := \int_U g(r,\omega,u,x) m^\alpha(du)$. Denote by $\mathcal{A}(t,\omega)$ the collection of all control terms α with initial condition (t,ω), then one has the following controlled martingale problem:

$$\sup_{\alpha \in \mathcal{A}(t,\omega)} \mathbb{E}^{\mathbb{P}^\alpha}\big[X^\alpha_\cdot\big]. \tag{3.2.8}$$

Theorem 3.2.3. *Assume that there is a countable subset $\mathbb{G}_0 \subset \mathbb{G}$ which defines the same martingale problem as $\mathbb{G}$, and the set $\mathcal{A}(t,\omega)$ is non-empty. Then the family*

$$\mathcal{P}(t,\omega) := \big\{\mathbb{P}^\alpha \circ (X^\alpha_\cdot)^{-1} \ : \alpha \in \mathcal{A}(t,\omega)\big\}$$

satisfies Assumption 3.2.1. Consequently, one has the DPP for the controlled martingale problem (3.2.8) associated to the generator $\mathbb{G}$.

Remark 3.2.3. *The above result holds still in a more general context, where $\Omega = \mathbb{D}(\mathbb{R}_+, E)$ is the canonical space of càdlàg paths taking value in a Polish space E, with an appropriate generator $\mathbb{G}$. One can also consider an enlarged canonical space $\Omega \times \mathbb{R}_+$ to study the optimal control/stopping problems.*

To go back to the controlled diffusion processes problem, it is enough to define

$$\mathbb{G} := \big\{\big(\varphi,\ \mathcal{L}\varphi\big)\ : \varphi \in C_c^\infty(\mathbb{R})\big\},\quad \text{with } \mathcal{L}\varphi := \mu(t,\omega,u)D\varphi(x) + \frac{1}{2}\sigma^2(t,\omega,u)D^2\varphi(x).$$

In particular, the formulation (3.2.8) with all control terms provides a weak formulation of the control problem; the formulation (3.2.8) with all relaxed control terms provides a relaxed formulation of the control problem. Finally, by consider the law the couple (X^ν, B), the strong formulation (3.2.2) of the control problem can also be reformulated as the controlled martingale problem above. In this context, we have the following result.

Theorem 3.2.4. (i) *Assume that the coefficient functions μ and σ are Borel measurable and the controlled SDE above (3.2.2) has at least one solution. Then, the strong, weak and relaxed formulation of the controlled diffusion processes problem all satisfies Assumption 3.2.1.*
(ii) *Under some further regularity conditions on μ, σ and ξ, one can approximate any relaxed control term by strong controls and hence all the three formulations have the same value function.*

Remark 3.2.4. *A very nice property of the relaxed formulation is that the set of all measures induced by the relaxed controls is closed. In many situations, one only needs to check the tightness or relative compactness of a sequence of control terms, and then their limits point is still a relaxed control term. This fact is essentially used in the numerical approximation methods for control problem by Kushner and Dupuis [144]. This property has also been essentially used to study the mean field game problems in a series paper of Carmona, Delarue and Lacker [145, 49, 146], etc.*

3.3 Second order BSDEs, estimation and decomposition of super-solution of BSDEs

To obtain a probabilistic representation for PDEs or path-dependent PDEs in a more general form than the classical Bellman equation (3.1.2), one could consider the problem of controlling a family of BSDEs, which leads to the so-called second order BSDE (2BSDE) of Soner, Touzi and Zhang [186].

Let $\Omega := \mathbb{C}([0,T],\mathbb{R}^d)$ be the canonical space with canonical filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ and canonical process B, let $\mathbb{F}^{\mathbb{P}} = (\mathcal{F}_t^{\mathbb{P}})_{t\in[0,T]}$ be the $\mathbb{P}$-augmented filtration given $\mathbb{P}$ on Ω. We will consider the probability measures $\mathbb{P}$ on Ω, under which B is a semi-martingale with canonical decomposition

$$dB_t = b_t^{\mathbb{P}}dt + dB_t^{c,\mathbb{P}},\quad \text{and}\quad d\langle B^{c,\mathbb{P}}\rangle_t = \hat{a}_t dt\quad \text{under } \mathbb{P},$$

where $B^{c,\mathbb{P}}$ is a continuous local martingale with quadratic variation $\langle B^{c,\mathbb{P}}\rangle_t = \langle B\rangle_t$. Let τ be a $\mathbb{F}$-stopping time, ξ a $\mathcal{F}_\tau^{\mathbb{P}}$-measurable variable, and $f : [0,T]\times\Omega\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ the generator function, we denote by $\mathcal{Y}^{\mathbb{P}}(\tau,\zeta)$ the Y-part of the solution to the BSDE

$$\mathcal{Y}_s = \zeta - \int_s^\tau f(r, B_{r\wedge\cdot}, \mathcal{Y}_s, (\hat{a}_r^{1/2})^\top \mathcal{Z}_r, \hat{a}_r, b_r^{\mathbb{P}})dr - \int_s^\tau \mathcal{Z}_s dB_r^{c,\mathbb{P}} - \int_s^\tau d\mathcal{M}_r,\ \mathbb{P}\text{-a.s.} \quad (3.3.1)$$

Here a solution to the above BSDE is a triple $(\mathcal{Y}, \mathcal{Z}, \mathcal{M})$ in an appropriate space such that (3.3.1) holds, where in particular $\mathcal{M}$ is a martingale orthogonal to $B^{c,\mathbb{P}}$. The wellposedness of the BSDE is ensured under the integrability condition

$$\mathbb{E}^{\mathbb{P}}[|\zeta|^p] < \infty \quad \text{for some} \quad p > 1,$$

and some integrability condition on f and the standard Lipschitz condition of f in (y, z), see e.g. El Karoui and Huang [86].

We will consider a family of measures sets $(\mathcal{P}(t,\omega))_{(t,\omega)\in[0,T]\times\mathbb{R}}$ satisfying Assumption 3.2.1, then given $\xi \in \mathcal{F}_T$, we study the following optimization problem:

$$\widehat{\mathcal{Y}}_t(\omega) \;:=\; \sup_{\mathbb{P}\in\mathcal{P}(t,\omega)} \mathbb{E}^{\mathbb{P}}\big[\mathcal{Y}_t^{\mathbb{P}}(T,\xi)\big]. \tag{3.3.2}$$

In particular, take the controlled diffusion processes in (3.2.3) as example, we can expect that $\widehat{\mathcal{Y}}_t(\omega)$ provides a representation of the nonlinear PDE, in a Markovian context,

$$\partial_t v(t,x) + \sup_{u\in U}\Big(f(\cdot, v, \sigma\partial_x v, \mu(\cdot,u), \sigma^2(\cdot,u)) + \frac{1}{2}\sigma^2(\cdot,u)\partial^2_{xx} v\Big)(t,x) = 0;$$

or of the nonlinear PPDE, in a non-Markovian context,

$$\partial_t v(t,\omega) + \sup_{u\in U}\Big(f(\cdot, v, \sigma\partial_\omega v, \mu(\cdot,u), \sigma^2(\cdot,u)) + \frac{1}{2}\sigma^2(\cdot,u)\partial^2_{\omega\omega} v\Big)(t,\omega) = 0,$$

with terminal condition $v(T,\omega) = \xi(\omega)$ for $\omega \in \Omega$.

2BSDE without regularity on (t,ω) In Possamaï, Tan and Zhou [172], we proved a dynamic programming principle for the control problem (3.3.2) using the measurable selection technique as in Theorem 3.2.2. The measurable selection technique allows to avoid some technical regularity conditions on f and ξ in the original paper of [186].

Theorem 3.3.1. *Assume that the family $(\mathcal{P}(t,\omega))_{(t,\omega)\in[0,T]\times\Omega}$ satisfies Assumption 3.2.1, $\xi : \Omega \to \mathbb{R}$ is Borel measurable, and the generator $f(t,x,y,z,b,a)$ is Borel measurable and Lipschitz in (y,z). Assume in additional some integrality conditions on ξ and f. Then one has the DPP: for all $(t,\omega) \in [0,T]\times\Omega$ and $\mathbb{F}$-stopping time taking value on $[t,T]$,*

$$\widehat{\mathcal{Y}}_t(\omega) \;:=\; \sup_{\mathbb{P}\in\mathcal{P}(t,\omega)} \mathbb{E}^{\mathbb{P}}\big[\mathcal{Y}_t^{\mathbb{P}}(\tau, \widehat{\mathcal{Y}}_\tau)\big].$$

Remark 3.3.1. *The above result is an extension of the DPP result in the linear case in Theorem 3.2.2. To adapt the previous proof, a key step is to construct the solution of the BSDE such that $\mathbb{P} \mapsto \mathbb{E}^{\mathbb{P}}\big[\mathcal{Y}_t^{\mathbb{P}}(T,\xi)\big]$ is Borel measurable.*

Once we have the above dynamic programming principle result, we can follow the same routine in [186] to obtain a wellposedness result for the 2BSDE:

1. First, the dynamic programming principle in Theorem 3.3.1 provides a super-martingale property of process $(\widehat{\mathcal{Y}}_t)_{t\in[0,T]}$ in the nonlinear sense, that is,

$$\widehat{\mathcal{Y}}_t(\omega) \;\geq\; \mathbb{E}^{\mathbb{P}}\big[\mathcal{Y}_t^{\mathbb{P}}(\tau, \widehat{\mathcal{Y}}_\tau)\big|\mathcal{F}_t\big](\omega), \;\; \mathbb{P}\text{-a.s.} \;\; \text{for all } \mathbb{P} \in \mathcal{P}_0.$$

2. Consider the right-continuous version $\widehat{\mathcal{Y}}_t^+$ of $\widehat{\mathcal{Y}}_t$, which is optional w.r.t. the universally augmented filtration $\mathbb{F}^{U,+}$, one hence obtain a strong super-martingale $\widehat{\mathcal{Y}}^+$ in sense that

$$\widehat{\mathcal{Y}}_\sigma^+ \;\geq\; \mathcal{Y}_\sigma^{\mathbb{P}}(\tau, \widehat{\mathcal{Y}}_\tau^+) \text{ } \mathbb{P}\text{-a.s. for all } \mathbb{F}^{U,+}\text{-stopping times } \sigma \leq \tau, \text{ and all } \mathbb{P} \in \mathcal{P}_0.$$

3. Under each fixed $\mathbb{P}$, we consider the reflected BSDE with generator f and càdlàg obstacle $\widehat{\mathcal{Y}}^+$, it follows the Doob-Meyer decomposition for the super-martingale:

$$\widehat{\mathcal{Y}}_t^+ = \xi - \int_t^T f(s, B_{s\wedge\cdot}, \widehat{\mathcal{Y}}_s^+, (\hat{a}_s^{1/2})^\top Z_s^{\mathbb{P}})ds - \int_t^T Z_s^{\mathbb{P}} \cdot dB_s^{c,\mathbb{P}} - \int_t^T dM_s^{\mathbb{P}} + \int_t^T dK_s^{\mathbb{P}},$$

where $M^{\mathbb{P}}$ is a càdlàg martingale orthogonal to $B^{c,\mathbb{P}}$, and $K^{\mathbb{P}}$ is an predictable non-decreasing process, w.r.t. the filtration $\mathbb{F}^{U,+}$.

4. By considering the co-quadratic variation $\langle \widehat{\mathcal{Y}}^+, B \rangle$, one can aggregate the family $(Z^{\mathbb{P}})_{\mathbb{P}\in\mathcal{P}_0}$ into a unique process Z, and it follows a solution to the 2BSDE:

$$\begin{aligned}\widehat{\mathcal{Y}}_t^+ \;&=\; \xi - \int_t^T f(s, B_{s\wedge\cdot}, \widehat{\mathcal{Y}}_s^+, (\hat{a}_s^{1/2})^\top Z_s)ds - \int_t^T Z_s \cdot dB_s^{c,\mathbb{P}} \\ &\qquad - \int_t^T dM_s^{\mathbb{P}} + \int_t^T dK_s^{\mathbb{P}}, \; \mathcal{P}_0\text{-q.s.} \end{aligned} \quad (3.3.3)$$

Theorem 3.3.2. *Under the same technical conditions in Theorem 3.3.1, the 2BSDE (3.3.3) has a unique solution in an appropriate space.*

Remark 3.3.2. *In the above routine, some technical questions motivates the next two works.*

- *First, as we would like to consider a general family of semi-martingale measures $\mathbb{P}$ in $\mathcal{P}(t,\omega)$, for which the augmented filtration $\mathbb{F}^{\mathbb{P}}$ would not be quasi-leftcontinuous in general. Nevertheless, this quasi-leftcontinuous condition is essentially assumed in the previous literature for the a priori estimates and hence the wellposedness of the RBSDE as in Step 3. This motivates us to consider the wellposedness of RBSDEs under general filtration, or more essentially, the a priori estimates to the super-solution of BSDEs.*

- *Without the right-continuous regularization, the original process $\widehat{\mathcal{Y}}$ has already a super-martingale property as shown Step 1. A natural question is that whether one has the Doob-Meyer decomposition without assuming the right-continuity of the super-martingale.*

A priori estimates for super-solution of BSDEs under general filtration Most of the literature on BSDEs, reflected BSDEs (RBSDEs), etc. remains in the context of the Brownian filtration, or in a general filtration by assuming the quasi-leftcontinuous condition, in order to obtain the a priori estimates of super-solutions of the BSDEs. Technically, this condition avoids the jumps of a martingale at predictable times, a key argument used in the classical forward approach to deduce the a priori estimates. The a priori estimates allows to control the norm of all other terms by the norm of the Y-term for a super-solution of the BSDE. It plays an essential role for the wellposedness of the BSDE and RBSDE, and used in various situations such as constrained BSDE in [60], weak BSDE [34], and 2BSDE [186, 172], etc.

In Bouchard, Possamaï, Tan and Zhou [172], we provides another approach to deduce the a priori estimates for super-solutions of the BSDEs. In particular, it allows to bypass the previous technical conditions. Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space, where $\mathbb{F}$ satisfies the usual conditions, and it equipped with a standard d-dimensional Brownian motion W. We have a random variable $\xi \in \mathbb{L}^p$ for some $p > 1$ and a generator function $f : [0,T] \times \Omega \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ such that $f(t, \omega, y, z)$ is Lipschitz in (y, z), and we consider the so-called super-solution of BSDE

$$Y_t = \xi - \int_t^T f_s(Y_s, Z_s) ds - \int_t^T Z_s \cdot dW_s - \int_t^T dM_s + \int_t^T dK_s, \ \mathbb{P}\text{-a.s.}, \tag{3.3.4}$$

where M is a martingale orthogonal to W and K is predictable non-decreasing process.

We also introduce some space with the given $p > 1$ and $\alpha \geq 0$:

- $\mathbb{S}^p$ denotes the space of all optional process Y such that $\|Y\|^p_{\mathbb{S}^p} := \mathbb{E}\big[\sup_{0\leq s\leq T} |Y_t|^p\big] < \infty$.
- $\mathbb{M}^{p,\alpha}$ denotes the space of all c̀dlàg martingales M such that $\|M\|^p_{\mathbb{M}^{p,\alpha}} := \mathbb{E}\Big[\Big(\int_0^T e^{\alpha t} d[M]_t\Big)^{p/2}\Big] < \infty$.
- $\mathbb{H}^{p,\alpha}$ denotes the space of all predictable process such that $\|Z\|^p_{\mathbb{H}^{p,\alpha}} := \mathbb{E}\Big[\Big(\int_0^T e^{\alpha t} \|Z_t\|^2 dt\Big)^{p/2}\Big] < \infty$.
- $\mathbb{I}^{p,\alpha}$ denotes the space of all predictable process K with bounded variation and $K_0 = 0$, and such that $\|K\|^p_{\mathbb{I}^{p,\alpha}} := \mathbb{E}\Big[\Big(\int_0^T e^{\alpha s/2} d\mathrm{TV}(K)_s\Big)^p\Big] < \infty$. Here $\mathrm{TV}(K)$ means the total variation of process K. Denote by $\mathbb{I}^{p,\alpha}_+$ the subset of $\mathbb{I}^{p,\alpha}$ containing all nondecreasing paths.

Recall that in the classical linear case, any uniformly integrable strong super-martingale has the following classical Doob-Meyer(-Mertens) decomposition

$$X_t = X_0 + M_t - K_t - I_t, \tag{3.3.5}$$

where M is a càdlàg martingale, K is a predictable right-continuous non-decreasing process, and I is a predictable quasi-leftcontinuous non-decreasing process, with $M_0 = K_0 = I_0 = 0$.

Remark 3.3.3. *The classical continuous time version of the Doob-Meyer decomposition is on the càdlàg supermartingale. Mertens [152] was the first to obtains the decomposition for general strong supermartingale, which is automatically làdlàg, see also Dellacherie and Meyer [67] for an alternative proof.*

Lemma 3.3.1 (Meyer)**.** *There is a constant $C_p > 0$, such that for any strong supermartingale $X \in \mathbb{S}^p$ with Doob-Meyer decomposition (3.3.5), one has*

$$\|K\|_{\mathbb{I}^p} + \|I\|_{\mathbb{I}^p} + \|M\|_{\mathbb{M}^{p,\alpha}} \ \leq \ C_p \|X\|_{\mathbb{S}^p}.$$

Using the above estimate in Meyer [153], we will consider the nonlinear case and obtain an extended result. Given $\xi \in \mathbb{L}^p$ and a generator function f satisfying the standard Lipschitz condition, we say $(Y, Z, M, K) \in \mathbb{S}^p \times \mathbb{H}^p \times \mathbb{M}^p \times \mathbb{I}^p_+$ is a super-solution of the associated BSDE if (3.3.4) holds. Let us also define $f_0(t, \omega) := f(t, \omega, 0, 0)$, which can be viewed as a process on Ω.

