



HAL
open science

Deterministic Mean Field Games with Jumps

Annette Dumas

► **To cite this version:**

Annette Dumas. Deterministic Mean Field Games with Jumps. Mathematics [math]. Lyon 1, 2024. English. NNT: . tel-04805842

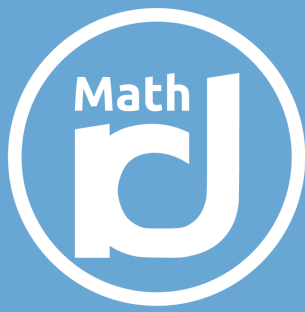
HAL Id: tel-04805842

<https://theses.hal.science/tel-04805842v1>

Submitted on 26 Nov 2024

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

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



Institut
Camille
Jordan

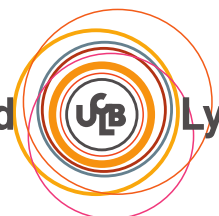
Laboratoire de recherche en mathématiques Lyon/Saint-Étienne

Jeux à champ moyen déterministes avec sauts



Annette Dumas
Thèse de doctorat

Université Claude Bernard



Lyon 1



Université Claude Bernard  Lyon 1

THÈSE de DOCTORAT DE L'UNIVERSITÉ DE LYON

Opérée au sein de :

l'Université Claude Bernard Lyon 1

Ecole Doctorale ED 512

InfoMaths

Spécialité de doctorat : Mathématiques

Soutenue publiquement le 26 Septembre 2024, par :

Annette Dumas

Jeux à champ moyen déterministes avec sauts

Devant le jury composé de :

Ariela Briani

Maître de Conférences, Université de Tours

Pierre Cardaliaguet

Professeur des Universités, Université Paris-Dauphine

Simon Masnou

Professeur des Universités, Université Lyon 1

Nadia Oudjane

Ingénieure de Recherche, EDF R&D - Palaiseau

Nelly Pustelnik

Directrice de Recherche, CNRS Lyon

Filippo Santambrogio

Professeur des Universités, Université Lyon 1

Francisco Silva

Maître de Conférences, Université de Limoges

Examinatrice

Rapporteur

Examineur

Examinatrice

Examinatrice

Directeur de thèse

Rapporteur

Remerciements

Je tiens à adresser mes remerciements aux personnes qui m'ont aidée à réaliser ce projet.

Merci tout d'abord à Filippo, mon directeur de thèse, de m'avoir fait confiance pour travailler sur ce sujet de thèse qui est très intéressant et passionnant. Merci de m'avoir guidée durant ces trois années. Tu as toujours été là pour répondre à toutes mes questions aussi bêtes soient-elles, mathématiques ou non, malgré tes multiples responsabilités. Je suis très fière de t'avoir comme directeur de thèse. Ces quelques mots ne suffisent pas à mesurer ma reconnaissance.

Merci aux membres du jury Ariela Briani, Simon Masnou, Nadia Oudjane, Nelly Pustelnik d'avoir accepté d'évaluer mon travail. Merci en particulier à Pierre Cardaliaguet et à Francisco Silva d'avoir pris le temps de lire mon manuscrit.

Merci à Clément Berger de m'avoir donné l'idée d'utiliser les algorithmes proximaux pour faire les simulations numériques de notre problème. On ne s'est vu que très peu de fois, mais à chaque fois c'est très productif.

Merci à Roland Denis de m'avoir aidée à rendre le code de l'algorithme plus rapide. Grâce à tes conseils, j'ai pu faire des simulations que je n'arrivais pas à faire avant.

Merci aux doctorants du bureau 118 de me donner envie de venir au laboratoire tous les jours et de mettre une ambiance détendue, drôle et sérieuse à la fois. A Matthieu, le M, qui mérite une demi-page de remerciements. Ton arrivée au bureau a été déterminante pour l'atmosphère. Ton humour est capable de désarmer n'importe qui et de créer des liens inattendus. A Guillaume (la famille !) pour les discussions de maths (je ne suis pas très sûre de t'avoir aidé, en tout cas tu m'as bien aidée) et les discussions de films, de jeux, etc. Merci à tous les deux pour les séances Google Maps qui permettent de voyager sans bouger. A Luca, Adrien et Julien, pour votre bonne humeur. A Dimitri, ancien membre du bureau, pour tes nombreux conseils et ta bienveillance. A Hugo, nouveau membre, pour perpétuer la tradition Auvergnate de ce bureau.

Merci à tous les autres doctorants d'avoir fait de ces trois années les meilleures années de ma vie jusqu'ici. A Sébastien d'être une figure rassurante. Merci pour les discussions de maths et de la vie en général, et aussi d'avoir relu une partie de cette thèse. A Mariane, briseuse de barrières, mon adversaire préférée aux échecs, même si on n'aime toutes les deux plus ça. J'aime beaucoup ta spontanéité, ton honnêteté et tes questions intéressantes. Merci de m'avoir appris quelques mots importants en Libanais. A Léa, pour ta gentillesse et ta positivité. A Wissam, le soleil de l'ICJ, toujours là, même derrière les nuages. Merci pour ta bonne humeur, les quizz et ton intelligence à toute épreuve. A Luca, sans 's', ta patience et ton ouverture d'esprit sont un exemple pour moi. A Thibault, frère de thèse, j'admire ta franchise et ta très grande intelligence aussi bien en maths qu'en perception des gens. Merci pour ton aide apportée en applications économiques et en transport optimal. A Antoine, pour être un modèle inatteignable de travail et de motivation. A tous ceux avec qui on a partagé des moments joyeux ensemble, Baptiste, Mouksit, Fanch, Yvon, Guillaume, Moreno, Kaï, Pierrick, Martín (j'ai toujours pas compris ta blague, désolée), Pablo, Rémi, Solal, Lorenzo, Nhan-Trung,... Ainsi qu'aux plus anciens désormais diplômés, Tibo, Martin, Jorge, Simon, Valentin,...

Merci à tous ceux que j'ai rencontrés dans les conférences et qui ont rendu des moments potentiellement stressants très chaleureux.

J'aimerais remercier également des personnes qui m'ont marquée avant la thèse sans qui je ne serais peut-être pas là.

Merci Clément et Vianney d'avoir été des phares durant ma scolarité à l'ENS de Lyon.

Merci à Madame Pélisse, ma professeuse de lycée, pour de nombreuses choses. Tout d'abord pour l'enseignement des mathématiques à travers des questions et non à travers des réponses. Ensuite de m'avoir conseillée de ne pas faire la remise à niveau scientifique après mon bac ES et d'aller directement en fac de maths à l'aide d'un livre de maths prêté pendant les vacances. Merci à Claire Debord de m'avoir encouragée à passer le second concours de l'ENS de Lyon et à Yanick Heurteaux de m'avoir aidée à écrire le petit mémoire sur les objets fractals. Merci à mes maîtres de stage Christophe Lacave et Kang-Ling Liao de m'avoir donné goût à la recherche.

Merci à mes parents et en particulier Papa de m'inspirer la curiosité pour toute chose. Merci Hélène et Matthieu de m'avoir appris à profiter de la vie et de m'avoir aidée à organiser le pot de thèse. Merci Karen, Lucas et Laurie pour votre bonne humeur et votre joie.

Enfin, merci Paul de m'avoir appris à savoir ce qui était important dans la vie et de me supporter à travers les bons et les mauvais moments.

Contents

Introduction en français	7
Introduction en anglais	13
1 Mean Field Games	19
1.1 Classical Mean Field Games	19
1.2 Variational Mean Field Games	21
1.3 Mean Field Games with jumps	24
1.4 Mixed Mean Field Games	27
1.4.1 The jump case	33
1.4.2 The kinetic case	35
2 Equivalence between Eulerian and Lagrangian	37
2.1 Equivalence between Eulerian and Lagrangian problems	38
2.2 Banach fixed point	41
3 Equivalence between optimizers and Nash equilibria	45
3.1 The jump case	47
3.2 The kinetic case	51
4 Regularity for Variational Mean Field Games with jumps	55
4.1 Lipschitz regularity in the Euclidean setting	56
4.2 Lipschitz regularity in the L^1 setting	61
4.3 Infinite horizon problems	70
4.4 Regularity in space	71
5 Numerical simulations	77
5.1 Introduction to non-smooth optimization algorithms	78
5.1.1 The Proximal Gradient method	78
5.1.2 The proximal operator	80
5.1.3 FISTA	82
5.1.4 Dual Proximal Gradient method	83
5.2 Numerical simulations	86
5.2.1 1D in time and periodic	87
5.2.2 1D in time and 1D in space, periodic in time	89
5.2.3 1D in time, 1D in space and non periodic in time	91
5.3 Numerical simulations for the density dependent case	94
6 Multi-species Modeling	101
6.1 A two-population model	101
6.2 Land Use Between Firms and Workers	102

Introduction en français

La physique statistique a été introduite au milieu du 19^{ème} siècle par Boltzmann et Maxwell pour décrire la dynamique d'un gaz composé d'un grand nombre de particules. L'approche microscopique consistant à suivre individuellement chaque particule a été délaissée au profit d'une approche macroscopique.

En terme de planification sociale ou urbaine, il est tentant à l'échelle d'une ville avec un grand nombre d'habitants d'adopter un point de vue macroscopique à l'aide de la physique statistique pour modéliser les dynamiques de populations. Cependant, à la différence des particules, chaque individu est différent et évolue en fonction de son expérience et des rencontres qu'il fait, ce qui rend complexe toute tentative de modélisation de la société. Une revue de la littérature en physique statistique appliquée aux dynamiques sociales [CFL09] montre que l'idée d'approcher la société à l'aide de la physique statistique a connu un intérêt croissant depuis les années 2000, mais le bilan montre un déséquilibre entre les preuves empiriques et les modélisations théoriques dont la dernière est favorisée. De plus, très peu de liens entre données et théories ont été faits.

Malgré cela, la théorie des jeux à champ moyen a émergé en 2006, grâce à Lasry et Lions et au même moment grâce à Caines, Huang et Malhamé. Leurs travaux consistent à étendre la notion d'équilibre de Nash pour N joueurs à un nombre infini de joueurs où chaque joueur a une influence négligeable sur le jeu, mais l'ensemble des stratégies forme une densité globale à la manière de la physique statistique, d'où le choix du terme *champ moyen*. De nombreuses applications existent dans les modèles économiques dont quelques unes par exemple sont présentées dans la thèse de Guéant [Gué09].

Les travaux de cette thèse sont motivés par la modélisation de la dynamique de la population au sein d'une ville où chaque habitant déménage d'un endroit à un autre. De ce fait, la stratégie d'un joueur ici est une courbe dépendant du temps qui est constante par morceaux à valeurs dans l'espace qui est la ville. Cette stratégie est optimisée en fonction du nombre de déménagements effectués et d'une densité globale donnée de la population. Si chaque joueur résout ce problème d'optimisation, cela génère un nouvel ensemble de stratégies qui forme une autre densité de la population. Si la densité initiale et la densité finale sont égales, alors cette densité est un point fixe d'un certain opérateur. De cette façon, nous avons une idée d'approche pour montrer l'existence d'un équilibre de Nash pour un jeu avec un continuum de joueurs dont les trajectoires sont discontinues et minimisant le nombre de ces discontinuités. Nous aurions voulu faire plus de modélisations, mais les jeux à champs moyen de ce type sont nouveaux et cela nous a fait rencontrer de nombreuses propriétés non-triviales et intéressantes. En plus d'une formulation rigoureuse de l'aspect point fixe en lien avec une formulation variationnelle où il faut choisir un espace fonctionnel adapté des trajectoires BV, nous obtenons une formulation variationnelle faisant apparaître un terme de longueur d'une courbe à valeurs dans l'espace de mesures muni ici de la norme TV . D'un point de vue numérique, le terme de longueur implique l'utilisation d'un terme non-lisse mais proximal qui est la norme L^1 de la vitesse et qui est appliqué à l'image de l'inconnue par un opérateur linéaire (ici, la dérivée).