Theorem 3.3.3. (i) *For every $\alpha > 0$, there is some constant $C_{\alpha,p} > 0$ such that for all super-solutions (Y, Z, M, K) of the BSDE, one has*

$$\|Z\|^p_{\mathbb{H}^{p,\alpha}} + \|M\|^p_{\mathbb{M}^{p,\alpha}} + \|K\|^p_{\mathbb{I}^{p,\alpha}} \;\leq\; C_{\alpha,p}\Big(\|Y\|^p_{\mathbb{S}^p} + \|f_0\|^p_{\mathbb{H}^{p,\alpha}_1}\Big).$$

(ii) *Let ξ^1, ξ^2 be two terminal conditions, f^1, f^2 be two generator functions, and (Y^i, Z^i, M^i, K^i) $i = 1, 2$ be the super-solution of the associated BSDEs. Denote by $(\delta Y, \delta Z, \delta M, \delta K)$ their difference. Then for all $\alpha > 0$, there is some constant $C'_{p,\alpha}$ such that*

$$\|\delta Z\|^p_{\mathbb{H}^{p,\alpha}} + \|\delta(M-K)\|^p_{\mathbb{M}^{p,\alpha}} \;\leq\; C'_{p,\alpha}\Big(\|\delta Y\|^p_{\mathbb{S}^p} + \|\delta Y\|^{p/2\wedge(p-1)}_{\mathbb{S}^p} + \|\delta f_0(Y^1, Z^1)\|^p_{\mathbb{H}^{p,\alpha}}\Big).$$

Doob-Meyer's decomposition for (làdlàg) strong super-martingales The procedure above Theorem 3.3.2 is a very classical routine for solving optimal control/stopping problems. Nevertheless, in the case with one reference probability, it is possible to avoid the technical right-continuous regularization in Step 2 and use directly the Doob-Meyer decomposition. This is for example the case of the optimal stopping problem in El Karoui [85]. Indeed, for any stopping time $\tau \in \mathcal{T}$, one can define the value function as a random variable S_τ. By the dynamic programming principle, it follows that family $(S_\tau)_{\tau\in\mathcal{T}}$ is a super-martingale system, i.e. $S_\sigma \geq \mathbb{E}[S_\tau|\mathcal{F}_\sigma]$, $\mathbb{P}$-a.s. Then by Dellacherie and Lenglart [68], one can aggregate this system into a unique optional process X such that $X_\tau = S_\tau$, $\mathbb{P}$-a.s. for every stopping time τ. Moreover, X is a strong super-martingale, which is làdlàg, and one has the Doob-Meyer(-Mertens) decomposition, see Remark 3.3.3.

In the BSDE context, Peng [166] provided a Doob-Meyer decomposition for the so-called càdlàg $\mathcal{E}$-super-martingales in the context of the Brownian filtration. As pointed later by El Karoui, this is in fact a direct consequence of the reflected BSDE using the $\mathcal{E}$-supermartingale as obstacle. In Bouchard, Possamaï and Tan [38], we try to extend this Doob-Meyer's decomposition for general strong $\mathcal{E}$-super-martingales. In particular, using our a prioiri estimates in the previous work [39], we do not need to assume any more the quasi-leftcontinuous condition on the filtration.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a completed probability space, equipped with a filtration satisfying the usual conditions and a standard Brownian motion. We denote by $\mathcal{T}$ the collection of all stopping times taking value in $[0, T]$. Let $\sigma \leq \tau$ be two stopping times, and $\xi \in \mathbb{L}^p(\mathcal{F}_\tau)$, we define

$$\mathcal{E}_{\sigma,\tau}[\xi] := Y_\sigma, \quad \text{with } Y_t = \xi - \int_{t\wedge\tau}^{\tau} f_s(Y_s, Z_s)ds - \int_{t\wedge\tau}^{\tau} Z_s \cdot dW_s - \int_{t\wedge\tau}^{\tau} dM_s,$$

where $f : [0, T] \times \Omega \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ is the generator function, M is a martingale orthogonal to W. The wellposedness of the above BSDE is ensured by the $\mathbb{L}^p(\mathcal{F}_\tau)$ condition on ξ.

Theorem 3.3.4. *Let $\{S(\tau), \tau \in \mathcal{T}\}$ be a family such that $S(\tau) \in \mathbb{L}^p(\mathcal{F}_\tau)$, $S(\tau) = S(\tau')$ a.s. on $\tau = \tau'$, and $S(\sigma) \geq \mathcal{E}_{\sigma,\tau}[S(\tau)]$ for every $\sigma \leq \tau$.*

(i) *Assume that $\{S(\tau), \tau \in \mathcal{T}\}$ is uniformly integrable. Then there is an optional process X such that $S(\tau) = X_\tau$ for every $\tau \in \mathcal{T}$.*

(ii) *Assume in addition that esssup$\{S(\tau), \tau \in \mathcal{T}\} \in \mathbb{L}^p$, there there exists $Z \in \mathbb{H}^{p,0}$, $A \in \mathbb{I}^{p,0}_+$ and a càdlàg martingale $M \in \mathbb{M}^{p,0}$ orthogonal to W such that for all $\sigma \leq \tau \in \mathcal{T}$,*

$$X_\sigma = X_\tau - \int_\sigma^\tau f_s(X_s, Z_s)ds + \int_\sigma^\tau dA_s - \int_\sigma^\tau Z_s \cdot dW_s - \int_\sigma^\tau dM_s.$$

Remark 3.3.4. (i) *In a parallel work of Grigorova, Imkeller, Offen, Ouknine and Quenez [104], and in the context of the Brownian filtration with $p = 2$, the authors obtained a wellposedeness result of reflected BSDE with right upper-semicontinuous obstacle. In particular, their result induces the above Doob-Meyer decomposition in their context, which is less general than ours in Theorem 3.3.4.*

(ii) *As applications, we provide an optional decomposition for $\mathcal{E}$-supermaritingale systems, as well as a Dual formulation for minimal super-solution of BSDEs with constraints on the gains process studies by Cvitanić, Karatzas and Soner [60]. With the above Doob-Meyer decomposition, one avoids the tedious right-continuous regularization step as in Step 2 above Theorem 3.3.2.*

3.4 Numerical approximations

The numerical methods have been developed along the stochastic control theory. It plays an important role for the application of the optimal control theory, as soon as an explicit solution is not available. There are mainly two approaches: the probabilistic approach and the PDE approach, to prove the convergence.

For the probabilistic approach, Kushner and Dupuis [144] have constructed a controlled Markov chain to approximate the controlled diffusion process. Using the weak convergence argument, the value of the controlled Markov chain problem converges to the value of the controlled diffusion processes problem. Moreover, their controlled Markov chain system can be interpreted as a finite difference numerical scheme. The weak convergence technique provides only a general convergence result. Recently, Dolinsky [70] used the strong invariance principle technique in Sakhanenko [184], and obtained a convergence rate for a class of control problems.

The PDE approach goes back to the seminal paper of Barles and Souganidis [9]. Recall that, in general, the value function of a control problem can be characterized by a Bellman equation in form (3.1.2) and in sense of viscosity solution. In [9], Barles and Souganidis introduced three sufficient conditions: consistency, monotonicity and stability conditions, to ensure the convergence of a numerical scheme. Because of the lack of regularity of the viscosity solution of a PDE, the convergence rate is more difficulty. Krylov [142] introduced a perturbation technique on the Bellman equation, which allows to construct a smooth sub and super-solution closed to the original unique viscosity solution of the equation. By analyzing the local error of the numerical scheme on the sub and super-solution, and then optimizing the perturbation parameters, one obtains a convergence rate. This convergence rate is later improved by Krylov [143], Barles and Jakobsen [8], etc.

Further, with the development of the BSDE theory, the numerical methods for BSDEs have been studied a lot. In particular, it provides numerical schemes for a class of control problems, where only the drift coefficient function is controlled. Moreover, it could work in a general non-Markovian context. Bally and Pagès [6], Ma, Protter, San Martín and Torres [149] were the first to study the numerical methods for BSDEs when the generator function depends only on y. For general BSDEs whose generator depends on (y, z), the breakthroughs have been made by Bouchard and Touzi [37], Zhang [199], where a convergence rate has been obtained. Let us stay in the Markov context as in BSDE (3.1.3) and present their numerical scheme: let $0 = t_0 < t_1 < \cdots < t_n = T$ be a time discretization of $[0, T]$, with $t_i := i\Delta t$ and $\Delta t := T/n$, one first define a forward SDEs by its Euler scheme

$$X^\Delta_{t_{i+1}} \;=\; X^\Delta_{t_i} \;+\; \sigma(t_i, X^\Delta_{t_i})\Delta W_{i+1}, \quad \text{with} \;\; \Delta W_{i+1} := W_{t_{i+1}} - W_{t_i}.$$

Then one can compute (Y^Δ, Z^Δ) on the discrete time grid $(t_i)_{0\le i\le n}$ by a backward iteration: $Y^\Delta_{t_n} := g(X^\Delta_{t_n})$, and

$$Z^\Delta_{t_i} \;:=\; \mathbb{E}_{t_i}\Big[Y^\Delta_{t_{i+1}}\frac{\Delta W_{i+1}}{\Delta t}\Big], \qquad Y^\Delta_{t_i} \;:=\; \mathbb{E}_{t_i}\big[Y^\Delta_{t_{i+1}}\big] + f\big(t_i, X^\Delta_{t_i}, \mathbb{E}_{t_i}[Y^\Delta_{t_{i+1}}], Z^\Delta_{t_i}\big)\Delta t, \tag{3.4.1}$$

where $\mathbb{E}_{t_i}[\cdot] := \mathbb{E}[\cdot|\mathcal{F}_{t_i}]$ denotes the conditional expectation. In practice, one can use a simulation-regression method such as in [148, 103, 42], etc. to estimate the conditional expectation terms. Since then, a large stream of literature on the numerical methods for BSDEs has been generated.

Let us come back to the concrete numerical schemes of the optimal control problem. The most classical scheme should the finite difference scheme. To obtain the finite difference scheme for the one-dimensional case with HJB equation (3.1.2), the first step is to approximate the derivatives by their discrete counterpart on a discrete grid $\{(t_i, x_j)\}_{i,j}$:

$$\partial_t v(t_i,x_j) \approx \frac{v(t_i,x_j) - v(t_{i-1},x_j)}{\Delta t}, \;\; \partial_x v(t_i,x_j) \approx \frac{v(t_i,x_{j+1}) - v(t_i,x_j)}{\Delta x} \approx \frac{v(t_i,x_j) - v(t_i,x_{j-1})}{\Delta x},$$

and

$$\partial^2_{xx} v(t_i,x_j) \;\approx\; \frac{v(t_i,x_{j+1}) - 2v(t_i,x_j) + v(t_i,x_{j-1})}{2\Delta x^2}. \tag{3.4.2}$$

Next, plugging the above approximating into the original equation (3.1.2), it follows the the scheme

$$\begin{aligned} v(t_{i-1},x_j) \;=\; \sup_{u\in U}\Big\{ \Big(\frac{\mu(\cdot,u)\Delta t}{\Delta x} + \frac{\sigma^2(\cdot,u)\Delta t}{2\Delta x^2}\Big) v(t_i,x_{j+1}) + \frac{\sigma^2(\cdot,u)\Delta t}{2\Delta x^2} v(t_i,x_{j-1}) \\ + \Big(1 - \frac{\mu(\cdot,u)\Delta t}{\Delta x} - \frac{\sigma^2(\cdot,u)\Delta t}{\Delta x^2}\Big) v(t_i,x_j) + f(t_i,x_j,u)\Delta t \Big\}. \end{aligned} \tag{3.4.3}$$

In the three convergence criteria of Barles and Souganidis [9], the consistency is ensured by the approximation in and above (3.4.2), the monotonicity will be ensured by the so-called

CFL condition

$$1 - \frac{\mu(\cdot,u)\Delta t}{\Delta x} - \frac{\sigma^2(\cdot,u)\Delta t}{\Delta x^2} \;\geq\; 0. \tag{3.4.4}$$

This is exactly the **same condition** in Kushner and Dupuis [144] to ensure that the three coefficients before $v(t_i,x_{j+1})$, $v(t_i,x_j)$ and $v(t_i,x_{j-1})$ are all positive and with sum 1 to obtain a probability function, so that one can interpret the finite difference scheme as a control Markov chain system: the system moves from (t_{i-1},x_j) to $\{(t_i,x_{j+1}),\ (t_i,x_j),\ (t_i,x_{j-1})\}$ with different probabilities. In high dimensional case, the finite difference scheme is more difficult to keep monotone, except in the diagonal dominated case. Bonnans, Ottenwaelter and Zidani [27, 30] introduced a generalized finite difference scheme, where the coefficients can be computed in $d=2$ case to obtain a monotone scheme.

To solve higher dimensional problems, one may expect to use Monte-Carlo methods. Inspired from the numerical scheme (3.4.1) of BSDEs (which is a semilinear PDE in the Markov case), Fahim, Touzi and Warin [93] introduced a new scheme for a class of fully nonlinear parabolic PDEs:

$$\partial_t v + \frac{1}{2}\sigma^2(\cdot)D^2 v + F(\cdot,v,Dv,D^2v) \;=\; 0, \qquad v(T,\cdot) = g(\cdot). \tag{3.4.5}$$

The numerical scheme is as follows: let $Y^\Delta_{t_n} := g(X^\Delta_{t_n})$, we then compute $Y^\Delta_{t_i}$ using the backward iteration:

$$\begin{cases} Z^\Delta_{t_i} \;:=\; \mathbb{E}_{t_i}\Big[Y^\Delta_{t_{i+1}}\big(\sigma(t_i,X^\Delta_{t_i})^\top\big)^{-1}\frac{\Delta W_{i+1}}{\Delta t}\Big], \\ \Gamma^\Delta_{t_i} \;:=\; \mathbb{E}_{t_i}\Big[Y^\Delta_{t_{i+1}}\big(\sigma(t_i,X^\Delta_{t_i})^\top\big)^{-1}\frac{\Delta W^2_{i+1}-\Delta t}{\Delta t^2}\sigma(t_i,X^\Delta_{t_i})^{-1}\Big], \\ Y^\Delta_{t_i} \;:=\; \mathbb{E}_{t_i}\big[Y^\Delta_{t_{i+1}}\big] + F\big(t_i,X^\Delta_{t_i},\mathbb{E}_{t_i}[Y^\Delta_{t_{i+1}}],Z^\Delta_{t_i},\Gamma^\Delta_{t_i}\big)\Delta t. \end{cases} \tag{3.4.6}$$

Under some technical conditions, they show that the above scheme satisfies the consistency, monotonicity and stability conditions of Barles and Souganidis [9] and hence obtain a convergence result. Restrict to the HJB equations, they also apply the perturbation techniques of Krylov [142] to obtain a convergence rate. An extension for degenerate nonlinear PDEs has been made in Tan [190]. For similar nonlinear PDEs, more monotone numerical schemes have been proposed, let us cite the semi-Lagrangian scheme in Debrabant and Jakobsen [65], the trinomial tree scheme of Guo, Zhang and Zhuo [108], the switching system scheme of Kharroubi, Langrené and Pham [138], etc. Notice that all the above schemes have been analysed in a Markov context, with the numerical analysis technique of Barles and Souganidis [9], or the Perturbation technique as in Krylov [142] and Barles and Jacobsen [8].

My work in this subject concentrates mainly on the generalization of the classical numerical analysis techniques to the non-Markovian case. As a consequence, it allows to extend all those numerical schemes introduced in a Markov context to the non-Markovian context.

A numerical scheme for non-Markovian stochastic control problems In the paper of Tan [191], we worked on the numerical scheme of Fahim, Touzi and Warin [93], that introduced for nonlinear PDEs. As recalled above for the finite difference scheme, the CFL condition (3.4.4) used to ensure the monotonicity condition of Barles and Souganidis [9] is in fact equivalent to the technical condition used in Kushner and Dupuis [144] for the existence of the controlled Markov chain. The main contribution of [191] is to provide a re-interpretation of the numerical scheme of Fahim, Touzi and Warin [93] as a controlled Markov chain in the context of an optimal control problem, under the same technical monotone conditions. This allows to extend it for a class of non-Markovian control problem in form of (3.2.2). For the non-Markovian control problem (3.2.2), our scheme is given as follows: First, we introduce an uncontrolled volatility function $\sigma_0 : [0,T] \times \Omega \to \mathbb{R}$, and simulate the associated solution of SDE $dX_t = \sigma_0(t, X_{t\wedge\cdot})dW_t$ by its Eurler scheme

$$X^{\Delta}_{t_{i+1}} \;=\; X^{\Delta}_{t_i} \;+\; \sigma_0(t, \widehat{X}^{\Delta}_{t_i\wedge\cdot})\Delta W_{i+1},$$

where $\widehat{X}^{\Delta}$ denotes the continuous time path obtained by linear interpolation of $(X^{\Delta}_{t_0}, \cdots, X^{\Delta}_{t_i})$. Recall that μ and σ are the controlled drift and volatility coefficient functions for SDE (3.2.1), we introduce

$$b^{t,\omega}_u := \mu(t,\omega,u), \quad a^{t,\omega}_u := \sigma^2(t,\omega,u) - \sigma_0^2(t,\omega) \;\text{ and }\; F(t,\omega,z,\gamma) := \sup_{u\in U}\big(b^{t,\omega}_u z + \frac{1}{2}a^{t,\omega}_u \gamma\big).$$

As a non-Markovian extension of Fahim, Touzi and Warin's [93] scheme, our numerical solution is computed in the backward iteration: Let $Y^{\Delta}_{t_n} := \xi(\widehat{X}^{\Delta}_{\cdot})$, and then define $Y^{\Delta}_{t_i}$ by induction:

$$\begin{cases} Z^{\Delta}_{t_i} \;:=\; \mathbb{E}_{t_i}\Big[Y^{\Delta}_{t_{i+1}}\big(\sigma_0(t_i,\widehat{X}^{\Delta}_{t_i})^{\top}\big)^{-1}\frac{\Delta W_{i+1}}{\Delta t}\Big], \\ \Gamma^{\Delta}_{t_i} \;:=\; \mathbb{E}_{t_i}\Big[Y^{\Delta}_{t_{i+1}}\big(\sigma_0(t_i,\widehat{X}^{\Delta}_{t_i})^{\top}\big)^{-1}\frac{\Delta W^2_{i+1}-\Delta t}{\Delta t^2}\sigma_0(t_i,\widehat{X}^{\Delta}_{t_i})^{-1}\Big], \\ Y^{\Delta}_{t_i} \;:=\; \mathbb{E}_{t_i}\big[Y^{\Delta}_{t_{i+1}}\big] + F\big(t_i,\widehat{X}^{\Delta}_{t_i}, Z^{\Delta}_{t_i}, \Gamma^{\Delta}_{t_i}\big)\Delta t. \end{cases} \tag{3.4.7}$$

Notice that the above scheme can be implemented together with a simulation-regression technique to estimate the conditional expectation terms, similar to the numerical schemes of the BSDEs (3.4.1).