Dans un premier temps, le Chapitre 1 introduit la théorie des jeux à champ moyen qui consiste à prendre un jeu à N joueurs et faire tendre N vers l'infini. À l'aide de la théorie du contrôle optimal, nous obtenons une solution (u, ρ) d'un système de deux équations: la première est une équation d'Hamilton-Jacobi dont u est la fonction valeur qui dépend de la densité et la deuxième est une équation de continuité décrivant le flot de la densité dont la vitesse dépend de l'Hamiltonien appliqué au gradient de u . Dans cette thèse, nous adoptons le point de vue variationnel des jeux à champ moyen dont l'idée provient initialement de la reformulation de Benamou et Brenier du problème de transport optimal comme un problème de minimisation d'une énergie cinétique. Dans notre cas, le coût de transport est trivial, c'est-à-dire que cela coûte 1 de changer d'endroit quelque soit la distance. Les jeux à champ moyen avec sauts ont été introduits par Bertucci en étendant le contrôle impulsif au cas champ moyen avec une infinité de joueurs. Néanmoins, l'aspect stochastique semble cruciale dans les travaux de Bertucci pour montrer l'existence d'une solution au système, tandis que la formulation variationnelle de notre problème nous permet de considérer les trajectoires déterministes. Dans la Section 1.4, nous montrons l'existence d'un

équilibre de Nash \bar{Q} comme optimiseur d'un problème où chaque agent résout

$$\inf_{\substack{\gamma: [0,T] \rightarrow \Omega \\ \gamma(0)=x_0}} J_{\bar{Q}}(\gamma) := A(\gamma) + \int_0^T \frac{dI}{dm}(e_t \# \bar{Q})(\gamma(t)) dt + \int_0^T F(t, \gamma(t), e_t \# \bar{Q}) dt + \psi_T(\gamma(T)),$$

où $\Omega \subset \mathbb{R}^n$ représente la ville. La fonction A décrit soit le nombre de sauts, soit l'énergie cinétique. Le problème contient deux termes dépendant de la densité $e_t \# \bar{Q}$: le premier (dI/dm) est variationnel, c'est-à-dire que nous pouvons calculer son intégrale première; le second (F) est non-variationnel, autrement dit, nous ne pouvons calculer son intégrale première, mais sa continuité suffit pour obtenir des résultats d'existence et de régularité de la solution. De ce fait, nous appelons ce type de problème *jeux à champ moyen mixtes*. Le dernier terme est une pénalisation en temps final. En notant \mathcal{C} l'espace des trajectoires, nous utilisons le théorème de point fixe de Kakutani pour montrer que la multifonction

$$\bar{Q} \mapsto \operatorname{argmin}_{Q, e_0 \# Q = m_0} \left\{ \int_{\mathcal{C}} A(\gamma) dQ(\gamma) + \int_0^T I(e_t \# Q) dt + \int_{\mathcal{C}} \int_0^T F(t, \gamma(t), e_t \# \bar{Q}) dt dQ(\gamma) + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ(\gamma) \right\}$$

admet un point fixe dans $\mathcal{P}(\mathcal{C})$. Il reste à démontrer que ce point fixe est bien une mesure supportée sur les trajectoires optimales, mais ce sera l'objet du Chapitre 3. Ce Chapitre 1 est aussi bien une introduction au sujet qu'une présentation de résultats nouveaux (ceux du cas mixte) qui constituent la Section 2 de [DS24].

Dans le Chapitre 2, nous réécrivons le problème

$$\min_{Q \in \mathcal{P}(\mathcal{C})} \left\{ \int_{\mathcal{C}} A(\gamma) dQ(\gamma) + \int_0^T I(e_t \# Q) dt + \int_{\mathcal{C}} \int_0^T V(t, \gamma(t)) dt dQ(\gamma) + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ(\gamma) \right\}$$

sous la forme Eulérienne

$$\min_{\rho: [0,T] \rightarrow \mathcal{P}(\Omega)} \mathbf{A}(\rho) + \int_0^T I(\rho(t)) dt + \int_0^T \int_{\Omega} V(t, \cdot) d\rho(t) dt + \int_{\Omega} \psi_T d\rho(T).$$

Nous considérons les cas suivants:

- $\mathbf{A}(\rho) = \frac{1}{2} \int_0^T \int_{\Omega} |\dot{\rho}| dx dt$ lorsque A décrit le nombre de sauts S .
- $\mathbf{A}(\rho) = \int_0^T \frac{1}{p} |\rho'|_{W_p}(t)^p dt$ où A est l'énergie cinétique $\int_0^T \frac{|\gamma'|^p}{p} dt$ et $|\cdot|_{W_p}$ est la dérivée métrique selon la distance de Wasserstein d'ordre $p > 1$.

Précédemment, nous avons montré l'existence d'un point fixe qui est une mesure de probabilité \bar{Q} sur l'ensemble des trajectoires utilisées par les agents. Ce point de vue correspond au point de vue Lagrangien, car nous suivons individuellement chaque particule. Dans la reformulation Eulérienne, pour chaque temps t nous observons l'allure de la densité qui cette fois est une mesure de probabilité sur l'espace Ω , en prenant la mesure image de Q par l'application d'évaluation au temps t , i.e. $\rho_t = e_t \# Q$. Dans le cas où l'énergie correspond au nombre de sauts, le nouveau problème consiste à minimiser la longueur de la courbe $\rho: [0, T] \rightarrow \mathcal{P}(\Omega)$. Cette réécriture, démontrée dans la Section 4 de [DS24], provient d'une propriété importante en transport optimal, où lorsque le coût de transport est trivial, c'est-à-dire $c(x, y) = \mathbb{1}_{x \neq y}$, alors nous avons ceci:

$$\inf_{\pi \in \Pi(\mu, \nu)} \int_{\Omega \times \Omega} \mathbb{1}_{x \neq y} d\pi(x, y) = \frac{1}{2} \|\mu - \nu\|_{TV}.$$

L'intérêt de cette reformulation est multiple. En effet, nous pourrions étudier la régularité de la solution ρ dans le Chapitre 4 et utiliser des algorithmes d'optimisation non-lisses pour effectuer des simulations numériques dans le Chapitre 5. Dans le cas de l'énergie cinétique, l'équivalence Eulérienne-Lagrangienne est déjà connue et présentée par exemple dans [BCS17]. Après avoir montré l'existence d'une mesure d'équilibre dans la Section 1.4 pour le problème Lagrangien, nous montrons également dans la Section 2.2 (qui se trouve aussi dans [DS24, Section 5]) l'existence d'une mesure d'équilibre pour le problème Eulérien en définissant

$$\begin{aligned} \mathcal{F}: L^2_{t,x} &\rightarrow L^2_{t,x} \\ \bar{\rho} &\mapsto \operatorname{argmin}_{\rho} \mathbf{A}(\rho) + \int_0^T I(\rho(t)) dt + \int_0^T \int_{\Omega} V(t, \bar{\rho}(t), \cdot) d\rho(t) dt + \int_{\Omega} \psi_T d\rho(T). \end{aligned}$$

La fonction I est définie par

$$I(\rho) = \begin{cases} \int_{\Omega} f(\rho(x)) dx & \text{si } \rho \ll \mathcal{L}^d, \\ +\infty & \text{sinon,} \end{cases}$$

avec f une fonction c_0 -convexe. Si V est un opérateur Lipschitzien en fonction de $\bar{\rho}$ tel que $\frac{\text{Lip}(V)}{c_0} < 1$, alors \mathcal{F} est une contraction, ce qui permet d'utiliser le théorème de point fixe de Banach. Ce dernier offre un moyen d'approcher numériquement la solution lorsque le problème est mixte (cf. Section 5.3).

Dans le Chapitre 3 (et aussi dans [DS24, Section 3]), nous montrons rigoureusement que la mesure \bar{Q} obtenue comme point fixe dans la Section 1.4 est bien un équilibre de Nash, c'est-à-dire que \bar{Q} est concentré sur les courbes minimisant $J_{\bar{Q}}$. Ceci équivaut à

$$\forall Q, \int_{\mathcal{C}} J_{\bar{Q}}(\gamma) d\bar{Q}(\gamma) \leq \int_{\mathcal{C}} J_{\bar{Q}}(\gamma) dQ(\gamma). \quad (1)$$

Ainsi, cette mesure est supportée sur les trajectoires optimales des joueurs en sachant que les autres joueurs sont distribués aussi selon cette mesure. La preuve consistant à dériver

$$\mathcal{U}_{\bar{Q}}(Q) = \int_{\mathcal{C}} A(\gamma) dQ(\gamma) + \int_0^T I(e_t \# Q) dt + \int_{\mathcal{C}} \int_0^T F(t, \gamma(t), e_t \# \bar{Q}) dt dQ(\gamma)$$

dans une certaine direction au sens de Gâteaux permet d'obtenir (1) lorsque Q vérifie $\int I(e_t \# Q) < \infty$. Pour montrer que (1) est vraie pour toute mesure Q , il faut approcher Q par une mesure Q_r obtenue en remplaçant chaque courbe par une translation (à modification près en $t = 0$) et en moyennisant selon le paramètre de translation. De ce fait, Q_r vérifie bien (1) et en passant à la limite $r \rightarrow 0$, nous obtenons que \bar{Q} est un équilibre de Nash. Notre preuve a été construite en particulier pour le cas où les trajectoires sont constantes par morceaux, car la preuve contenue dans [CMS16] et inspirée des travaux d'Ambrosio et Figalli [AF09] pour les équations d'Euler incompressible ne marche pas ici. En effet, la difficulté réside dans la non-séparabilité de l'espace BV et la non-continuité de cette fonctionnelle avec la convergence forte. Une fois la nouvelle approche établie pour le cas avec discontinuité, nous pouvons aussi retrouver un résultat connu pour les jeux à champ moyen *classiques* avec une nouvelle preuve. Cela demande une hypothèse sur les exposants (i.e. $q < 1 + \frac{p'}{d}$ où f a une croissance polynomiale d'exposant au plus q et p' est l'exposant conjugué de p), mais cette hypothèse n'est pas nouvelle. Elle apparaît en particulier dans de nombreux articles au début de l'analyse des MFG variationnels d'ordre 1 (par exemple [CG15]). L'approche pour montrer que \bar{Q} est bien un équilibre de Nash demande des propriétés de régularité (continuité et bornitude de la densité), mais ces propriétés ont été obtenues grâce à la formulation Eulérienne du problème.