To prove the convergence, we adapted Kushner and Dupuis's [144] weak convergence technique in our non-Markovian context. When the controlled coefficient functions μ and σ do not dependent on (t, X_t), we also adapted Dolinsky's [70] strong invariance principle approach to provide a convergence rate result.

Weak approximation of 2BSDEs With the above numerical scheme for non-Markovian control problem, it is natural to consider the numerical approximation of the second order BSDEs, which is basically a control problem on a family of BSDEs. It would be easy to extend formally the above schemes (3.4.1) and (3.4.7) for BSDEs and non-Markov control problem to the case of 2BSDEs. However, the main difficulty here is to extend the weak convergence technique to this nonlinear context. From another point of view, the weak convergence would be an interesting property for the 2BSDE. In the work of Possamaï and Tan [171], our aim is to develop a weak convergence property for the 2BSDE as an extension of the classical Donsker's theorem.

The classical Donsker's [76] theorem is given as follows: Let $(X_i)_{i\geq 1}$ be a sequence of i.i.d. random variables with mean 0 and variance 1, we define the sum $S_n := \sum_{i=1}^n X_i$ and the scaled sum $W_t^n := S_{\lfloor nt \rfloor}/\sqrt{n}$, $t \geq 0$, then the process $W_\cdot^n$ converges weakly to a standard Brownian motion $W_\cdot$. In particular, given a bounded continuous variable $\xi : \Omega \to \mathbb{R}$ with canonical space $\Omega := \mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)$ of càdlàg paths, one obtains that $\mathbb{E}[\xi(W_\cdot^n)] \to \mathbb{E}[\xi(W_\cdot)]$ as $n \to \infty$. Recall that the solution of BSDE with terminal condition ξ can be considered as a nonlinear expectation on ξ. This motivates Briand, Delyon and Mémin's [44] work to provide an extension of the Donsker's theorem to the BSDE case. Namely, let $(W^n)_{n\geq 1}$ be a sequence as above which converges weakly to the Brownian motion W, and let (Y, Z) (resp. (Y^n, Z^n)) be the solution of the BSDE (3.1.3) generated by W (resp. W^n), it is proved in [44] that (Y^n, Z^n) converges to (Y, Z) in a weak sense. For the G-expectation introduced by Peng [167], Dolinsky, Nutz and Soner [71] studied its weak convergence property. As the G-expectation is the supremum of expectations under a family of probability measures on Ω, by considering the supremum of expectation under an appropriate family of probability measures on a discrete time canonical space, they obtained an weak approximation result as in Donsker's theorem. Notice that the G-expectation is in fact a simple non-Markovian control problem, where the drift function is 0, and the volatility function is bounded between two constants $\sigma_0 < \sigma_1$. Their proof is in fact in the same spirit of the weak convergence argument as in Kushner and Dupuis [144].

In view of the above results, our main objective in [171] is to extend the Donsker's type result for the 2BSDE. Notice that the solution of a 2BSDE is in fact the supremum of solutions of BSDEs under a family of probability measures $\mathcal{P}$, see (3.3.3), where $\mathcal{P}$ could be a family of probability measures induced by the controlled diffusion processes as in (3.2.3). Given a family of discrete time controlled processes $(M_\cdot^{\nu,\pi})_\pi$, where π denotes the time discretization with time step converging to 0, such that

- Any sequence of controlled processes $(M_\cdot^{\nu_n,\pi_n})_{n\geq 0}$ with $|\pi_n| \to 0$ is tight on the space of continuous time pahs, and any continuous time limit lies in $\mathcal{P}$.
- Any probability in $\mathcal{P}$ can be approximated weakly by the discrete time controlled processes in the family $(M_\cdot^{\nu,\pi})_\pi$.

Let us consider the BSDEs generated by the family $(M_\cdot^{\nu,\pi})_\pi$:

$$\mathcal{Y}_t^{\nu,\pi} = \xi(M_\cdot^{\nu,\pi}) - \int_t^T f(s, M_{s\wedge\cdot}^{\nu,\pi}, \mathcal{Y}_s^{\nu,\pi}, \mathcal{Z}_s^{\nu,\pi}, \nu_s)ds - \int_t^T \mathcal{Z}_s^{\nu,\pi} dM_s^{\nu,\pi} - \int_t^T dN_s^{\nu,\pi}, \tag{3.4.8}$$

and denote $Y_0^\pi := \sup_\nu \mathcal{Y}_0^{\nu,\pi}$. Using results and techniques as in Ma, Protter, San Martín and Torres [149] and Briand, Delyon and Mémin [44], we have the following convergence result:

Theorem 3.4.1. *Under some regularity conditions, we have* $\liminf_{|\pi|\to 0} Y_0^\pi \geq Y_0$*, where* Y *is the* Y*-part of the solution to the 2BSDE* (3.3.3)*. Assume in addition that* f *does not dependent on* z*, then*

$$\lim_{|\pi|\to 0} Y_0^\pi = Y_0.$$

Remark 3.4.1. (i) *In the above result, the final convergence result is only obtained when f is independent of z. The main technical reason is that one cannot have the convergence of Z under the augmented filtration $\mathbb{F}^{\mathbb{P}}$ for a general probability $\mathbb{P}$. Indeed, the process Z in the solution of the BSDE is intuitively the integral part in the martingale representation theorem, counter-examples can be found in e.g. [130].*

(ii) *As applications, we can choose an appropriate family $(M_{\cdot}^{\nu,\pi})$ such that the discrete time BSDE (3.4.8) can be reformulated as a numerical scheme, and the above result ensures its convergence. In particular, it covers the finite-difference scheme as in (3.4.3), and the Fahim-Touzi-Warin scheme as in (3.4.6), etc.*

Numerical analysis for path-dependent PDEs The weak convergence techniques could be very powerful in numerical analysis for non-Markovian control problems, but it has its limit in many cases such as the 2BSDE, the differential games, etc. With the development of the viscosity solution theory of the path-dependent PDE (PPDE) in [82], etc., it is natural to try to extend the seminal work of Barles and Souganidis [9] to the path-dependent case. As the PPDE describes the value function for a large class of non-Markovian control problems, the 2BSDE, the differential games, etc., such an extension could provide new numerical schemes and convergence results for these non-Markovian problems. This is the main objective of the work in Ren and Tan [174].

Let us first recall the seminal work of Barles and Souganidis [9] on the monotone scheme for numerical approximation of the viscosity solution of the PDE. We will restrict to the parabolic PDE:

$$\mathsf{L}v(t,x) \;:=\; -\partial_t v(t,x) - G_0(\cdot, v, \partial_x v, \partial^2_{xx} v)(t,x) \;=\; 0, \quad \text{on} \quad [0,T)\times\mathbb{R}^d, \quad (3.4.9)$$

with the terminal condition $u(T,\cdot) = g$. A function u is said to be a viscosity super-solution (resp. sub-solution) of PDE (3.4.9) if for any point $(t,x) \in (0,T)\times\mathbb{R}^d$ and any function $\phi^{\alpha,\beta,\gamma}(s,y) := \alpha s + \beta\cdot y + \frac{1}{2}\gamma : (yy^\top)$ such that $(u - \phi^{\alpha,\beta,\gamma})(s,y)$ has a local minimum (resp. maximum) at (t,x), one has $-\alpha - G_0(t,x,u(t,x),\beta,\gamma) \geq 0$ (resp. ≤ 0). Then u is a viscosity solution of (3.4.9) if it is at the same time super- and sub-solutions, see e.g. Crandall, Ishii and Lions [59].

Assumption 3.4.1. (i) *The terminal condition g is bounded continuous.*
(ii) *The function G_0 is continuous and $G_0(t,x,y,z,\gamma)$ is non-decreasing in γ.*
(iii) *PDE (3.4.9) admits a comparison principle for bounded viscosity solution, i.e. if u, v are bounded viscosity subsolution and supersolution to PDE (3.4.9), respectively, and $u(T,\cdot) \leq v(T,\cdot)$, then $u \leq v$ on $[0,T]\times\mathbb{R}^d$.*

For any $t \in [t,T)$ and $h \in (0, T-t]$, let $\mathbb{T}^{t,x}_h$ be an operator on the set of bounded functions defined on $\mathbb{R}^d$. For $n \geq 1$, denote $h := \frac{T}{n} < T - t$, $t_i = ih$, $i = 0,1,\cdots,n$, let the numerical solution be defined by

$$u^h(T,x) := g(x), \quad u^h(t,x) := \mathbb{T}^{t,x}_h[u^h(t+h,\cdot)], \quad t\in[0,T), \quad i = n,\cdots,1.$$

Assumption 3.4.2. (i) *Consistency: for any $(t,x)\in[0,T)\times\mathbb{R}^d$ and any smooth function $\phi \in C^{1,2}([0,T)\times\mathbb{R}^d)$,*

$$\lim_{(t',x',h,c)\to(t,x,0,0)} \frac{(c+\phi)(t',x') - \mathbb{T}^{t',x'}_h\big[(c+\phi)(t'+h,\cdot)\big]}{h} \;=\; \mathsf{L}\phi(t,x).$$

(ii) *Monotonicity:* $\mathbb{T}_h^{t,x}[\phi] \le \mathbb{T}_h^{t,x}[\psi]$ *whenever* $\phi \le \psi$.
(iii) *Stability:* u^h *is bounded uniformly in* h *whenever* g *is bounded.*
(iv) *Boundary condition:* $\lim_{(t',x',h)\to(T,x,0)} u^h(t',x') = g(x)$ *for any* $x \in \mathbb{R}^d$.

Theorem 3.4.2 (Barles and Souganidis)**.** *Let the generator function* G_0 *in* (3.4.9) *and the terminal condition* g *satisfy Assumption 3.4.1, and the numerical scheme* $\mathbb{T}_h^{t,x}$ *satisfies Assumption 3.4.2. Then the parabolic PDE* (3.4.9) *has a unique bounded viscosity solution* u *and the numerical solution* u^h *converges to* u *locally uniformly as* $h \to 0$.

Remark 3.4.2. *In Barles and Souganidis [9], the results are given for more general PDEs instead of the parabolic PDEs.*

Let us now consider the PPDE defined on $[0,T]\times\Omega$, where $\Omega = \mathbb{C}([0,T],\mathbb{R}^d)$ is the canonical space with continuous paths on $[0,T]$,

$$- \partial_t u(t,\omega) \; - \; G\big(\cdot, u, \partial_\omega u, \partial^2_{\omega\omega} u\big)(t,\omega) \; = \; 0, \quad \text{for all } \; (t,\omega) \in [0,T) \times \Omega, \qquad (3.4.10)$$

with the terminal condition $u(T,\cdot) = \xi(\cdot)$. The main difference in the definition of viscosity solution of PDE and that of PPDE is that the test functions ϕ are dominated by u for every (t,x) in the PDE case, but are dominated in expectation in the PPDE case. More precisely, the test functions to define the viscosity solution of PPDE is given as follows. Let $\mathcal{P}$ be the family of all semi-martingale measures $\mathbb{P}$ on Ω under which the canonical process B has the canonical decomposition $B_t = B_0 + A_t^{\mathbb{P}} + M_t^{\mathbb{P}}$ with $\|\frac{dA_t^{\mathbb{P}}}{dt}\| \le L$ and $\|\frac{d\langle M^{\mathbb{P}}\rangle_t}{dt}\| \le L$. We denote by $\overline{\mathcal{E}}[\cdot] := \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}^{\mathbb{P}}[\cdot]$ and $\underline{\mathcal{E}}[\cdot] := \inf_{\mathbb{P}\in\mathcal{P}} \mathbb{E}^{\mathbb{P}}[\cdot]$, and the introduce the class of test functions:

$$\underline{\mathcal{J}}u(t,\omega) := \Big\{(\alpha,\beta,\gamma) : u(t,\omega) = \max\nolimits_{\tau\in\mathcal{T}_{H_\delta}} \overline{\mathcal{E}}[u_\tau^{t,\omega} - \phi_\tau^{\alpha,\beta,\gamma}], \text{ for some } \delta > 0\Big\},$$
$$\overline{\mathcal{J}}u(t,\omega) := \Big\{(\alpha,\beta,\gamma) : u(t,\omega) = \min\nolimits_{\tau\in\mathcal{T}_{H_\delta}} \underline{\mathcal{E}}[u_\tau^{t,\omega} - \phi_\tau^{\alpha,\beta,\gamma}], \text{ for some } \delta > 0\Big\},$$

where $H_\delta(\omega') := \delta \wedge \inf\{s \ge 0 : |\omega'_s| \ge \delta\} \in \mathcal{T}^+$. Notice that in above, the optimal stopping problems described the dominance of u by ϕ in sense of expectation.

Definition 3.4.1. *Let* $u : [0,T]\times\Omega \to \mathbb{R}$ *be a bounded uniformly continuous functions.*
(i). *u is a $\mathcal{P}$-viscosity subsolution (resp. supersolution) of the PPDE* (3.4.10)*, if at any point* $(t,\omega) \in [0,T)\times\Omega$ *it holds for all* $(\alpha,\beta,\gamma) \in \underline{\mathcal{J}}u(t,\omega)$ *(resp.* $\overline{\mathcal{J}}u(t,\omega)$*) that*

$$-\alpha - G(t,\omega,u(t,\omega),\beta,\gamma) \;\le \text{ (resp. } \ge \text{) } 0.$$

(ii). *u is a $\mathcal{P}$-viscosity solution of the PPDE* (3.4.10)*, if u is both a $\mathcal{P}$-viscosity subsolution and a $\mathcal{P}$-viscosity supersolution of* (3.4.10).

Let us consider a numerical scheme denoted by $\mathbb{T}$: for each $(t,\omega) \in [0,T)\times\Omega$ and $0 < h \le T - t$, $\mathbb{T}_h^{t,\omega}$ is a function from $\mathbb{L}^0(\mathcal{F}_{t+h})$ to $\mathbb{R}$, the numerical scheme is given by the backward iteration: $u^h(T,\cdot) = \xi(\cdot)$ and

$$u^h(t,\omega) \;\; := \;\; \mathbb{T}_h^{t,\omega}\big[u^h(t+h,\cdot)\big].$$

Remember that for the viscosity solution of the PPDE, the domination of test function ϕ on value function u is only given in sense of expectation, our key step is to find a good reformulation of the monotonicity condition in Barles and Souganidis [9] in our context.

Definition 3.4.2. *Let $\{U_i, i \geq 1\}$ be a sequence of independent random variables defined on a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$, such that each U_i follows the uniform distribution on $[0,1]$. Let $h > 0$, K be a subset of a metric space, $\mathbb{F}_h : K \times [0,1] \to \mathbb{R}$ be a Borel measurable function such that for all $k \in K$ we have*

$$|\tilde{\mathbb{E}}\big[\mathbb{F}_h(k,U)\big]| \ \leq \ Lh, \quad \mathrm{Var}\big[\mathbb{F}_h(k,U)\big] \leq Lh \quad \text{and} \quad \tilde{\mathbb{E}}\big[\mathbb{F}_h(k,U)^3\big] \leq Lh^{3/2}.$$

Denote the filtration $\tilde{\mathbb{F}} := \{\tilde{\mathcal{F}}_i, i \in \mathbb{N}\}$, where $\tilde{\mathcal{F}}_n := \sigma\{U_i, i \leq n\}$. Let $\mathcal{K} = \{\nu : \nu_{ih}$ is $\tilde{\mathcal{F}}_i$-measurable and takes values in K for all $i \in \mathbb{N}\}$. For all $\nu \in \mathcal{K}$, we define

$$X_0^{h,\nu} := 0, \quad X_{ih}^{h,\nu} \ = \ X_{(i-1)h}^{h,\nu} \ + \ \mathbb{F}_h(\nu_{ih}, U_i) \ \text{ for } \ i \geq 1.$$

Further, we denote by $\hat{X}^{h,\nu} : [0,T] \times \tilde{\Omega} \to \Omega$ the linear interpolation of the discrete process $\{X_{ih}^{h,\nu}, i \in \mathbb{N}\}$ such that $\hat{X}_{ih}^{h,\nu} = X_{ih}^{h,\nu}$ for all i. Finally, for any function $\phi \in \mathbb{L}^0(\mathcal{F})$, we define the nonlinear expectation:

$$\underline{\mathcal{E}}_h[\phi] \ := \ \inf_{\nu\in\mathcal{K}} \tilde{\mathbb{E}}\Big[\phi\big(\hat{X}^{h,\nu}\big)\Big] \quad \text{and} \quad \overline{\mathcal{E}}_h[\phi] \ := \ \sup_{\nu\in\mathcal{K}} \tilde{\mathbb{E}}\Big[\phi\big(\hat{X}^{h,\nu}\big)\Big].$$

Our conditions on the numerical scheme $\mathbb{T}$ are as follows

Assumption 3.4.3. (i) Consistency: for every $(t,\omega) \in [0,T) \times \Omega$ and $\phi \in C_0^{1,2}(\mathbb{R}^+ \times \mathbb{R}^d)$,

$$\begin{aligned}&\lim_{(t',\omega',h,c)\to(t,0,0,0)} \frac{\phi(t', (\omega \otimes_t \omega')_{t'}) + c - \mathbb{T}_h^{t',\omega\otimes_t\omega'}\big[\phi(t'+h,\cdot)+c\big]}{h} \\ &= \ \partial_t\phi(0,\omega_t) - G(t,\omega,\phi(0,\omega_t), D\phi(0,\omega_t), D^2\phi(0,\omega_t)).\end{aligned}$$

(ii) Monotonicity: there exists a nonlinear expectation $\underline{\mathcal{E}}_h$ as in Definition 3.4.2 such that, for any $\varphi, \psi \in \mathbb{L}^0(\mathcal{F}_{t+h})$ we have

$$\mathbb{T}_h^{t,\omega}[\varphi] - \mathbb{T}_h^{t,\omega}[\psi] \geq -h\bar{\delta}(h) \quad \text{whenever} \quad \inf_{0\leq\alpha\leq L} \underline{\mathcal{E}}_h\big[e^{\alpha h}(\phi-\psi)^{t,\omega}\big] \geq 0.$$

where $\bar{\delta} : \mathbb{R}^+ \to \mathbb{R}$ is a function such that $\lim_{h\downarrow 0} \bar{\delta}(h) = 0$.
(iii) Stability: u^h are uniformly bounded and uniformly continuous in (t,ω), uniformly in h.