Ces régularités sont rassemblées dans le Chapitre 4 et apparaissent également dans l'article [DS23]. Nous commençons par étudier la régularité de la solution $\gamma: [0, 1] \rightarrow \mathbb{R}^d$ à valeurs dans un espace de dimension finie de

$$\min \left\{ TV(\gamma; [0, 1]) + \int_0^1 F(t, \gamma(t)) dt + \psi_0(\gamma(0)) + \psi_1(\gamma(1)) : \gamma \in BV([0, 1]; \mathbb{R}^d) \right\}. \quad (2)$$

En approchant ce problème par

$$\min \left\{ \int_0^1 (L_\varepsilon(\gamma'(t)) + F(t, \gamma(t))) dt + \alpha_\varepsilon(\psi_0(\gamma(0)) + \psi_1(\gamma(1))) : \gamma \in H^1([0, 1]) \right\},$$

en utilisant le système d'Euler-Lagrange et en maximisant une fonction appropriée, nous obtenons une borne uniforme sur la dérivée de la solution

$$|\gamma'_\varepsilon| \leq \max \left\{ \frac{C_0}{c_0}, \frac{\alpha_\varepsilon \varepsilon}{\sqrt{1 - \alpha_\varepsilon^2}} \right\}. \quad (3)$$

En appliquant le théorème d'Arzelà-Ascoli et la semi-continuité de la variation totale TV , la suite γ_ε converge à une sous-suite près vers la solution de (2). De plus, en passant à la limite dans l'inégalité (3), la solution vérifie

$$|\gamma'| \leq C,$$

ce qui est une autre façon de dire que la solution est Lipschitz en temps. De plus, lorsque $d = 1$, la solution est soit constante, soit elle est un point critique de F . Ceci nous permet d'avoir une idée de l'allure de la courbe. Ensuite, nous montrons une régularité similaire pour la solution du problème

$$\min_{\substack{\rho: [0, T] \rightarrow L^1(\Omega) \\ \rho \geq 0 \\ \forall t \in [0, T], \int_{\Omega} \rho(t, x) dx = 1}} \int_0^T \int_{\Omega} (|\dot{\rho}(t, x)| + V(t, x)\rho(t, x) + f(\rho(t, x))) dx dt + \psi_0(\rho(0)) + \psi_T(\rho(T)), \quad (4)$$

où cette fois ρ est une courbe à valeurs dans un espace de dimension infinie. Nous approchons ρ par une courbe dans un espace de dimension finie $\rho_n : [0, T] \rightarrow \mathbb{R}^n$, puis nous utilisons des techniques analogues qui permettent d'obtenir

$$\sup_{t \in [0, T]} \int_{\Omega} |\dot{\rho}(t, x)|^2 dx \leq C,$$

autrement dit, $\rho :]0, T[\rightarrow L^2(\Omega)$ est Lipschitz. Avec des ajustements supplémentaires, nous avons aussi que la solution de

$$\min_{\substack{\rho : [0, +\infty[\rightarrow L^1(\Omega) \\ \rho \geq 0 \\ \forall t \in \mathbb{R}_+, \int_{\Omega} \rho(t, x) dx = 1}} \int_0^{\infty} e^{-rt} \int_{\Omega} (|\dot{\rho}(t, x)| + V(t, x)\rho(t, x) + f(\rho(t, x))) dx dt + \psi_0(\rho(0))$$

est Lipschitz en temps à valeurs dans $L^2(\Omega)$. Lorsque le problème a des conditions de Dirichlet en temps, il est possible de montrer que la solution ρ admet la même régularité que les données du problème (c'est-à-dire les conditions aux bords et la fonction V) en espace presque partout.

Les simulations numériques de la solution de (4) sont rassemblées dans le Chapitre 5. Tout d'abord, nous introduisons les différents algorithmes d'optimisation existants en terminant par l'algorithme que nous utilisons, à savoir la méthode de gradient proximal duale rapide. Il est connu que la suite obtenue par cet algorithme converge à une vitesse d'ordre $O(1/k)$ vers la solution du problème (cf. [BT14]). Pour différents exemples de V et différentes conditions aux bords de $[0, T]$, nous effectuons quelques simulations de la solution de (4). Ces résultats numériques apparaissent dans [DS23, Section 6]. Dans le cas du problème mixte, c'est-à-dire lorsque la solution $\bar{\rho}$ qui minimise

$$\min_{\substack{\rho \in L^2([0, T] \times \Omega) \\ \rho \geq 0 \\ \forall t \in [0, T], \int_{\Omega} \rho(t, x) dx = 1}} \mathcal{V}_{\bar{\rho}}(\rho) = \int_0^T \int_{\Omega} \left(\lambda |\dot{\rho}(t, x)| + V(\bar{\rho})(t, x)\rho(t, x) + c_0 \frac{\rho(t, x)^2}{2} \right) dx dt + \psi_0(\rho(0)) + \psi_T(\rho(T))$$

est un point fixe d'un opérateur, alors il est possible de réaliser des simulations numériques de ce dernier grâce au théorème de point fixe de Banach qui fournit un moyen de le calculer. Notons que nous choisissons des exemples $V(\bar{\rho})$ non-variationnels et qui dépendent de $\bar{\rho}$ de façon non-autonome (pour que la solution ne soit pas constante en temps). Nous considérons aussi un exemple dépendant de toute l'histoire, c'est-à-dire $V(\rho)(t, x) \equiv \int_0^t \rho(s, x) dx$, qui peut avoir un intérêt pour la modélisation. Ces exemples sont présentés dans [DS24, Section 6].

Le dernier Chapitre 6 de cette thèse est une présentation de deux applications du modèle de jeux à champ moyen déterministes avec sauts dans les cas à plusieurs populations. Le premier exemple est un modèle de deux populations où les individus d'un groupe sont moins enclins à vivre proches des individus de l'autre groupe. Grâce à la formulation mixte du problème,

$$\mathcal{G} : L^2([0, T] \times \Omega) \times L^2([0, T] \times \Omega) \rightarrow L^2([0, T] \times \Omega) \times L^2([0, T] \times \Omega)$$

$$(\bar{\rho}_1, \bar{\rho}_2) \mapsto \left(\begin{array}{c} \operatorname{argmin}_{\substack{\rho_1 \geq 0 \\ \forall t, \int \rho_1(t) = \frac{1}{2}}} \mathcal{V}_{\bar{\rho}_1, \bar{\rho}_2}^1(\rho_1); \quad \operatorname{argmin}_{\substack{\rho_2 \geq 0 \\ \forall t, \int \rho_2(t) = \frac{1}{2}}} \mathcal{V}_{\bar{\rho}_1, \bar{\rho}_2}^2(\rho_2) \end{array} \right)$$

avec

$$\mathcal{V}_{\bar{\rho}}^1(\rho) := \int_0^T \int_{\Omega} \lambda |\dot{\rho}| + V_1(\bar{\rho})\rho + f(\rho) dx dt + \psi_0(\rho(0)) + \psi_T(\rho(T))$$

et $\mathcal{V}_{\bar{\rho}}^2(\rho) := \int_0^T \int_{\Omega} \lambda |\dot{\rho}| + V_2(\bar{\rho})\rho + f(\rho) dx dt + \psi_0(\rho(0)) + \psi_T(\rho(T))$,

nous obtenons des simulations numériques de la solution grâce au théorème de point fixe de Banach. Le deuxième exemple est un modèle inspiré de Carlier et Ekeland [CE07] sur l'aménagement du territoire selon les travailleurs et les entreprises. Dans leur article, ils interprètent le salaire perçu par les travailleurs φ et le salaire payé par les entreprises ψ comme des potentiels de Kantorovitch dont le coût de transport correspond aux frais de trajets pour se rendre sur le lieu de travail. Autrement dit, le problème de transport est

$$\inf_{\gamma \in \Pi(\mu, \nu)} \int_{\bar{\Omega} \times \bar{\Omega}} c(x, y) d\gamma(x, y)$$

avec μ décrivant la densité des travailleurs et ν la densité des entreprises, et ce problème est équivalent à

$$\sup_{\varphi, \psi} \left\{ \int_{\bar{\Omega}} \psi d\nu - \int_{\bar{\Omega}} \varphi d\mu ; \psi(y) - \varphi(x) \leq c(x, y), \forall (x, y) \in \bar{\Omega} \times \bar{\Omega} \right\}.$$

Nous considérons que chaque travailleur souhaite minimiser

$$\min_{\gamma(0)=x_0} 2\lambda_1 S(\gamma) + \int_0^T [-\varphi_{\mu,\nu}(\gamma(t)) + V_{0,1}(t, \gamma(t)) + f'(\mu)(t, \gamma(t))] dt + \phi_T(\gamma(T))$$

et chaque entreprise souhaite minimiser

$$\min_{\gamma(0)=x_0} 2\lambda_2 S(\gamma) + \int_0^T [\psi_{\mu,\nu}(\gamma(t)) + V_{0,2}(t, \gamma(t)) + f'(\nu)(t, \gamma(t))] dt + \phi_T(\gamma(T)).$$

Un équilibre de Nash (μ, ν) vérifierait le problème Eulérien suivant

$$\begin{aligned} \min_{\substack{\mu, \nu \geq 0 \\ \int \mu_t = \int \nu_t = \frac{1}{2}}} & \int_0^T \int_{\Omega} \lambda_1 |\dot{\mu}| + \lambda_2 |\dot{\nu}| + V_{0,1}\mu + V_{0,2}\nu + f(\nu) + f(\mu) dx dt + \int_0^T W_c(\mu_t, \nu_t) dt \\ & + \Phi_0(\mu(0)) + \Phi_0(\nu(0)) + \Phi_T(\mu(T)) + \Phi_T(\nu(T)), \end{aligned}$$

où $W_c(\mu, \nu) = \inf_{\gamma \in \Pi(\mu, \nu)} \int_{\bar{\Omega} \times \bar{\Omega}} c(x, y) d\gamma(x, y)$. Il est plus facile de calculer les potentiels de Kantorovitch grâce à l'algorithme de Sinkhorn. C'est pourquoi nous utilisons la régularisation de la distance de Wasserstein

$$\begin{aligned} W_{c,\varepsilon}(\mu, \nu) &= \inf_{\gamma \in \Pi(\mu, \nu)} \int_{\Omega \times \Omega} c(x, y) d\gamma(x, y) - \varepsilon \int \gamma(\log \gamma - 1) \\ &= \sup_{\varphi, \psi} \int_{\Omega} \psi(y) d\nu(y) - \int_{\Omega} \varphi(x) d\mu(x) - \varepsilon \int_{\Omega \times \Omega} e^{\frac{\psi(y) - \varphi(x) - c(x, y)}{\varepsilon}} d(\mu \otimes \nu)(x, y), \end{aligned}$$

qui permet aussi d'éliminer la contrainte $\psi(y) - \varphi(x) \leq c(x, y)$. Nous proposons trois façons différentes d'approcher le problème numériquement. Ce problème, dans sa version la plus simple, est variationnel. Pourtant, la première approche consiste à ignorer cela et appliquer un théorème de point fixe comme précédemment pour le problème mixte avec la fonction \mathcal{G} en définissant

$$\begin{aligned} \mathcal{V}_{\bar{\mu}, \bar{\nu}}^1(\mu) &:= \int_0^T \int_{\Omega} \lambda_1 |\dot{\mu}| - \varphi_{\bar{\mu}, \bar{\nu}} \mu + f(\mu) dx dt + \Phi_0(\mu(0)) + \Phi_T(\mu(T)) \\ \text{et } \mathcal{V}_{\bar{\mu}, \bar{\nu}}^2(\nu) &:= \int_0^T \int_{\Omega} \lambda_2 |\dot{\nu}| + \psi_{\bar{\mu}, \bar{\nu}} \nu + f(\nu) dx dt + \Phi_0(\nu(0)) + \Phi_T(\nu(T)). \end{aligned}$$

Des tentatives de simulations numériques ont été effectuées, mais les résultats ne sont pas encore convaincants et ne sont pas inclus dans cette thèse. La seconde approche exploite la structure variationnelle pour appliquer la méthode de gradient proximale duale dans deux cas différents: soit $W_{c,\varepsilon}$ est perçue comme une perturbation de la fonction uniformément convexe, soit comme une perturbation de la fonction convexe non-lisse. La troisième approche utilise également la formulation variationnelle en l'écrivant sous la forme

$$\inf_{\mu, \nu} \sup_{\varphi, \psi} \mathcal{L}(\mu, \nu, \varphi, \psi),$$

ce qui laisse la possibilité d'appliquer l'algorithme d'Uzawa, esquivant ainsi le calcul des potentiels de Kantorovitch avec l'algorithme de Sinkhorn. Par manque de temps, nous n'avons pas encore exploré ces pistes. Néanmoins, elle méritent des études plus approfondies.