Our main result is the following convergence of the monotone scheme for PPDE (3.4.10).

Theorem 3.4.3. *Assume that*

- *G and ξ in PPDE (3.4.10) satisfies some technical conditions;*
- *the numerical scheme $\mathbb{T}_h$ satisfies Assumption 3.4.3;*
- *the comparison of the viscosity sub- and super-solutions of PPDE (3.4.10) holds true, i.e. if u, v are $\mathcal{P}$-viscosity subsolution and supersolution to PPDE (3.4.10), respectively, and $u(T,\cdot) \leq v(T,\cdot)$, then $u \leq v$ on $[0,T] \times \Omega$.*

Then PPDE (3.4.10) admits a unique bounded $\mathcal{P}$-viscosity solution u, and

$$u^h \ \to \ u \quad \text{locally uniformly,} \quad \text{as} \quad h \to 0.$$

Remark 3.4.3. *Zhang and Zuo [200] were the first to make the extension of Barles and Souganidis's result to the PPDE case. However, their convergence conditions are not satisfied by most of the classical numerical schemes, such as the finite difference scheme, the Fahim-Touzi-Warin scheme, etc. Our reformulation of the conditions allows to cover all the numerical schemes (to the best of our knowledge) in the context of stochastic control theory. Our proof is also quite different since the assumptions are different.*

3.5 Perspectives

As summarized at the beginning of the chapter, the stochastic control problem may have numerous different formulations. My previous work concentrates mainly on the non-Markovian control problems in a very standard formulation: the controlled diffusion processes problem, in a finite horizon, without any constraint on the control processes. It would be quite interesting to extend some techniques as well as results to the variated formulations of the non-Markovian control problem.

For a concrete project, with Dylan Possamaï and Fabrice Djete, we are interested in the McKean-Vlasov control problem, which is notably studied in a series of recent papers of Hûyen Pham and his co-authors. As a first objective, we aim to deduce a dynamic programming principle for the McKean-Vlasov control problem under minimal conditions, using the measurable selection argument. Secondly, we will consider its numerical approximation, which would be important for its applications.

Chapitre 4

Branching diffusion representation for nonlinear PDEs

4.1 Introduction

The branching stochastic process has always been an important topic in mathematics. It is notably motivated by its applications in biology to study the evolution of the population of some living species. For example, the most elementary discrete time branching process, Galton-Watson process, was initially introduced in 19th century to study the probability of the extinction of family names. Embedding the Galton-Watson process into a continuous time framework, where each particle has an independent life time of exponential distribution, one obtains a continuous time Markov branching process. The life time of particles may follow other distributions than the exponential distribution, then it loses the Markov property and becomes an age-dependent process. Let us refer to Athreya and Ney [3] for a presentation of the basic properties of different basic branching stochastic processes.

With the development of the diffusion process theory, it is quite natural to consider the branching diffusion process, by introducing a diffusion movement for each particle in the branching system. Moreover, it is related to a class of semilinear parabolic PDEs, the KPP equations, as extension of the classical Feynmann-Kac formula. This was initially explored by Skrokohod [185], Watanabe [197], Mckean [151], etc., and developed by Dynkin [80], Le Gall [147], etc. It has also been studied as examples of measure valued Markov process, see e.g. Dawson [63], Roelly and Rouault [180], El Karoui and Roelly [88], etc. Combining with the stochastic control theory, the optimal control of branching diffusion processes has also been studied by Nisio [158], Claisse [51], etc.

It is quite classical to use the KPP equation to characterize the branching process system. Nevertheless, the idea to use branching process systems to represent the solution of PDEs and hence to obtain a Monte-Carlo method for PDEs seems to be new. To the best of our knowledge, Rasulov, Raimova and Mascagni [173] were the first to consider the Monte-Carlo method for KPP type equations by branching processes, although no rigorous convergence analysis was provided. Motivated by some high dimensional nonlinear equation with application in finance, Henry-Labordère [111] introduced a Monte-Carlo simulation method by the so-called "marked" branching process. Sufficient conditions for the convergence has been provided. In Henry-Labordère, Tan and Touzi [115], we tried to extend this algorithm for a class of path-dependent PDE (or equivalently nonMarkovian BSDEs), and provide a more rigorous convergence analysis. Similar ideas have also been used in Bossy, Champagnat, Leman, Maire, Violeau and Yvinec [31] to solve a class of nonlinear Poisson-Boltzmann equations using branching process systems. Notice that the above methods are all studied for the KPP type equations, where the nonlinearity of the PDE is a polynomial of value function u. To introduce the derivative Du into the nonlinearity, in Henry-Labordère, Oudjane, Tan, Touzi and Warin [113], we introduced a Malliavin type weight into the representation formula and obtained a representation result for a larger class of semilinear PDEs.

These ideas could also be used to deduce new simulation schemes for stochastic differential equations (SDEs). To estimate the expected value of a functional of the solution of SDEs by Monte Carlo method, a key step is to simulate the SDEs. For one-dimensional homogeneous SDEs with constant volatility function, Beskos and Roberts [23] introduced an exact simulation technique, based on a Girsanov measure change technique together with a rejection algorithm. However, for more general SDEs, an exact simulation method is still not available, and one needs to use a time discretization simulation method. Then the global error of the Monte Carlo method decomposes into two parts: the discretization error and the statistical error. The most elementary discrete scheme should be Euler scheme, where the discretization error has been initially analyzed by Talay and Tubaro [189]. Since then, many new discretization technique as well as error analysis result have been studied, for which we can refer to e.g. Kloeden and Platen [140], Graham and Talay [102], etc. Generally, to reduce the discretization error, one needs to use a finer time grid, which requires however more simulation effort. In other words, under a fixed computation effort, there will be less simulated samples of the SDEs if one uses a finer time grid, which induces more important statistical error. Recently, a new efficient simulation method, named multilevel Monte (MLMC) algorithm, has been introduced by Giles [101] (see also the statistical Romberg method of Kebaier [136]). The idea is to rewrite a high order discretized scheme as the sum of a low order discretized scheme and their differences. The low order term has big variance but can be simulated with little computation effort, one simulate many samples to reduce the statistical error; the differences terms need large computation effort but has a very small variance, one can simulate less samples to obtain a good estimation. With a good tradeoff on the levels and number of simulated samples, one can obtain an estimator with small discretization error as well as small statistical error.

As an extension of the MLMC method, Rhee and Glynn [179] considered a randomized level method and obtained an unbiased simulation method of SDEs, that is, the obtained estimator has the same expected value as the functional of solutions of the SDE. More recently, Bally and Kohatsu-Higa [5] provided a probabilistic interpretation for the PDE parametrix method. In particular, they obtained another unbiased simulation estimator for SDEs, where the estimator consists in simulating the SDE by a Euler scheme on a Poisson random discrete time grid, and then multiplying the final functional with a corrective weight function. By restricting our branching process analysis to the linear PDE case and using a freezing coefficient technique, we also obtained a unbiased simulation method for linear PDEs, or equivalently for the stochastic differential equations (SDEs). Our estimator is similar to that of Bally and Kohatsu-Higa [5], while the weight functions are quite different since they are obtained by different arguments. In particular, the estimator in [5] has an infinite variance, but we achieved to control the variance explosion in our estimator.

Coming back to the branching diffusion process algorithm, we notice that the considered semilinear PDEs (or PPDEs) should have a polynomial nonlinearity to have the representation result. For general semilinear PDEs, one needs to approximate the nonlinear part by a polynomial in order to use this algorithm. Nevertheless, to achieve a good approximation, a high order polynomials is required, but it may makes the branching process estimator explode. In Bouchard, Tan, Warin and Zou [40], we introduce a local polynomial approximation technique together with a Picard iteration to implement the branching process algorithm.

4.2 Branching diffusion, semilinear PDEs and Monte Carlo methods

Representation of a class of semi-linear PDEs by branching diffusions In Henry-Labordère, Oudjane, Tan, Touzi and Warin [113], we aim to provide a branching diffusion representation for the following semi-linear PDE:

$$\partial_t u + \mu \cdot Du + \frac{1}{2}\sigma\sigma^\top : D^2 u + f(\cdot, u, Du) = 0, \text{ on } [0,T) \times \mathbb{R}^d, \text{ and } u(T,.) = g, \quad (4.2.1)$$

where $g : \mathbb{R}^d \longrightarrow \mathbb{R}$ is bounded Lipschitz, f is given by

$$f(t,x,y,z) \;\; := \sum_{\ell=(\ell_0,\ell_1,\cdots,\ell_m)\in L} c_\ell(t,x)\; y^{\ell_0} \prod_{i=1}^m \big(b_i(t,x)\cdot z\big)^{\ell_i},$$

with some set $L \subset \mathbb{N}^{m+1}$, and functions $(c_\ell)_{\ell\in L}$ and $(b_i)_{i=1,\cdots,m}$. For every $\ell = (\ell_0, \ell_1, \cdots, \ell_m) \in L$, denote $|\ell| := \sum_{i=0}^m \ell_i$.

Let us first introduce an age-dependent branching diffusion process. Let $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ be a distribution density function, $(p_\ell)_{\ell\in L}$ be a probability mass function (i.e. $p_\ell \geq 0$ and $\sum_{\ell\in L} p_\ell = 1$). The age-dependent process starts from one particle indexed by (1), and marked by $\theta_{(1)} = 0$, performing a diffusion process movement $X^{(1)}_\cdot$ with generator $\mathcal{L} := \mu \cdot D + \frac{1}{2}\sigma\sigma^\top : D^2$ and initial condition $(0, x_0)$. Denote also by $W^{(1)}$ the associated Brownian motion in the definition of diffusion process $X^{(1)}$. Assume that the arrival time $T_{(1)}$ of (1) is of density function ρ. At the arrival time, the particle branches into $|\ell|$ offsprings with probability p_ℓ, indexed by $(1,1), \cdots (1, |\ell|)$. In other words, one has the random variable $I_{(1)}$ such that $\mathbb{P}[I_{(1)} = \ell] = p_\ell$. Among these offspring particles, ℓ_i particles carry the mark i, $i = 0, \ldots, m$. Then regardless of its mark, each descendant particle performs the same but independent branching process as the initial particle.

We denote by $\mathcal{K}_t$ (resp. $\mathcal{K}^n_t$) the set of all living particles (resp. of generation n) in the system at time t, and $\overline{\mathcal{K}}_t$ (resp. Kcb^n_t) the set of all particles which have been alive (resp. of generation n) before time t. Similarly to $\theta_{(1)}$, $T_{(1)}$ and $I_{(1)}$, we denote by θ_k the mark, by T_k the default time, and by $k-$ the parent for any particle $k \in \overline{\mathcal{K}}_T$. As a result, the birth time of k is T_{k-}. The position of particle k is denoted by $X^k_\cdot$ with the associated Brownian motion W^k. Denote $\Delta W^k_\cdot := W^k_{\cdot\wedge T_{k-}} - W^k_{T_{k-}}$.

Our key argument is an automatic differentiation technique on the underlying diffusion $\overline{X}^{t,x}_s$ defined by

$$\overline{X}^{t,x}_s \;=\; x + \int_t^s \mu\big(r, \overline{X}^{t,x}_r\big)dr + \int_t^s \sigma\big(r, \overline{X}^{t,x}_r\big)dW_r, \quad s \in [t,T],$$

where W is a d-dimensional Brownian motion.

Assumption 4.2.1. *There is a measurable functional $\overline{\mathcal{W}}(t,s,x,(W_r - W_t)_{r\in[t,s]})$ satisfying $(t,x) \mapsto \overline{\mathcal{W}}(t,s,x,(W_r - W_t)_{r\in[t,s]})$ is continuous, and for any $s \in [t,T]$ and bounded continuous function $\phi : \mathbb{R}^d \to \mathbb{R}$, one has*

$$\partial_x \mathbb{E}\big[\phi\big(\overline{X}^{t,x}_s\big)\big] \;=\; \mathbb{E}\Big[\phi\big(\overline{X}^{t,x}_s\big)\overline{\mathcal{W}}(t,s,x,(W_r - W_t)_{r\in[t,s]})\Big].$$

Remark 4.2.1. *In case $(\mu,\sigma) \equiv (\mu_0,\sigma_0)$ for some constant $(\mu_0,\sigma_0) \in \mathbb{R}^d \times \mathbb{M}^d$, where the matrix σ_0 is not degenerate, an example of such automatic differentiation function is given by*

$$\overline{\mathcal{W}}\big(t,s,x,(W_r - W_t)_{r\in[t,s]}\big) \;:=\; (\sigma_0^\top)^{-1}\frac{W_s - W_t}{s-t}.$$

For general coefficient functions (μ,σ) satisfying some regularity and non-degeneracy conditions, one can find such functional $\overline{\mathcal{W}}$ using Malliavin calculus.

With the above notations, we finally introduce

$$\psi \;:=\; \Big[\prod_{k\in\mathcal{K}_T}\frac{g(X^k_T) - g(X^k_{T_{k-}})\mathbf{1}_{\{\theta_k\neq 0\}}}{\overline{F}(\Delta T_k)}\mathcal{W}_k\Big]\Big[\prod_{k\in\overline{\mathcal{K}}_T\setminus\mathcal{K}_T}\frac{c_{I_k}(T_k,X^k_{T_k})}{p_{I_k}}\frac{\mathcal{W}_k}{\rho(\Delta T_k)}\Big], \qquad (4.2.2)$$

where $\overline{F}(t) := \int_t^\infty \rho(s)ds$, and

$$\mathcal{W}_k \;:=\; \mathbf{1}_{\{\theta_k=0\}} + \mathbf{1}_{\{\theta_k\neq 0\}}\; b_{\theta_k}(T_{k-}, X^k_{T_{k-}}) \cdot \overline{\mathcal{W}}\big(T_{k-}, T_k, X^k_{T_{k-}}, \Delta W^k_\cdot\big).$$

Theorem 4.2.1. (i) *Assume that the PDE (4.2.1) has a smooth solution u, then under technical condition, one has $u(0,x_0) = \mathbb{E}[\psi]$.*
(ii) *Replacing the initial condition $(0,x_0)$ of the above branching diffusion system by (t,x), and denote by $\psi_{t,x}$ the corresponding estimator. Under technical condition, the function $u(t,x) := \mathbb{E}[\psi_{t,x}]$ is well defined and is a viscosity solution of PDE (4.2.1).*

Proof. We will only provide a formal proof for item (i). For every $n \geq 1$, let us introduce

$$\psi_n \;:=\; \Big[\prod_{k\in\cup_{j=1}^n\mathcal{K}^j_T}\frac{g(X^k_T) - g(X^k_{T_{k-}})\mathbf{1}_{\{\theta_k\neq 0\}}}{\overline{F}(\Delta T_k)}\mathcal{W}_k\Big]\Big[\prod_{k\in\cup_{j=1}^n(\overline{\mathcal{K}}^j_T\setminus\mathcal{K}^j_T)}\frac{c_{I_k}(T_k,X^k_{T_k})}{p_{I_k}}\frac{\mathcal{W}_k}{\rho(\Delta T_k)}\Big]$$
$$\Big[\prod_{k\in\overline{\mathcal{K}}^{n+1}_T}\Big(\mathbf{1}_{\{\theta_k=0\}}u + \sum_{i=1}^m \mathbf{1}_{\{\theta_k=i\}}b_i\cdot Du\Big)(T_{k-},X^k_{T_{k-}})\Big]. \qquad (4.2.3)$$

First, given a solution u of the PDE (4.2.1), it follows from the Feynma-Kac formula that

$$\begin{aligned} u(0,x_0) &= \mathbb{E}\Big[\frac{1}{\overline{F}(T)} g\big(\overline{X}_T^{0,x_0}\big)\overline{F}(T) + \int_0^T \frac{1}{\rho(s)} f\big(\cdot,u,Du\big)\big(s,\overline{X}_s^{0,x_0}\big)\rho(s)ds\Big] \\ &= \mathbb{E}\Big[\frac{1}{\overline{F}(T_{(1)})} g\big(X_T^{(1)}\big)\mathbf{1}_{\{T_{(1)}=T\}} + \\ &\qquad \frac{1}{\rho(T_{(1)})}\Big(\frac{c_{I_{(1)}}}{p_{I_{(1)}}} u^{I_{(1),0}} \prod_{i=1}^m \big(b_i\cdot Du\big)^{I_{(1),i}}\Big)\big(T_{(1)},X_{T_{(1)}}^{(1)}\big)\mathbf{1}_{\{T_{(1)}<T\}}\Big] \qquad (4.2.4) \\ &= \mathbb{E}[\psi_1]. \end{aligned}$$

In the above, the last equality is in fact a simple reformulation of the expression for ψ_1.

Next, we use Assumptions 4.2.1 to obtain that

$$D_x u(0,x_0) = \mathbb{E}\Big[\psi_1 \overline{\mathcal{W}}\big(0,T_{(1)},x_0,\Delta W_\cdot^{(1)}\big)\Big]. \qquad (4.2.5)$$

For $k \in \overline{\mathcal{K}}_T^2$, change the initial condition from $(0,x_0)$ to $(T_{k-},X_{T_{k-}}^k) = (T_{(1)},X_{T_{(1)}}^{(1)})$ in formula (4.2.4) and (4.2.5), and plugging them into the definition of ψ_1 in (4.2.3), it follows that $u(0,x_0)=\mathbb{E}\big[\psi_2\big]$.

To conclude, it is enough to iterate the above procedure to prove that

$$u(0,x_0) = \mathbb{E}\big[\psi_n\big], \quad \text{for all } n \ge 1,$$

and by taking the limit, one obtains

$$u(0,x_0) = \lim_{n\to\infty} \mathbb{E}\big[\psi_n\big] = \mathbb{E}\big[\psi\big].$$

□

Unbiased simulation of SDEs In Henry-Labordère, Tan and Touzi [116], we restrict to the linear case and then obtain an unbiased simulation method for SDEs. Let us consider the SDE

$$X_0 = x_0, \qquad dX_t \;=\; \mu(t,X_t)\,dt \;+\; \sigma_0\,dW_t,$$

where σ_0 is a non-degenerate constant matrix. Our aim is to estimate

$$u(0,x_0) \;=\; \mathbb{E}[g(X_T)],$$

which, by Feynmann-Kac formula, is the solution of the linear PDE

$$\partial_t u + \frac{1}{2}\sigma_0\sigma_0^\top : D^2 u + \mu(\cdot)\cdot Du = 0, \quad u(T,\cdot)=g(\cdot).$$

Assumption 4.2.2. *The drift function $\mu(t,x)$ is bounded continuous in (t,x), uniformly $\frac{1}{2}$-Hölder in t and uniformly Lipschitz in x, i.e. for some constant $L>0$,*

$$\big|\mu(t,x)-\mu(s,y)\big| \;\le\; L\Big(\sqrt{|t-s|}+\big|x-y\big|\Big), \quad \forall (s,x),(t,y)\in[0,T]\times\mathbb{R}^d. \qquad (4.2.6)$$

Let $\beta > 0$ be a fixed positive constant, $(\tau_i)_{i>0}$ be a sequence of i.i.d. $\mathcal{E}(\beta)$-exponential random variables, which is independent of the Brownian motion W. We define

$$T_k \;:=\; \Big(\sum_{i=1}^{k} \tau_i\Big) \wedge T,\; k \geq 0, \quad \text{and} \quad N_t \;:=\; \max\big\{k : T_k < t\big\}.$$

Then $(N_t)_{0\leq t\leq T}$ is a Poisson process with intensity β and arrival times $(T_k)_{k>0}$, and $T_0 = 0$. For simplicity, denote

$$\Delta W_{T_k} \;:=\; \Delta W^k_{\Delta T_k} \;=\; W_{T_k} - W_{T_{k-1}}, \quad k > 0.$$

Define $\hat{X}$ by $\hat{X}_0 = x_0$ and

$$\hat{X}_{T_{k+1}} \;:=\; \hat{X}_{T_k} \;+\; \mu\big(T_k, \hat{X}_{T_k}\big)\Delta T_{k+1} \;+\; \sigma_0 \Delta W_{T_{k+1}}, \quad k = 0, 1, \cdots, N_T.$$

In the present case, the increment $\hat{X}_{T_{k+1}} - \hat{X}_{T_k}$, conditioning on $(T_k, \hat{X}_{T_k}, \Delta T_{k+1})$, is Gaussian. Then following Remark 4.2.1, we obtain the Malliavin weight function $\widehat{\mathcal{W}}^1_\theta\big(\cdot, \delta t, \delta w\big) := (\sigma_0^T)^{-1}\frac{\delta w}{\delta t}$. Our estimator is given by

$$\hat{\psi} \;:=\; e^{\beta T}\,\Big[g\Big(\hat{X}_T\Big) - g\Big(\hat{X}_{T_{N_T}}\Big)\mathbf{1}_{\{N_T>0\}}\Big]\;\beta^{-N_T}\prod_{k=1}^{N_T}\overline{\mathcal{W}}^1_k, \tag{4.2.7}$$

with

$$\overline{\mathcal{W}}^1_k \;:=\; \frac{\big(\mu(T_k, \hat{X}_{T_k}) - \mu(T_{k-1}, \hat{X}_{T_{k-1}})\big)\cdot(\sigma_0^T)^{-1}\Delta W_{T_{k+1}}}{\Delta T_{k+1}}. \tag{4.2.8}$$

Theorem 4.2.2. *Suppose that Assumption 4.2.2 holds true, and g is Lipschitz. Then for all intensity constant $\beta > 0$, one has*

$$\mathbb{E}\big[\big(\hat{\psi}\big)^2\big] \;<\; \infty; \quad \textit{and moreover,} \quad V_0 \;=\; u(0, x_0) \;=\; \mathbb{E}\big[\hat{\psi}\big].$$

Remark 4.2.2. (i) *The proof of Theorem 4.2.2 follows by the same arguments as in Theorem 4.2.1, together with a freezing coefficient argument.*

(ii) *Extensions have been made in [77] for more general SDEs, which is obtained essentially from an important sampling technique. see also Andersson and Kohatsu-Higa [2] as extension of Bally and Kohatsu-Higa's [5] estimator.*

Solving BSDE using polynomial generators approximation and branching processes Notice that the generator f of the semilinear PDE (4.2.1) has a polynomial structure, and the technical conditions to ensure the convergence of the Monte Carlo method requires essentially a small time horizon, or small coefficient functions in f. In practice, one would need to solve the semilinear PDE, or BSDEs with general Lipschitz generator f, which is wellposed for an arbitrary time horizon. To obtain a good approximation of f by polynomials f_n, one would like to use some high-order polynomial with arbitrary coefficients, but it may lead to variance explosion in practice. In Bouchard, Tan, Warin and Zou [40], we introduced an local polynomial approximation method together with a Picard iteration implementation technique to overcome this difficulty.

Let us consider the BSDE:

$$Y_\cdot = g(X_T) + \int_\cdot^T f(X_s, Y_s)\,ds - \int_\cdot^T Z_s\,dW_s,$$

with generator $f(x, y)$ Lipschitz in y and

$$X = X_0 + \int_0^\cdot \mu(X_s)\,dt + \int_0^\cdot \sigma(X_s)\,dW_s.$$

Equivalently, one has $Y_0 = u(0, X_0)$, where u is the unique viscosity solution of

$$\partial_t u + \mu \cdot Du + \frac{1}{2}\sigma\sigma^\top : D^2 u + f(\cdot, u) = 0, \qquad u(T, \cdot) = g(\cdot).$$

Our first step is to approximate f by some local polynomials $f_{\ell_\circ}$ (in y):

$$f_{\ell_\circ} : (x, y, y') \in \mathbb{R}^d \times \mathbb{R} \times \mathbb{R} \mapsto \sum_{j=1}^{j_\circ} \sum_{\ell=0}^{\ell_\circ} a_{j,\ell}(x) y^\ell \varphi_j(y'), \tag{4.2.9}$$

in which $(a_{j,\ell}, \varphi_j)_{\ell \le \ell_\circ, j \le j_\circ}$ is a family of bounded continuous functions satisfying

$$|a_{j,\ell}| \le C_{\ell_\circ}\,,\ |\varphi_j(y_1') - \varphi_j(y_2')| \le L_\varphi |y_1' - y_2'| \text{ and } |\varphi_j| \le 1, \tag{4.2.10}$$

for all $y_1', y_2' \in \mathbb{R}$, $j \le j_\circ$ and $\ell \le \ell_\circ$, for some constants $C_{\ell_\circ}, L_\varphi \ge 0$. Without loss of generality, we assume that $\ell_\circ \ge 2$. We then consider the BSDE with generator $f_{\ell_\circ}$:

$$\bar{Y}_\cdot = g(X_T) + \int_\cdot^T f_{\ell_\circ}(X_s, \bar{Y}_s, \bar{Y}_s)\,ds - \int_\cdot^T \bar{Z}_s\,dW_s.$$

Whenever $f_{\ell_\circ}(x, y, y)$ is a good approximation of $f(x, y)$, it is very classical to see that $\bar{Y}$ is a good approximation of Y.

To approximate $\bar{Y}$, we suggest a Picard iteration: Let us first fix $h > 0$ small enough and such that $N_h := T/h \in \mathbb{N}$, and define $t_i = ih$. Then set $\bar{Y}^0 := \mathrm{y}(t, X_t)$ for some deterministic function y, and then define $\bar{Y}^m$ given $\bar{Y}^{m-1}$ as follows:

- Let $\bar{Y}_T^m := g(X_T)$ be the terminal condition.
- On each interval $[t_i, t_{i+1}]$, define $(Y_\cdot^m, Z_\cdot^m)$ as the solution on $[t_i, t_{i+1}]$ of

$$Y_\cdot^m \ = \ \bar{Y}_{t_{i+1}}^m + \int_\cdot^{t_{i+1}} f_{\ell_\circ}(X_s, Y_s^m, \bar{Y}_s^{m-1}) ds - \int_\cdot^{t_{i+1}} Z_s^m dW_s. \tag{4.2.11}$$

- Let $\bar{Y}^m := Y^m$ on $(t_i, t_{i+1}]$, and $\bar{Y}_{t_i}^m := (-M) \vee Y_{t_i}^m \wedge M$, where M is some a priori constant dominating Y.

Theorem 4.2.3. *Under technical conditions, $\bar{Y}^m$ converges to $\bar{Y}$ with a convergence rate $C_\varepsilon \varepsilon^m$ with some constant C_ε for any $\varepsilon > 0$.*

Remark 4.2.3. (i) *Notice that given y', the function $y \mapsto f_{\ell_\circ}(x, y, y')$ is a polynomial, and hence Y^m in (4.2.11) can be computed by its branching diffusion representation formula. A good choice $h > 0$ as time discretization parameter could ensure the convergence of branching Monte Carlo method.*

(ii) *For implementation of the algorithm, thanks to the time discretization, we do not need really to repeat the Picard iteration, only once is enough.*

4.3 Perspectives

The link between branching diffusion process and the nonlinear PDE provides not only new Monte-Carlo methods for nonlinear equations, but also has its theoretical interest. As for the application in Monte-Carlo approximation of nonlinear PDEs, it is still limited to the small time horizon or/and small nonlinearity coefficients case. New efforts are still needed to obtain more stable and efficient algorithm. Moreover, our actual result is only presented for a class of semilinear PDEs. Using the same Mallaivin type weight for the second order derivative of the value function, it could be immediately extended to the fully nonlinear case, where the nonlinearity depends also on D^2u. Nevertheless, it is observed that such an estimator in the simplest case will not be integrable, and one can not even define the expectation. In an on-going project with Pierre Henry-Labordère, Nizar Touzi and Xavier Warin, we use an antithetic variable kind method to obtain a new estimator. Moreover, numerical tests show that it could work in a fully nonlinear case. We are still woking on it to understand how to control the variance theoretically.

The branching process has been largely used in biological mathematics to model the dynamic of the population. Nevertheless, there has been very little work using branching processes in economical or financial modelling. In a current project with Julien Claisse and Zhenjie Ren, we are interested in a mean-field game problem where the population could be dynamic because of the immigration, emigration and branching behaviour of the individuals. We use the branching process to model the dynamics of the population in the game.

Appendix A

Appendix

A.1 Convergence of measures and the topology

In this section, we provide a summary of different notions related to the convergence of measures as well as the associated topology, which have been used in different places in this HDR thesis. Most of the the results can be found in Billinsley [25, 26], Jacod and Shiryaev [128], etc. In this section, we assume that $(\Omega, \mathcal{F})$ is an abstract measurable space, E is a topological space with the associated Borel σ-field $\mathcal{E}$, so that $(E, \mathcal{E})$ is a measurable space.

Weak convergence

Let $\mathfrak{B}(E)$ denote the collection of all probability measures on $(E, \mathcal{E})$ and $(\mu_n)_{n\geq 1}$ be a sequence in $\mathfrak{B}(E)$, we introduce some definitions related to the weak convergence.

Definition A.1.1. (i). *We say that $(\mu_n)_{n\geq 1}$ converges weakly to a probability measure μ_∞ if $\mu_n(f) \to \mu_\infty(f)$ for all $f \in C_b(E)$, where $C_b(E)$ denotes the collection of all bounded continuous functions defined on E, and $\mu(f)$ denotes the integration of function f w.r.t. measure μ.*
(ii). *We say that the set $(\mu_n)_{n\geq 1}$ is tight if for any $\varepsilon > 0$, there is a compact set K_ε such that $\mu_n(K_\varepsilon) \geq 1 - \varepsilon$ for all $n \geq 1$.*
(iii). *We say that the set $(\mu_n)_{n\geq 1}$ is relatively compact if for any subsequence $(\mu_{n_k})_{k\geq 1}$, there is a subsubsequence $(\mu_{n_{k_i}})_{i\geq 1}$ which converges weakly to some probability measure.*

In many context, it would be nice to see a probability measure $\mu \in \mathfrak{B}(E)$ as an element in a vector space and the natural candidate of such a vector space would be the space $\mathfrak{M}(E)$ of all finite signed measures on $(E, \mathcal{E})$. The above notion of weak convergence can be defined on $\mathfrak{M}(e)$ by exactly the same way. One has the following results on the tightness and the relative compactness of the measures on E.

Proposition A.1.1. (i). *If E is a Hausdoff topological space, then $(\mu_n)_{n\geq 1}$ is tight, then it is relatively compact.*
(ii). *If E is a Polish space, then the tightness of $(\mu_n)_{n\geq 1}$ is equivalent to its relative compactness.*
(iii). *If E is a Polish space, then the weak convergence induces a metrizable topology on $\mathfrak{B}(E)$ as well as on $\mathfrak{M}(E)$ such that they are both Polish spaces. Moreover, the dual space of $\mathfrak{M}(E)$ can be identify to be $C_b(E)$, i.e. any linear continuous form $\phi : \mathfrak{M}(E) \to \mathbb{R}$ is in form $\phi(\mu) = \mu(f)$ for some $f \in C_b(E)$.*

Wasserstein convergence

Although the Wasserstein convergence topology can be defined on general Polish space, we restrict here to the case $E = \mathbb{R}^d$. Let $p \geq 1$, we denote by $\mathfrak{B}_p(\mathbb{R}^d)$ the space of all probability measures μ on $\mathbb{R}^d$ such that $\int_{\mathbb{R}^d} |x|^p \mu(dx) < \infty$, and by $\mathcal{C}_p(\mathbb{R}^d)$ the space of all continuous functions $\phi : \mathbb{R}^d \to \mathbb{R}$ such that $\sup_{x \in \mathbb{R}^d} \frac{|\phi(x)|}{1+|x|^p} < \infty$. Similarly, we denote by $\mathfrak{M}_p(\mathbb{R}^d)$ the collection of all finite signed measures μ on $\mathbb{R}^d$ such that $\int_{\mathbb{R}^d} |x|^p |\mu|(dx) < \infty$

Definition A.1.2. *We say that $(\mu_n)_{n\geq 1} \subset \mathfrak{M}_p(\mathbb{R}^d)$ converges to $\mu_\infty \in \mathfrak{M}_p(\mathbb{R}^d)$ in $p-$Wasserstein topology if $\mu_n(f) \to \mu_\infty(f)$ for all $f \in \mathcal{C}_p(\mathbb{R}^d)$.*

Proposition A.1.2. (i). *The space $\mathfrak{B}_p(\mathbb{R}^d)$ equipped with $p-$Wasserstein topology is a Polish space.*
(ii). *The dual space of the vector space $\mathfrak{M}_p(\mathbb{R}^d)$ can be identified to $\mathcal{C}_p(\mathbb{R}^d)$.*

Item (ii) in Proposition A.1.2 should compare to item (iii) in Proposition A.1.1.

Stable convergence

In some context, one would like to keep an abstract measurable (or probability) space $(\Omega, \mathcal{F})$ and consider at the same the weak convergence of the E-valued random variables (processes) in this measurable (or probability) space. A powerful tool in this context should be the stable convergence topology on the space of measures on the product space $(\overline{\Omega}, \overline{\mathcal{F}}) := (\Omega \times E, \mathcal{F} \otimes \mathcal{E})$ introduced by Jacod and Mémin [129]. Let us denote by $BC_b(\overline{\Omega})$ the collection of all bounded measurable functions $\xi : \overline{\Omega} \to \mathbb{R}$ such that $e \mapsto \xi(\omega, e)$ is continuous for every fixed $\omega \in \Omega$.

Definition A.1.3. *Let $(\overline{\mathbb{P}}_n)_{n\geq 1}$ be a sequence of probability measures defined on $(\overline{\Omega}, \overline{\mathcal{F}})$. We say that $(\overline{\mathbb{P}}_n)_{n\geq 1}$ converges to $\overline{\mathbb{P}}_\infty$ under the stable convergence topology if $\mathbb{E}^{\overline{\mathbb{P}}_n}[\xi] \to \mathbb{E}^{\overline{\mathbb{P}}_\infty}[\xi]$ for every $\xi \in BC_b(\overline{\Omega})$.*

Let $\mathbb{P}$ be a fixed probability measure on $(\Omega, \mathcal{F})$, denote by $\mathfrak{B}(\overline{\Omega}, \mathbb{P})$ the collection of all probability measures $\overline{\mathbb{P}}$ on $(\overline{\Omega}, \overline{\mathcal{F}})$ such that $\overline{\mathbb{P}}|_\Omega = \mathbb{P}$.