Introduction en anglais

Statistical physics was introduced in the mid-19th century by Boltzmann and Maxwell to describe the dynamics of a gas composed of a large number of particles. The microscopic approach, which involves tracking each particle individually, was abandoned in favor of a macroscopic approach. In terms of social or urban planning, it is tempting to adopt a macroscopic viewpoint to model population dynamics using statistical physics in a city with a large number of inhabitants. However, unlike particles, each individual is different and evolves based on their experiences and interactions, making any attempt to model society complex. A review of the literature on statistical physics applied to social dynamics [CFL09] shows that the idea of approaching society using statistical physics has gained increasing interest since the 2000s, but there is an imbalance between empirical evidence and theoretical modeling, with the latter being favored. Moreover, very few links between data and theories have been made.

Despite this, the theory of Mean Field Games emerged in 2006, thanks to Lasry and Lions and simultaneously to Caines, Huang, and Malhamé. Their work extends the notion of Nash equilibrium for N players to an infinite number of players where each player has a negligible influence on the game, but the set of strategies forms a global density similar to statistical physics, hence the term *Mean Field*. There are numerous applications in economic models, some of which are presented in Guéant's thesis [Gué09].

The work of this thesis is motivated by modeling the population dynamics within a city where each inhabitant moves from one place to another. Thus, the strategy of a player here is a piecewise constant curve in time with values in the city. This strategy is optimized based on the number of moves made and a given global population density. If each player solves this optimization problem, it generates a new set of strategies that forms another population density. If the initial density and the final density are equal, then this density is a fixed point of a certain operator. This gives us an idea of an approach to show the existence of a Nash equilibrium for a game with a continuum of players whose trajectories are discontinuous and minimize the number of these discontinuities. We wanted to do more modeling, but Mean Field Games of this type are new and led us to encounter many non-trivial and interesting properties. In addition to a rigorous formulation of the fixed-point aspect linked to a variational formulation where an appropriate functional space for BV trajectories must be chosen, we obtain a variational formulation featuring a curve length term valued in the space of measures endowed with the TV norm. From a numerical point of view, the length term implies the use of a non-smooth but proximal term, which is the L^1 norm of the velocity applied to the unknown image by a linear operator (here, the derivative).

Firstly, Chapter 1 introduces the theory of Mean Field Games, which consists of taking a game with N players and letting N tend to infinity. Using optimal control theory, we obtain a solution (u, ρ) of a system of two equations: the first is a Hamilton-Jacobi equation where u is the value function depending on the density, and the second is a continuity equation describing the flow of the density whose velocity depends on the Hamiltonian applied to the gradient of u . In this thesis, we adopt the variational point of view of Mean Field Games, initially inspired by Benamou and Brenier's reformulation of the optimal transport problem as a kinetic energy minimization problem. In our case, the transport cost is trivial, meaning it costs 1 to change places regardless of the distance. Mean Field Games with jumps were introduced by Bertucci by extending impulsive control to the Mean Field case with an infinite number of players. However, the stochastic aspect seems crucial in Bertucci's work to show the existence of a solution to the system, while our variational formulation allows us to consider deterministic trajectories. In Section 1.4, we show the existence of a Nash equilibrium \bar{Q} as the optimizer of a problem where each agent solves

$$\inf_{\substack{\gamma: [0, T] \rightarrow \Omega \\ \gamma(0) = x_0}} J_{\bar{Q}}(\gamma) := A(\gamma) + \int_0^T \frac{dI}{dm}(e_t \# \bar{Q})(\gamma(t)) dt + \int_0^T F(t, \gamma(t), e_t \# \bar{Q}) dt + \psi_T(\gamma(T)),$$

where $\Omega \subset \mathbb{R}^n$ represents the city. The function A describes either the number of jumps or the kinetic energy. The problem contains two terms dependent on the density $e_t \# \bar{Q}$: the first (dI/dm) is variational, meaning we can calculate its first integral; the second (F) is non-variational, meaning we cannot calculate

its first integral, but its continuity is sufficient to obtain existence and regularity results for the solution. Therefore, we call this type of problem *mixed Mean Field Games*. The last term is a final time penalty. By noting \mathcal{C} the space of trajectories, we use Kakutani's fixed-point theorem to show that the multifunction

$$\bar{Q} \mapsto \operatorname{argmin}_{Q, e_0 \# Q = m_0} \left\{ \int_{\mathcal{C}} A(\gamma) dQ(\gamma) + \int_0^T I(e_t \# Q) dt + \int_{\mathcal{C}} \int_0^T F(t, \gamma(t), e_t \# \bar{Q}) dt dQ(\gamma) + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ(\gamma) \right\}$$

admits a fixed point in $\mathcal{P}(C)$. It remains to demonstrate that this fixed point is indeed a measure supported by the optimal trajectories, which will be the subject of Chapter 3. Chapter 1 is both an introduction to the subject and a presentation of new results (those of the mixed case) that constitute Section 2 of [DS24].