Proposition A.1.3. (i). *A sequence $(\overline{\mathbb{P}}_n)_{n\geq 1}$ converges to $\overline{\mathbb{P}}_\infty$ under the stable convergence topology if and only if $\mathbb{E}^{\overline{\mathbb{P}}_n}[\xi] \to \mathbb{E}^{\overline{\mathbb{P}}_\infty}[\xi]$ for every bounded measurable variable $\xi : \Omega \to R$ and every bounded continuous variable $\xi : E \to \mathbb{R}$.*
(ii). *For every probability measure $\mathbb{P} \in \mathfrak{B}(\Omega)$, the set $\mathfrak{B}(\overline{\Omega}, \mathbb{P})$ is a closed subspace of $\mathfrak{B}(\overline{\Omega})$ under the stable convergence topology. If, in addition, Ω is a Polish space with its Borel $\sigma-$field, then restricted on $\mathfrak{B}(\overline{\Omega}, \mathbb{P})$, the stable convergence is equivalent to the weak convergence topology.*

Item (ii) of Proposition A.1.3 could be very useful in the case where one considers the joint distributions on a product space with a fixed marginal distribution.

Topologies on the Skorokhod space

In many cases, we are interested in the weak convergence of the stochastic processes, where we usually identify E as the Skorokhod space, i.e. the space of càdlàg paths $E = \mathbb{D}([0,1],\mathbb{R}^d)$. The above notions of convergence of measures depend on the continuous functions on E, which depends essentially on the topology equipped on the space E. We recall here 4 different topologies defined on E below, which are sequential topologies induced by different notions of convergence: let $(\mathbf{x}_n)_{n\geq 1}$ be a sequence of elements in E,

- (The uniform convergence topology) we say $\mathbf{x}_n \to \mathbf{x}_\infty$ under the uniform convergence topology if $\|\mathbf{x}_n - \mathbf{x}_\infty\| := \sup_{0\leq t\leq 1} \|\mathbf{x}_n(t) - \mathbf{x}_\infty(t)\| \to 0$ as $n \to \infty$.

- (The Skorokhod topology) we say $\mathbf{x}_n \to \mathbf{x}_\infty$ under the Skorokhod topology if

$$\inf_{\lambda\in\Lambda} \max\{\|\lambda - I\|, \|\mathbf{x}_n - \mathbf{x}_\infty \circ \lambda\|\} \to 0, \tag{A.1.1}$$

  where I is the identical function on $[0,1]$ and Λ is the collection of all strictly increasing, continuous bijection from $[0,1]$ to $[0,1]$.

- (The S-topology) we say $\mathbf{x}_n \to^S \mathbf{x}_\infty$ if for any $\varepsilon > 0$, there exists a sequence $(\mathbf{v}_n^\varepsilon)_{n\geq 1}$ of functions of finite variation, and such that $\sup_{0\leq t\leq 1} \|\mathbf{x}_n(t) - \mathbf{v}_n^\varepsilon(t)\| \leq \varepsilon$ and $\int_0^1 f(t)\cdot d\mathbf{v}_n^\varepsilon(t) \to \int_0^1 f(t)\cdot d\mathbf{v}_\infty^\varepsilon(t)$ for all continuous functions $f : [0,1] \to \mathbb{R}^d$.

- (The pseudo-path topology of Meyer and Zheng [154]) we say $\mathbf{x}_n \to \mathbf{x}_\infty$ under the pseudo-path topology if $\int_0^1 f(\mathbf{x}_n(t))dt \to \int_0^1 f(\mathbf{x}_\infty(t))dt$ for all bounded continuous function $f : [0,1]\times\mathbb{R}^d$.

Remark A.1.1. (i). *The uniform convergence topology, Skorokhod topology and the pseudo-path topology are all metrizable topologies. In particular, the convergence in the three topologies is the equivalent to the above notions of convergence which induce the three topologies. Moreover, E is metrisable under the uniform convergence topology and the Skorokhod topology, and is a Borel subset in a Polish space under the pseudo-path topology. However, the S-topology, defined as the sequential topology induced by the convergence $\to^S$, is not metrizable. Moreover, the $\to^S$ induces the convergence under S-topology, but the reverse induction may not be true.*

(ii). *The uniform convergence topology is finer that the Skorokhod topology, which is finer than S-topology, and the pseudo-path topology is the sparsest one. The sparser the topology is, the less the open sets there are, and the more the compact sets there are, and hence the easier the tightness of the measures on it can be obtained. However, the sparser the topology is, the less the continuous functions there are. In practice, there is a tradeoff to make in order to choose a good topology to have the tightness (and hence the sequential compactness) and at the same time the most continuous functions possible.*

The Skorokhod topology is clearly the most important one on the Skorokhod space in many situation. In this HDR thesis, the S-topology has also been essentially used. Let us just recall some sufficient conditions to ensure the tightness of the distributions on E under the two different topologies.

Proposition A.1.4. (i). *All the four topologies induce the same Borel $\sigma-$field $\mathcal{E}$ on E.*
(ii). *A sequence of probability measures $(\mu_n)_{n\geq 1}$ on $(E, \mathcal{E})$ is tight w.r.t. the Skorokhod topology if and only if*

$$\lim_{C\to\infty} \limsup_{n\to\infty} \mu_n \big[\{ \mathbf{x} \in E \ : \|\mathbf{x}\| \geq C \} \big] \ = \ 0,$$

and

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mu_n \big[\{ \mathbf{x} \in E \ : \bar{\omega}'_{\mathbf{x}}(\delta) \geq \varepsilon \} \big] = 0, \quad \text{for all } \varepsilon > 0, \tag{A.1.2}$$

where $\bar{\omega}'_{\mathbf{x}}(\delta) := \inf_{\Pi_\delta} \max_{1\leq i\leq k} \sup_{s,t\in(t_{i-1},t_i)} |\mathbf{x}(s) - \mathbf{x}(t)|$ and Π_δ is the collection of all partition $\pi = \{0 = t_0 < t_1 < \cdots < t_k = 1\}$ such that all $t_i - t_{i-1} \geq \delta$.
(iii). *Let $(\mu_n)_{n\geq 1}$ be a sequence of probability measures on $(E, \mathcal{E})$, assume that the canonical process on canonical space E is a supermartingale under each μ_n and $\sup_{n\geq 1} \sup_{0\leq t\leq 1} \int_{\mathbb{R}^d} |\mathbf{x}(t)| \mu(d\mathbf{x}) < \infty$. Then $(\mu_n)_{n\geq 1}$ is tight w.r.t. the S-topology.*

Remark A.1.2. *In condition (A.1.2), $\bar{\omega}'_{\mathbf{x}}(\delta)$ could be small even if the path $\mathbf{x}$ has a big jump, since one can choose the partition in a way such that the jump times do not lie in any interval (t_i, t_{i+1}). However, because of the condition $t_i - t_{i-1} \geq \delta$, one could not have two consequent big jumps to make $\bar{\omega}'_{\mathbf{x}}(\delta)$ small. The intuition in condition (A.1.2) to ensure the tightness is that the process should not have two big consequent jumps, nor large volatility with big probability.*

To conclude, we also consider a subspace of the Skorokhod space, which is the space $A([0,1], \mathbb{R})$ of all non-decreasing càdlàg paths. An important topology on $A([0,1], \mathbb{R})$ is that induced by the Lévy metric.

$$d(\mathbf{x}, \mathbf{x}') := \inf\{\varepsilon > 0 \ : \mathbf{x}(t-\varepsilon) - \varepsilon \leq \mathbf{x}'(t) \leq \mathbf{x}(t+\varepsilon) + \varepsilon \text{ for all } t \in [0,1]\}.$$

Any non-decreasing càdlàg path on $[0,1]$ can be identified as a finite positive measure on $[0,1]$, then the topology induced by Lévy metric is equivalent to the corresponding weak convergence topology.

A.2 List of publications

Accepted papers:

1. X. Tan and N. Touzi, Optimal Transportation under Controlled Stochastic Dynamics, *Annals of Probability*, Vol. 41, No. 5, 3201-3240, 2013.

2. J.F. Bonnans and X. Tan, A model-free no-arbitrage price bound for variance options, *Applied Mathematics & Optimization*, Vol. 68, Issue 1, 43-73, 2013.

3. X.Tan, A splitting method for fully nonlinear degenerate parabolic PDEs, *Electron. J. Probab*, 18(15):1-24, 2013.

4. X. Tan, Discrete-time probabilistic approximation of path-dependent stochastic control problems, *Annals of Applied Probability*, 24(5):1803-1834, 2014.

5. P. Henry-Labordère, X. Tan and N. Touzi, A numerical algorithm for a class of BSDEs via branching process, *Stochastic Processes and their Applications*, 124(2):1112-1140, 2014.

6. D. Possamaï and X. Tan, Weak approximation of second order BSDEs, *Annals of Applied Probability*, 25(5):2535-2562, 2015.

7. J. Claisse, D. Talay and X. Tan, A pseudo-Markov property for controlled diffusion processes, *SIAM Journal on Control and Optimization*, 54(2):1017-1029, 2016.

8. P. Henry-Labordère, X. Tan and N. Touzi, An Explicit Martingale Version of the One-dimensional Brenier's Theorem with Full Marginals Constraint, *Stochastic Processes and their Applications*, 126(9):2800-2834, 2016.

9. G. Guo, X. Tan and N. Touzi, Optimal Skorokhod embedding under finitely-many marginal constraints, *SIAM Journal on Control and Optimization*, 54(4):2174-2201, 2016.

10. G. Guo, X. Tan and N. Touzi, On the monotonicity principle of optimal Skorokhod embedding problem., *SIAM Journal on Control and Optimization*, 54(5):2478-2489, 2016.

11. G. Guo, X. Tan and N. Touzi, Tightness and duality of martingale transport on the Skorokhod space, *Stochastic Processes and their Applications*, 127(3):927-956, 2017.

12. S. Källblad, X. Tan and N. Touzi, Optimal Skorokhod embedding given full marginals and Azema-Yor peacocks, *Annals of Applied Probability*, 27(2):686-719, 2017.

13. B. Bouchard, D. Possamaï and X. Tan, A general Doob-Meyer-Mertens decomposition for g-supermartingale systems, *Electronic Journal of Probability*, 21(36):1-21, 2016.

14. B. Bouchard, D. Possamaï, X. Tan and C. Zhou, A unified approach to a priori estimates for supersolutions of BSDEs in general filtrations, *Annales de l'Institut Henri Poincaré*, Probabilités et Statistiques, to appear.

15. Z. Ren and X. Tan, On the convergence of monotone schemes for path-dependent PDE, *Stochastic Processes and their Applications*, 127(6):1738-1762, 2017.

16. P. Henry-Labordère, X. Tan and N. Touzi, Unbiased simulation of stochastic differential equations, *Annals of Applied Probability*, to appear.

17. D. Possamaï, X. Tan and C. Zhou, Stochastic control for a class of nonlinear kernels and applications, *Annals of Probability*, to appear.

18. B. Bouchard, X. Tan, Y. Zou and X. Warin, Numerical approximation of BSDEs using local polynomial drivers and branching processes. *Monte Carlo Methods and Applications*, to appear.

Preprint

19. P. Henry-Labordère, N. Oudjane, X. Tan, N. Touzi and X. Warin, Branching diffusion representation of semilinear PDEs and Monte Carlo approximation.

20. A. Aksamit, S. Deng, J. Obłój and X. Tan, Robust pricing-hedging duality for American options in discrete time financial markets.

21. B. Bouchard, S. Deng and X. Tan, Super-replication with proportional transaction cost under model uncertainty.

22. B. Bouchard, X. Tan and X. Warin, Numerical approximation of general Lipschitz BSDEs with branching processes.

23. N. El Karoui and X. Tan, Capacities, measurable selection and dynamic programming Part I: abstract framework.

24. N. El Karoui and X. Tan Capacities, Measurable Selection and Dynamic Programming Part II: Application in Stochastic Control Problems.

A.3 CV

Employment

- **2013 - pres.**, Université Paris-Dauphine, PSL Research University, Assistant Professor.
- **2012 - 2013**, Université Paris-Dauphine, PSL Research University, ATER (temporary teaching/reseach position).
- **2010 - 2012**, Ecole Polytechnique, Teaching Assistant.
- **2009 - 2011**, AXA, Group Risk Management, Industrial consulting (1 day per week).

Education

- **Dec. 2017**, Université Paris-Dauphine, PSL Research University, **HDR thesis**.
- **2009 - 2011**, Ecole Polytechnique, CMAP, **PhD thesis** under supervision of Nizar Touzi and Frédéric Bonnnans.
- **2008 - 2009**, UPMC - X, Master 2 "Probabilités et Finance", **Master degree**.
- **2005 - 2009**, Ecole Polytechnique, **Engineer degree**.
- **2001 - 2005**, Peking University, School of Mathematical Science, **Bachelor degree**.

Teaching Activities

- Lectures at Univ. Paris Dauphine:
  - 2017 - present: **Derivatives pricing and no-arbitrage** (Masef, M2)
  - 2016 - present: **Optimisation and dynamic programming** (MIDO, M1)
  - 2014 - 2017: **Jump process** (Masef, M2)
  - 2013 - present: **Stochastic process and PDE** (ISF, M2)
  - 2012 - 2017: **Monte Carlo** (MIDO, M1)
- Small classes
  - At Ecole Polytechnique: Stochastic optimal control, C++, Scilab, etc.
  - At Paris-Dauphine: Analyse 3, Monte Carlo, Gestion de Risque, Mathématique de Risque, Processus Discret, Processus Poisson, Series Temporelles, Mouvement brownien et évaluation des actifs, etc.

Research Activities

- Supervision of master memories
  - Adel Cherchali (2015), Shuoqing Deng (2015), Hao Ngoc DANG (2017), Fabrice Djete (2017)
- Supervision of PhD thesis
  - Gaoyue Guo (2013 - 2016), co-supervised with Nizar Touzi.
  - Shuoqing Deng (2015 -), co-supervised with Bruno Bouchard
  - Fabrice Djete (2017 -), co-supervised with Dylan Possamaï
- Long stay visiting:
  - 2014/08, NUS, Singapore.
  - 2015/04, ENIT, Tunis.
  - 2016/04, Oxford University.

Bibliographie

[1] A. Aksamit, S. Deng, J. Obłój and X. Tan *Robust pricing-hedging duality for American options in discrete time financial markets*, preprint, 2017. (Cited on pages 15 and 22.)

[2] P. Andersson, and A. Kohatsu-Higa. *Unbiased simulation of stochastic differential equations using parametrix expansions.* Bernoulli 23.3: 2028-2057, 2017. (Cited on page 58.)

[3] K. B. Athreya and P. E. Ney, *Branching processes*, Springer-Verlag, New York, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196. (Cited on page 53.)

[4] J. Azéma and M. Yor, *Une solution simple au problème de Skorokhod*, in Séminaire de Probabilités, XIII, Vol 721 of Lecture Notes in Math., pages 90-115. Springer, Berlin, 1979. (Cited on page 15.)

[5] Bally, V., Kohatsu-Higa, A. : *A probabilistic interpretation of the parametrix method*, The Annals of Applied Probability 25.6: 3095-3138, 2015. (Cited on pages 54 and 58.)

[6] V. Bally and P. Pages, *Error analysis of the quantization algorithm for obstacle problems*, Stochastic Processes & Their Applications, 106(1):1-40, 2003. (Cited on page 45.)

[7] G. Barles, C. Daher, and M. Romano, *Convergence of numerical schemes for parabolic equations arising in finance theory*, Math. Models Methods Appl. Sci., 5(1):125-143, 1995. (Cited on page 33.)

[8] G. Barles and E.R. Jakobsen, *Error bounds for monotone approximation schemes for parabolic Hamilton-Jacobi-Bellman equations*, Math. Comp., 76(260):1861-1893(electronic), 2007. (Cited on pages 33, 45 and 46.)

[9] G. Barles and P.E. Souganidis, *Convergence of approximation schemes for fully nonlinear second order equations*, Asymptotic Anal., 4(3):271-283, 1991. (Cited on pages 33, 44, 45, 46, 47, 49 and 50.)

[10] D. Bartl. *Exponential utility maximization under model uncertainty for unbounded endowments.* arXiv preprint arXiv:1610.00999, 2016. (Cited on pages 15 and 30.)

[11] E. Bayraktar, Y. Huang, and Z. Zhou, *On hedging American options under model uncertainty.* SIAM Journal on Financial Mathematics, 6(1):425-447, 2015. (Cited on pages 22 and 24.)

[12] E. Bayraktar and Y. Zhang. *Fundamental theorem of asset pricing under transaction costs and model uncertainty.* Mathematics of Operations Research, 41 (3), 1039-1054, 2016. (Cited on pages 15 and 24.)

[13] D. Becherer and K. Kentia. *Good Deal Hedging and Valuation under Combined Uncertainty about Drift and Volatility*, 2017. (Cited on page 33.)

[14] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001. (Cited on page 29.)

[15] A. Bensoussan and J-L. Lions. *Applications of variational inequalities in stochastic control.* Vol. 12. Elsevier, 2011. (Cited on page 33.)

[16] M. Beiglböck, A. Cox and M. Huesmann : *Optimal Transport and Skorokhod Embedding*, Invent. Math., to appear. (Cited on pages 16, 20, 21, 25, 27 and 28.)

[17] M. Beiglböck, A. Cox and M. Huesmann : *The geometry of multi-marginal Skorokhod Embedding*, preprint, 2017. (Cited on page 25.)

[18] M. Beiglböck, P. Henry-Labordère and F. Penkner, *Model-independent bounds for option prices: a mass transport approach*. (Cited on pages 1, 5, 11 and 14.)

[19] M. Beiglböck, P. Henry-Labordère and N. Touzi, *Monotone Martingale Transport Plans and Skorohod Embedding*, Stochastic Processes and their Applications, to appear. (Cited on page 14.)

[20] M. Beiglböck and N. Juillet, *On a problem of optimal transport under marginal martingale constraints*, Ann. Probab., to appear. (Cited on pages 14, 16, 25 and 26.)

[21] M. Beiglböck, M. Nutz and N. Touzi: *Robust super-replication for time-invariant derivatives*, Annals of Probability, to appear. (Cited on page 14.)

[22] D.P. Bertsekas and S.E. Shreve, *Stochastic optimal control, the discrete time case*, volume 139 of *Mathematics in Science and Engineering*, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1978. (Cited on page 36.)

[23] Beskos, A., Roberts, G. O. : *Exact simulation of diffusions*, Ann. Appl. Probab. Volume 15, Number 4, pp. 2422-2444, 2005. (Cited on page 54.)

[24] S. Biagini, B. Bouchard, C. Kardaras and M. Nutz : *Robust fundamental theorem for continuous processes*, Mathematical Finance, to appear. (Cited on pages 18 and 30.)

[25] P. Billingsley, *Convergence of probability measures*, John Wiley & Sons Inc., New York, 1968. (Cited on page 61.)

[26] P. Billingsley, *Weak convergence of measures: Applications in probability*, Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1971. (Cited on page 61.)

[27] J.F. Bonnans, E. Ottenwaelter, and H. Zidani, *A fast algorithm for the two dimensional HJB equation of stochastic control*, M2AN Math. Model. Numer. Anal., 38(4):723–735, 2004. (Cited on page 46.)

[28] J.F. Bonnans and A. Shapiro, *Perturbation analysis of optimization problems*, Springer Series in Operations Research, Springer-Verlag, New York, 2000. (Cited on page 17.)

[29] J. F. Bonnans and X. Tan, *A model-free no-arbitrage price bound for variance options*, Applied Mathematics & Optimization, 68(1):43-73, 2013. (Cited on page 29.)

[30] J.F. Bonnans and H. Zidani, *Consistency of generalized finite difference schemes for the stochastic HJB equation*, SIAM J. Numer. Anal., 41(3):1008-1021(electronic), 2003. (Cited on page 46.)

[31] M. Bossy, N. Champagnat, H. Leman, S. Maire, L. Violeau and M. Yvinec, *Monte Carlo methods for linear and non-linear Poisson-Boltzmann equation*, ESAIM:Proceedings and Surveys, 48:420-446, 2015. (Cited on page 53.)

[32] B. Bouchard, *Introduction to stochastic control of mixed diffusion processes, viscosity solutions and applications in finance and insurance*, Lecture note, 2007. (Cited on page 33.)

[33] B. Bouchard, S. Deng and X. Tan *Super-replication with proportional transaction cost under model uncertainty*, preprint, 2017. (Cited on pages 15, 24 and 30.)

[34] Bouchard, B., Élie, R. and A., Réveillac. BSDEs with weak terminal condition, *The Annals of Probability*, 43(2), 2015, 572–604. (Cited on page 42.)

[35] B. Bouchard and M. Nutz. Arbitrage and duality in nondominated discrete-time models. *Ann. Appl. Probab.*, 25(2):823–859, 2015. (Cited on pages 14, 22, 24 and 30.)

[36] B. Bouchard and M. Nutz. *Consistent price systems under model uncertainty.* Finance and Stochastics, 20(1):83–98, 2016. (Cited on pages 15 and 24.)

[37] B. Bouchard and N. Touzi, *Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations*, Stochastic Process. Appl., 111(2):175-206, 2004. (Cited on page 45.)

[38] B. Bouchard, D. Possamaï, and X. Tan. A general Doob–Meyer–Mertens decomposition for $g-$supermartingale systems. *Electronic Journal of Probability*, 21(36):1–21, 2016. (Cited on page 43.)

[39] B. Bouchard, D. Possamaï, X. Tan, and C. Zhou. A unified approach to *a priori* estimates for supersolutions of BSDEs in general filtrations. *Annales de l'institut Henri Poincaré, Probabilités et Statistiques* (B), to appear, 2015. (Cited on page 43.)

[40] B. Bouchard, X. Tan, X. Warin, and Y. Zhou, *Numerical approximation of BSDEs using local polynomial drivers and branching processes*, preprint, 2017. (Cited on pages 55 and 58.)

[41] B. Bouchard and N. Touzi. Weak dynamic programming principle for viscosity solutions. *SIAM Journal on Control and Optimization*, 49(3):948–962, 2011. (Cited on page 35.)

[42] B. Bouchard and X. Warin, *Monte-Carlo valorisation of American options: facts and new algorithms to improve existing methods*, preprint, 2011. (Cited on page 45.)

[43] D.T. Breeden and R.H. Litzenberger : *Prices of state-contingent claims implicit in options prices*, J. Business, 51, 1978. (Cited on page 12.)

[44] Briand, P., Delyon, B., Mémin, J. (2002). On the robustness of backward stochastic differential equations, *Stochastic Process. Appl.*, 97:229–253. (Cited on page 48.)

[45] H. Brown, D. Hobson and L.C.G. Roggers : *Robust hedging of barrier options*, Mathematical Finance. 11, pp285-314, 2001. (Cited on page 16.)

[46] G. Brunick, and S. Shreve, *Mimicking an Itô process by a solution of a stochastic differential equation.* The Annals of Applied Probability, 23(4):1584-1628, 2013. (Cited on page 12.)

[47] M. Burzoni. Arbitrage and hedging in model-independent markets with frictions. *SIAM Journal on Financial Mathematics*, 7(1):812–844, 2016. (Cited on pages 24 and 25.)

[48] M. Burzoni, M. Frittelli, Z. Hou, M. Maggis, J. Obłój, *Pointwise arbitrage pricing theory in discrete-time*, preprint, 2016. (Cited on page 22.)

[49] R. Carmona, F. Delarue and D. Lacker, *Mean field games with common noise*, The Annals of Probability 44(6): 3740-3803, 2016. (Cited on page 39.)

[50] P. Cheridito, H.M. Soner, N. Touzi, and N. Victoir, *Second-order backward stochastic differential equations and fully nonlinear parabolic PDEs*, Comm. Pure Appl. Math., 60(7):1081-1110, 2007. (Cited on page 33.)

[51] J. Claisse, *Optimal control of branching diffusion processes: a finite horizon problem*, Annals of Applied Probability, to appear. (Cited on page 53.)

[52] J. Claisse, D. Talay, and X. Tan. *A pseudo–Markov property for controlled diffusion processes.* SIAM Journal on Control and Optimization, 54(2):1017–1029, 2016. (Cited on page 35.)

[53] A. Cox, and S. Källblad, *Model-independent bounds for Asian options: a dynamic programming approach.* SIAM Journal on Control and Optimization, to appear. (Cited on page 24.)

[54] A. Cox and J. Oblój : *Robust pricing and hedging of double no-touch options*, Finance and Stochastics, 15(3): 573–605, 2011. (Cited on page 16.)

[55] A. Cox and J. Oblój : *Robust hedging of double touch barrier options*, SIAM Journal on Financial Mathematics, 2: 141–182, 2011. (Cited on page 16.)

[56] A. Cox, D. Hobson and J. Obłój : *Pathwise inequalities for local time: applications to Skorokhod embeddings and optimal stopping*, Ann. Appl. Probab., 18(5): 1870–1896, 2008. (Cited on page 16.)

[57] A. Cox, J. Obłój and N. Touzi : *The Root solution to the multi-marginal embedding problem: an optimal stopping and time-reveral approach*, preprint, 2015. (Cited on page 16.)

[58] A. Cox and J. Wang : *Root's barrier: Construction, optimality and applications to variance options*, Annals of Applied Probability (2013) 23 (3), 859-894. (Cited on page 16.)

[59] M.G. Crandall, H. Ishii, and P-L. Lions, *User's guide to viscosity solutions of second order partial differential equations*, Bull. Amer. Math. Soc. (N.S.), 27(1):1-67, 1992. (Cited on page 49.)

[60] Cvitanić, J., Karatzas, I. and H.M., Soner. Backward stochastic differential equations with constraints on the gain-process, *The Annals of Probability*, 26, 1522–1551, 1998. (Cited on pages 42 and 44.)

[61] J. Cvitanić, D. Possamaï, and N. Touzi. Dynamic programming approach to principal–agent problems. *Finance and Stochastics*, to appear, 2015. (Cited on page 33.)

[62] M. Davis, J. Obłój and V. Raval : *Arbitrage Bounds for Weighted Variance Swap Prices*, Mathematical Finance 24(4): 821-854. (Cited on page 16.)

[63] D. Dawson, *Measure-valued Markov processes*, École d'été de probabilités de Saint-Flour XXI-1991: 1-260, 1993. (Cited on page 53.)

[64] H. De March, and N. Touzi, *Irreducible convex paving for decomposition of multi-dimensional martingale transport plans*, preprint, 2017. (Cited on page 14.)

[65] K. Debrabant and E.R. Jakobsen, *Semi-Lagrangian schemes for linear and fully nonlinear diffusion equations*, preprint, 2009. (Cited on page 46.)

[66] C. Dellacherie. Quelques résultats sur les maisons de jeu analytiques. *Séminaire de probabilités de Strasbourg*, XIX:222–229, 1985. (Cited on page 36.)

[67] C. Dellacherie and P.A. Meyer, *Probabilities and potential*, volume 29 of *North-Holland Mathematics Studies*, North-Holland Publishing Co., Amsterdam, 1978. (Cited on page 42.)

[68] C. Dellacherie and E. Lenglart. Sur des problèmes de régularisation, de recollement et d'interpolation en théorie des processus. In *Séminaire de Probabilités de Strasbourg XVI*, pages 298–313. Springer, 1982. (Cited on page 43.)

[69] J.D. Deuschel and D.W. Stroock, *Large deviations*, volume 137 of *Pure and Applied Mathematics*, Academic Press Inc., Boston, MA, 1989. (Cited on page 17.)

[70] Y. Dolinsky, *Numerical Schemes for G–Expectations*, Electronic Journal of Probability, 17(1):1-15, 2012. (Cited on pages 44 and 47.)

[71] Dolinsky, Y., Nutz, M., Soner, H.M. (2012). Weak approximations of $G-$expectations, *Stochastic Processes and their Applications*, 122(2):664–675. (Cited on page 48.)

[72] Y. Dolinsky and H.M. Soner : *Robust hedging and martingale optimal transport in continuous time*, Probability Theory and Related Fields, 160 (1?2):391-427, 2014. (Cited on pages 15, 17, 18 and 24.)

[73] Y. Dolinsky and H.M. Soner : *Martingale optimal transport in the Skorokhod space*, Stochastic processes and their applications, 125(10):3893-3931, 2015. (Cited on pages 15, 17 and 19.)

[74] Y. Dolinsky and H. M. Soner. *Robust hedging with proportional transaction costs.* Finance Stoch., 18(2):327–347, 2014. (Cited on pages 24 and 25.)

[75] Y. Dolinsky and H. M. Soner. *Convex duality with transaction costs*, Mathematics of Operations Research, 42(2):448-471, 2017. (Cited on page 24.)

[76] M.D. Donsker, *An invariance principle for certain probability limit theorems.* Memoirs of the American Mathematical Society, no. 6, 1951. (Cited on page 48.)

[77] M. Doumbia, N. Oudjane and X. Warin, *Computing expectations for general SDE with pure Monte Carlo methods*, preprint, 2016. (Cited on page 58.)

[78] B. Dupire, *Pricing and hedging with smiles.* Mathematics of derivative securities 1(1): 103-111, 1997. (Cited on page 11.)

[79] B. Dupire *Functional itô calculus*, SSRN papers, 2009. (Cited on page 33.)

[80] E.B. Dynkin, *An introduction to branching measure-valued processes.* No. 6. American Mathematical Soc., 1994. (Cited on page 53.)

[81] H. Edelsbrunner, *Algorithms in combinatorial geometry*, volume 10 of *EATCS Monographs on Theoretical Computer Science*, Springer-Verlag, Berlin, 1987. (Cited on page 29.)

[82] I. Ekren, C. Keller, N. Touzi, and J. Zhang. On viscosity solutions of path dependent PDEs. *The Annals of Probability*, 42(1):204–236, 2014. (Cited on pages 33 and 49.)

[83] I. Ekren, N. Touzi, and J. Zhang. Viscosity solutions of fully nonlinear parabolic path dependent PDEs: part I. *The Annals of Probability*, 44(2):1212–1253, 2016. (Cited on page 33.)

[84] I. Ekren, N. Touzi, and J. Zhang. Viscosity solutions of fully nonlinear parabolic path dependent PDEs: part II. *The Annals of Probability*, 44(4):2507–2553, 2016. (Cited on page 33.)

[85] N. El Karoui, *Les aspects probabilistes du contrôle stochastique*, Lecture Notes in Mathematics 876, Springer-Verlag, Berlin, 1981. (Cited on pages 32, 36 and 43.)

[86] N. El Karoui and S.-J. Huang. A general result of existence and uniqueness of backward stochastic differential equations. In N. El Karoui and L. Mazliak, editors, *Backward stochastic differential equations*, volume 364 of *Pitman research notes in mathematics*, pages 27–36. Longman, 1997. (Cited on page 40.)

[87] N. El Karoui, D. H. Nguyen, and M. Jeanblanc-Picqué, *Compactification methods in the control of degenerate diffusions: existence of an optimal control*, Stochastics, 20(3):169-219, 1987. (Cited on pages 32, 34 and 36.)

[88] N. El Karoui, and S. Roelly. *Propriétés de martingales, explosion et représentation de Lévy-Khintchine d'une classe de processus de branchement à valeurs mesures.* Stochastic processes and their applications 38.2: 239-266, 1991. (Cited on page 53.)

[89] N. El Karoui, S. Peng, and M.-C. Quenez. Backward stochastic differential equations in finance. *Mathematical Finance*, 7(1):1–71, 1997. (Cited on page 33.)

[90] S.N. Ethier and T.G. Kurtz, *Markov Processes: Characterization and Convergence*, Wiley Interscience, 2005. (Cited on page 38.)

[91] N. El Karoui and X. Tan. Capacities, measurable selection and dynamic programming part I: abstract framework. *arXiv preprint arXiv:1310.3363*, 2013. (Cited on page 36.)

[92] N. El Karoui and X. Tan. Capacities, measurable selection and dynamic programming part II: application in stochastic control problems. *arXiv preprint arXiv:1310.3364*, 2013. (Cited on page 36.)

[93] A. Fahim, N. Touzi, and X. Warin, *A probabilitic numerical method for fully nonlinear parabolic PDEs*, Ann. Appl. Probab., 21(4):1322-1364, 2010. (Cited on pages 46 and 47.)

[94] W. Fleming, *Generalized solutions in optimal stochastic control*, Differential Games and Control Theory, Kinston Conference 2, Lecture Notes in Pure and Applied Math. 30, Dekker, 1978. (Cited on page 32.)

[95] W. Fleming, and R. Rishel, *Deterministic and Stochastic Optimal Control*, Springer-Verlag, 1975. (Cited on page 33.)

[96] W. Fleming, and M. Soner, *Controlled Markov Processes and Viscosity Solutions*, Springer Verlag, 1993. (Cited on pages 33 and 35.)

[97] W.H. Fleming and P.E. Souganidis. On the existence of value functions of two-player, zero-sum stochastic differential games. Indiana Univ. Math. J. 38(2), 293-314, 1989. (Cited on pages 35 and 36.)

[98] M. Fuhrman and H. Pham *Randomized and backward SDE representation for optimal control of non-Markovian SDEs*, Annals of Applied Probability, vol 25(4), 2134-2167, 2015. (Cited on page 33.)

[99] A. Galichon, P. Henry-Labordère, and N. Touzi, *A stochastic control approach to no-arbitrage bounds given marginals, with an application to Lookback options*, Annals of Applied Probability, 2014. (Cited on pages 1, 5, 11 and 15.)

[100] P. Gassiat, H. Oberhauser and G. dos Reis : *Root's barrier, viscosity solutions of obstacle problems and reflected FBSDEs*, preprint, 2015. (Cited on page 16.)

[101] Giles, M. B. : *Multilevel Monte Carlo path simulation*, Oper. Res., 56(3):607-617, 2008. (Cited on page 54.)

[102] Graham, C. and Talay, D. *Stochastic Simulation and Monte Carlo Methods*, Mathematical Foundations of Stochastic Simulation. Vol. 68. Springer Science & Business Media, 2013. (Cited on page 54.)

[103] E. Gobet, J.P. Lemor, and X. Warin, *A regression-based Monte-Carlo method to solve backward stochastic differential equations*, Ann. Appl. Probab., 15(3):2172-2202, 2005. (Cited on page 45.)

[104] M. Grigorova, P. Imkeller, E. Offen, Y. Ouknine, and M.C. Quenez, *Reflected BSDEs when the obstacle is not right-continuous and optimal stopping.*, arXiv preprint arXiv:1504.06094, 2015. (Cited on page 44.)

[105] G. Guo, X. Tan and N. Touzi, *Tightness and duality of martingale transport on the Skorokhod space*, Stochastic Processes and their Applications, 127(3): 927-956, 2017. (Cited on pages 15 and 19.)

[106] G. Guo, X. Tan and N. Touzi, *Optimal Skorokhod embedding under finitely-many marginal constraints,* SIAM Journal on Control and Optimization, 54(4):2174-2201, 2016. (Cited on page 20.)

[107] G. Guo, X. Tan and N. Touzi, *On the monotonicity principle of optimal Skorokhod embedding problem*, SIAM Journal on Control and Optimization, 54(5):2478-2489, 2016. (Cited on page 27.)

[108] W. Guo, J. Zhang and J. Zhuo, *A Monotone Scheme for High Dimensional Fully Nonlinear PDEs*, Annals of Applied Probability, The Annals of Applied Probability, 25(3), 1540-1580, 2015. (Cited on page 46.)

[109] I. Gyöngy *Mimicking the one-dimensional marginal distributions of processes having an Itô differential*, Probab. Theory Relat. Fields, 71:501?516, 1986. (Cited on page 12.)

[110] U.G. Haussmann, *Existence of optimal Markovian controls for degenerate diffusions*, Stochastic differential systems (Bad Honnef, 1985), volume 78 of Lecture Notes in Control and Inform. Sci. 171-186, Springer, Berlin, 1986. (Cited on page 32.)

[111] P. Henry-Labordère, *Cutting CVA's Complexity*, Risk magazine (Jul 2012). (Cited on page 53.)

[112] P. Henry-Labordère, J. Oblój, P. Spoida and N. Touzi : *The maximum maximum of a martingale with given n marginals*, Annals of Applied Probability, to appear. (Cited on page 16.)

[113] P. Henry-Labordere, N. Oudjane, X. Tan, N. Touzi, X. Warin, *Branching diffusion representation of semilinear PDEs and Monte Carlo approximation*. arXiv preprint, 2016. (Cited on pages 54 and 55.)

[114] P. Henry-Labordère, X. Tan and N. Touzi : *An Explicit Martingale Version of the One-dimensional Brenier's Theorem with Full Marginals Constraint*, Stochastic processes and their applications, 126(9):2800-2834, 2016.. (Cited on pages 14, 16 and 26.)

[115] P. Henry-Labordere, X. Tan, N. Touzi *A numerical algorithm for a class of BSDEs via the branching process*. Stochastic Processes and their Applications. 28;124(2):1112-1140, 2014. (Cited on page 53.)

[116] P. Henry-Labordere, X. Tan, N. Touzi *Unbiased simulation of stochastic differential equations,* Annals of Applied Probability, to appear. (Cited on page 57.)

[117] P. Henry-Labordère and N. Touzi : *An Explicit Martingale Version of the one-dimensional Brenier Theorem*, Finance and Stochastics, to appear. (Cited on pages 16, 26 and 27.)

[118] Hirsch, F., Profeta, C., Roynette, B., Yor, M. : *Peacocks and Associated Martingales, with Explicit Constructions*, Springer (2011). (Cited on page 16.)

[119] D. Hobson, *Robust hedging of the lookback option*, Finance and Stochastics, 2:329-347, 1998. (Cited on pages 1, 5, 11, 16 and 20.)

[120] D. Hobson, *The Skorokhod Embedding Problem and Model-Independent Bounds for Option Prices*, Paris-Princeton Lectures on Mathematical Finance, 2010. (Cited on page 15.)

[121] D. Hobson, *Mimicking martingales*, Annals of Applied Probability, 26(4):2273-2303, 2016. (Cited on page 16.)

[122] D. Hobson and M. Klimmek : *Model independent hedging strategies for variance swaps*, Finance and Stochastics 16, 611-649, 2012 (DOI) 10.1007/s00780-012-0190-3. (Cited on page 16.)

[123] D. Hobson and M. Klimmek : *Robust price bounds for the forward starting straddle*, Finance and Stochastics Vol 19, Issue 1, 189-214, 2015. (Cited on page 16.)

[124] D. Hobson and M. Klimmek : *Maximising functionals of the joint law of the maximum and terminal value in the Skorokhod embedding problem*, Annals of Applied Probability Vol. 23, No. 5, p2020-2052 2013. (Cited on page 16.)

[125] D. Hobson and A. Neuberger : *Robust bounds for forward start options*, Mathematical Finance 22, 31-56, 2012. (Cited on page 16.)

[126] D. Hobson and A. Neuberger. *On the value of being American*, Finance and Stochastics, to appear. (Cited on page 22.)

[127] Z. Hou and J. Obłój : *On robust pricing-hedging duality in continuous time*, preprint, 2015. (Cited on page 15.)

[128] J. Jacod and A.N. Shiryaev, *Limit theorems for stochastic processes*, volume 228 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, Springer-Verlag, Berlin, 1987. (Cited on page 61.)

[129] J. JACOD and J. MEMIN, *Sur un type de convergence intermédiaire entre la convergence en loi et la convergence en probabilité*, Séminaire de Probabilité XV 1979/80, Lecture Notes in Mathematics, Vol. 850, 529-546, 1981. (Cited on pages 20 and 62.)

[130] Jacod, J., Mémin, J., Métivier, M. (1983). On tightness and stopping times, *Stoch. Proc. and their App.*, 14:109–146. (Cited on page 49.)

[131] A. Jakubowski : *A non-Skorohod topology on the Skorohod space*, Electronic journal of probability, 2(4):1-21, 1997. (Cited on page 19.)

[132] Yu. M. Kabanov and M. Safarian. *Markets with transaction costs*. Springer Finance. Springer, Berlin, 2009. (Cited on page 25.)

[133] S. Källblad, X. Tan and N. Touzi, *Optimal Skorokhod embedding given full marginals and Azema-Yor peacocks*, Annals of Applied Probability, 27(2):686-719, 2017. (Cited on pages 16 and 20.)

[134] L. Kantorovich, *On the translocation of masses*, C.R. (Doklady) Acad. Sci. URSS (N.S.), 37:199-201, 1942. (Cited on pages 12 and 13.)

[135] I. Karatzas and S.E. Shreve, *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*, Springer-Verlag, New York, second edition, 1991. (Cited on page 15.)

[136] Kebaier, A. : *Statistical Romberg extrapolation: A new variance reduction method and applications to option pricing*, Annals of Applied probability, 15(4):2681-2705, 2005. (Cited on page 54.)

[137] Hans G. Kellerer : *Markov-Komposition und eine Anwendung auf Martingale*, Mathematische Annalen (1972). (Cited on page 13.)

[138] I. Kharroubi, N. Langrené and H. Pham, *Discrete time approximation of fully nonlinear HJB equations via BSDEs with nonpositive jumps*, Annals of Applied Probability, 25.4: 2301-2338, 2015. (Cited on page 46.)

[139] I. Kharroubi and H. Pham *Feynman-Kac representation for Hamilton-Jacobi-Belllman IPDEs*, Annals of Probability, vol 43(4), 1823-1865, 2015 (Cited on page 33.)

[140] Kloeden, P., Platen, E. : *Numerical Solution of Stochastic Differential Equations*, Applications of Mathematics, Volume 23, Springer, 1992. (Cited on page 54.)

[141] N. Krylov, *Controlled diffusion processes*, Stochastic Modeling and Applied Probability, Vol. 14, Springer, 1980. (Cited on pages 33 and 35.)

[142] N. V. Krylov, *On the rate of convergence of finite-difference approximations for Bellmans equations with variable coefficients*, Probability theory and related fields 117.1: 1-16, 2000. (Cited on pages 33, 44 and 46.)

[143] N.V. Krylov, *The rate of convergence of finite-difference approximations for Bellman equations with Lipschitz coefficients*, Appl. Math. Optim., 52(3):365-399, 2005. (Cited on page 45.)

[144] H. Kushner, P. G. Dupuis *Numerical methods for stochastic control problems in continuous time*, Applications of Mathematics, 24, 1992. (Cited on pages 39, 44, 46, 47 and 48.)

[145] D. Lacker *Mean field games via controlled martingale problems: Existence of Markovian equilibria* Stochastic Processes and their Applications 125(7): 2856-2894, 2015. (Cited on page 39.)

[146] D. Lacker *A general characterization of the mean field limit for stochastic differential games* Probability Theory and Related Fields 165.3-4: 581-648, 2016. (Cited on page 39.)

[147] J.F. Le Gall, *Spatial branching processes, random snakes and partial differential equations.* Springer Science & Business Media, 1999. (Cited on page 53.)

[148] F.A. Longstaff and R.S. Schwartz, *Valuing American options by simulation: a simple least-square approach*, Review of Finnaical studies, 14:113-147, 2001. (Cited on page 45.)

[149] Ma, J., Protter, P., San Martin, J., Torres, S. (2002). Numerical method for backward stochastic differential equations, *Ann. Appl. Probab*, 12(1):302–316. (Cited on pages 45 and 48.)

[150] D. B. Madan, and M. Yor : *Making Markov martingales meet marginals: with explicit constructions*, Bernoulli 8(4):509-536, 2002. (Cited on page 16.)

[151] H. P. McKean, *Application of Brownian motion to the equation of Kolmogorov-Petrovskii-Piskunov*, Comm. Pure Appl. Math., Vol 28, 323-331, 1975. (Cited on page 53.)

[152] J.-F. Mertens. Théorie des processus stochastiques généraux applications, aux surmartingales. *Probability Theory and Related Fields*, 22(1):45–68, 1972. (Cited on page 42.)

[153] P.A. Meyer, *Une majoration du processus croissant naturel associé à une surmartingale*, Séminaire de probabilités (Strasbourg), 2:166–170, 1968. (Cited on page 43.)

[154] P.A. Meyer, and W.A. Zheng. *Tightness criteria for laws of semimartingales.* Annales de l'IHP Probabilités et statistiques. Vol. 20. No. 4. Elsevier, 1984. (Cited on page 63.)

[155] T. Mikami and M. Thieullen, *Duality theorem for the stochastic optimal control problem*, Stochastic Process. Appl., 116(12):1815-1835, 2006. (Cited on pages 15 and 18.)

[156] G. Monge, Mémoire sur la théorie des déblais et des remblais, Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année, 666-704, 1781. (Cited on pages 11 and 12.)

[157] A. Neuberger. *Bounds on the American option*, 2007, SSRN:http://ssrn.com/abstract=966333 (Cited on page 22.)

[158] M. Nisio. *Stochastic control related to branching diffusion processes.* J. Math. Kyoto Univ., 25(3):549?575, 1985. (Cited on page 53.)

[159] M. Nutz, F. Stebegg, and X. Tan *Multiperiod Martingale Transport*, preprint, 2017. (Cited on page 14.)

[160] M. Nutz and R. van Handel. Constructing sublinear expectations on path space. *Stochastic Processes and their Applications*, 123(8):3100–3121, 2013. (Cited on page 36.)

[161] J. Obłój, *The Skorokhod embedding problem and its offspring*, Probab. Surv., 1:321-390(electronic), 2004. (Cited on page 15.)

[162] J. Obłój, and P. Siorpaes, *Structure of martingale transports in finite dimensions*, preprint, 2017. (Cited on page 14.)

[163] E. Pardoux and S. Peng. *Adapted solution of a backward stochastic differential equation.* Systems & Control Letters 14(1):55-61, 1990. (Cited on page 33.)

[164] T. Parthasarathy, *Selection Theorems and their Applications*, Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1972. (Cited on page 36.)

[165] S. Peng, *A general stochastic maximum principle for optimal control problems.* Siam J. of Control and Optimization, 28 (4):966-979, 1990. (Cited on page 32.)

[166] S. Peng. Monotonic limit theorem of BSDE and nonlinear decomposition theorem of Doob-Meyer's type. *Probability Theory and Related Fields*, 113(4):473–499, 1999. (Cited on page 43.)

[167] Peng, S. (2007). G-Expectation, G-Brownian motion, and related stochastic calculus of Itô type, *Stochastic Analysis and Applications, Abel Symp. 2*, Springer Berlin, 541–567. (Cited on page 48.)

[168] E. Perkins, *The Cereteli-Davis solution to the H1-embedding problem and an optimal embedding in Brownian motion.* Seminar on stochastic processes, 1985. Birkhäuser Boston, 1986. (Cited on page 15.)

[169] H. Pham, *Continuous-time stochastic control and optimization with financial applications*, Springer, 2009. (Cited on page 33.)

[170] L.S. Pontryagin, *Mathematical theory of optimal processes.* CRC Press, 1987. (Cited on page 32.)

[171] D. Possamai and X. Tan, *Weak approximation of second order BSDEs*, Annals of Applied Probability, 25(5):2535-2562, 2015. (Cited on pages 47 and 48.)

[172] D. Possamaï, X. Tan, and C. Zhou. *Stochastic control for a class of nonlinear kernels and applications,* Annals of Probability, to appear. (Cited on pages 40 and 42.)

[173] A. Rasulov, G. Raimova, M. Mascagni, *Monte Carlo solution of Cauchy problem for a nonlinear parabolic equation*, Mathematics and Computers in Simulation, 80(6), 1118-1123, 2010. (Cited on page 53.)

[174] Z. Ren and X. Tan, On the convergence of monotone schemes for path-dependent PDE, Stochastic Processes and their Applications, to appear. (Cited on page 49.)

[175] Z. Ren, N. Touzi, and J. Zhang. Comparison of viscosity solutions of semi–linear path–dependent PDEs. *arXiv preprint arXiv:1410.7281*, 2014. (Cited on page 33.)

[176] Z. Ren, N. Touzi, and J. Zhang. An overview of viscosity solutions of path–dependent PDEs. In D. Crisan, B. Hambly, and T. Zariphopoulou, editors, *Stochastic analysis and applications 2014: in honour of Terry Lyons*, volume 100 of *Springer proceedings in mathematics and statistics*, pages 397–453. Springer, 2014. (Cited on page 33.)

[177] Z. Ren, N. Touzi, and J. Zhang. Comparison of viscosity solutions of fully nonlinear degenerate parabolic path–dependent PDEs. *SIAM Journal on Mathematical Analysis*, to appear, 2015. (Cited on page 33.)

[178] D. Revuz, and M. Yor. *Continuous martingales and Brownian motion* Vol. 293. Springer Science & Business Media, 2013. (Cited on page 15.)

[179] Rhee, C. H., Glynn, P. W. : *Unbiased estimation with square root convergence for SDE models*, Oper. Res., 63(5):1026-1043, 2015. (Cited on page 54.)

[180] S. Roelly and A. Rouault, *Construction et propriétés de martingales des branchements spatiaux interactifs.* International Statistical Review/Revue Internationale de Statistique: 173-189, 1990. (Cited on page 53.)

[181] D.H. Root, *The existence of certain stopping times on Brownian motion*, Ann. Math. Statist., 40:715-718, 1969. (Cited on page 15.)

[182] H. Rost, *Skorokhod stopping times of minimal variance*, in Séminaire de Probabilités, X (Première partie, Univ. Strasbourg, Strasbourg, année universitaire), pages 194-208. Lecture Notes in Math., Vol 511. Springer, Berlin, 1976. (Cited on page 15.)

[183] G. Royer *Martingale Optimal Transport and Utility Maximization*, PhD Thesis, Ecole Polytechnique, 2014. (Cited on page 30.)

[184] A.I. Sakhanenko, *A new way to obtain estimates in the invariance principle*, High Dimensional Probability II, 221-243, 2000. (Cited on page 44.)

[185] A.V. Skorokhod *Branching diffusion processes*, Theory of Probability & Its Applications, 9(3):445-449, 1964. (Cited on page 53.)

[186] H.M. Soner, N. Touzi, and J. Zhang. *Wellposedness of second order backward SDEs*. Probability Theory and Related Fields, 153(1-2):149–190, 2012. (Cited on pages 33, 39, 40 and 42.)

[187] V. Strassen, *The existence of probability measures with given marginals*, Ann. Math. Statist., 36:423-439, 1965. (Cited on page 13.)

[188] D.W. Stroock and S.R.S. Varadhan, *Multidimensional diffusion processes*, volume 233 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, Springer-Verlag, Berlin, 1979. (Cited on page 37.)

[189] Talay, D. and Tubaro, L. *Expansion of the global error for numerical schemes solving stochastic differential equations*. Stochastic Analysis and Applications, 8:483-509, 1990. (Cited on page 54.)

[190] X. Tan, *A splitting method for fully nonlinear degenerate parabolic PDEs*, Electron. J. Probab. 18(15):1-24, 2013. (Cited on page 46.)

[191] X. Tan, *Discrete-time probabilistic approximation of path-dependent stochastic control problems*, Annals of Applied Probability, 24(5):1803-1834, 2014. (Cited on pages 46 and 47.)

[192] X. Tan and N. Touzi, *Optimal Transportation under Controlled Stochastic Dynamics,* Annals of Probability, Vol. 41, No. 5, 3201-3240, 2013. (Cited on pages 1, 5, 15, 18 and 29.)

[193] S. Tang and X. Li, *Necessary conditions for optimal control of stochastic systems with random jumps.* Siam J. of Control and Optimization, 32 (5):1447-1475, 1994. (Cited on page 32.)

[194] N. Touzi, *Optimal stochastic control, stochastic target problems, and backward SDE*. Springer Science & Business Media, 2012. (Cited on pages 33 and 35.)

[195] P. Vallois, *Le probleme de Skorokhod sur IR: une approche avec le temps local.* Séminaire de Probabilités XVII 1981/82. Springer, Berlin, Heidelberg. 227-239, 1983. (Cited on page 15.)

[196] C. Villani, *Optimal transport: old and new.* Vol. 338. Springer Science & Business Media, 2008. (Cited on pages 14, 25 and 28.)

[197] S. Watanabe, *On the branching process for Brownian particles with an absorbing boundary*, J. Math. Kyoto Univ. 4(2), 385-398, 1964. (Cited on page 53.)

[198] J. Yong and X.Y. Zhou, *Stochastic Controls. Hamiltonian Systems and HJB Equations.* Vol. 43 in Applications of Mathematics Series. Springer-Verlag, New York, 1999. (Cited on page 33.)

[199] J. Zhang. *A numerical scheme for backward stochastic differential equations*, Ann. of Appl. Probab., 14(1):459–488, 2004. (Cited on page 45.)

[200] J. Zhang and J. Zhuo, *Monotone schemes for fully nonlinear parabolic path dependent PDEs*, Journal of Financial Engineering, 1, 2014. (Cited on page 51.)