In Chapter 2, we rewrite the problem

$$\min_{Q \in \mathcal{P}(C)} \left\{ \int_{\mathcal{C}} A(\gamma) dQ(\gamma) + \int_0^T I(e_t \# Q) dt + \int_{\mathcal{C}} \int_0^T V(t, \gamma(t)) dt dQ(\gamma) + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ(\gamma) \right\}$$

in Eulerian form

$$\min_{\rho: [0, T] \rightarrow \mathcal{P}(\Omega)} \mathbf{A}(\rho) + \int_0^T I(\rho(t)) dt + \int_0^T \int_{\Omega} V(t, \cdot) d\rho(t) dt + \int_{\Omega} \psi_T d\rho(T).$$

We consider the following cases:

- $\mathbf{A}(\rho) = \frac{1}{2} \int_0^T \int_{\Omega} |\dot{\rho}| dx dt$ when A describes the number of jumps S .
- $\mathbf{A}(\rho) = \int_0^T \frac{1}{p} |\rho'|_{W_p}(t)^p dt$ where A is the kinetic energy $\int_0^T \frac{|\gamma'|^p}{p} dt$ and $|\cdot|_{W_p}$ is the metric derivative of the $p > 1$ Wasserstein distance.

Previously, we showed the existence of a fixed point which is a probability measure \bar{Q} on the set of trajectories used by the agents. This viewpoint corresponds to the Lagrangian point of view, as we follow each particle individually. In the Eulerian reformulation, for each time t we observe the density, which this time is a probability measure on the space Ω , by taking the push-forward measure of Q by the evaluation map at time t , i.e., $\rho_t = e_t \# Q$. In the case where the energy corresponds to the number of jumps, the new problem is to minimize the length of the curve $\rho: [0, T] \rightarrow \mathcal{P}(\Omega)$. This rewriting, demonstrated in Section 4 of [DS24], comes from an important property in optimal transport, which claims that when the transport cost is trivial, i.e., $c(x, y) = \mathbb{1}_{x \neq y}$, then we have this result:

$$\inf_{\pi \in \Pi(\mu, \nu)} \int_{\Omega \times \Omega} \mathbb{1}_{x \neq y} d\pi(x, y) = \frac{1}{2} \|\mu - \nu\|_{TV}.$$

The interest in this reformulation is multiple. Indeed, we will be able to study the regularity of the solution ρ in Chapter 4 and use non-smooth optimization algorithms to perform numerical simulations in Chapter 5. In the kinetic case, the Eulerian-Lagrangian equivalence is already known and presented, for example, in [BCS17]. After showing the existence of an equilibrium measure in Section 1.4 for the Lagrangian problem, we also show in Section 2.2 (also found in [DS24, Section 5]) the existence of an equilibrium measure for the Eulerian problem by defining

$$\begin{aligned} \mathcal{F}: L_{t,x}^2 &\rightarrow L_{t,x}^2 \\ \bar{\rho} &\mapsto \operatorname{argmin}_{\rho} \mathbf{A}(\rho) + \int_0^T I(\rho(t)) dt + \int_0^T \int_{\Omega} V(t, \bar{\rho}(t), \cdot) d\rho(t) dt + \int_{\Omega} \psi_T d\rho(T). \end{aligned}$$

The function I is defined by

$$I(\rho) = \begin{cases} \int_{\Omega} f(\rho(x)) dx & \text{if } \rho \ll \mathcal{L}^d, \\ +\infty & \text{otherwise,} \end{cases}$$

with f a c_0 -convex function. If V is a Lipschitz operator with respect to $\bar{\rho}$ such that $\frac{\operatorname{Lip}(V)}{c_0} < 1$, then \mathcal{F} is a contraction, allowing the use of Banach's fixed-point theorem. This theorem provides a means to numerically approximate the solution when the problem is mixed (see Section 5.3).

In Chapter 3 (and also in [DS24, Section 3]), we rigorously show that the measure \bar{Q} obtained as a fixed point in Section 1.4 is indeed a Nash equilibrium, meaning that \bar{Q} is concentrated on the curves minimizing $J_{\bar{Q}}$. This is equivalent to

$$\forall Q, \int_{\mathcal{C}} J_{\bar{Q}}(\gamma), d\bar{Q}(\gamma) \leq \int_{\mathcal{C}} J_{\bar{Q}}(\gamma), dQ(\gamma). \quad (5)$$

Thus, this measure is supported on the optimal trajectories of the players, knowing that the other players are also distributed according to this measure. The proof consists of deriving

$$\mathcal{U}_{\bar{Q}}(Q) = \int_{\mathcal{C}} A(\gamma) dQ(\gamma) + \int_0^T I(e_t \# Q) dt + \int_{\mathcal{C}} \int_0^T F(t, \gamma(t), e_t \# \bar{Q}) dt dQ(\gamma)$$

in a certain direction in the sense of Gâteaux to obtain (5) when Q satisfies $\int I(e_t \# Q) < \infty$. To show that (5) is true for any measure Q , it is necessary to approximate Q by a measure Q_r obtained by replacing each curve by a translation (with a modification at $t = 0$) and averaging according to the translation parameter. Thus, Q_r indeed satisfies (5), and by taking the limit as $r \rightarrow 0$, we obtain that \bar{Q} is a Nash equilibrium. Our proof was constructed particularly for the case where the trajectories are piecewise constant, as the proof contained in [CMS16] and inspired by the works of Ambrosio and Figalli [AF09] for the incompressible Euler equations does not work here. Indeed, the difficulty lies in the non-separability of the BV space and the non-continuity of this functional with strong convergence. Once the new approach is established for the case with discontinuities, we can also retrieve a known result for *classical* mean field games with a new proof. This requires an assumption on the exponents (i.e., $q < 1 + \frac{p'}{d}$ where f has polynomial growth of exponent at most q and p' is the conjugate exponent of p), but this assumption is not new. It appears particularly in many articles at the beginning of the analysis of first-order variational MFGs (for example [CG15]). The approach to show that \bar{Q} is indeed a Nash equilibrium requires regularity properties (continuity and boundedness of the density), but these properties were obtained thanks to the Eulerian formulation of the problem.

These regularities are gathered in Chapter 4 and also appear in the article [DS23]. We begin by studying the regularity of the solution $\gamma: [0, 1] \rightarrow \mathbb{R}^d$ in a finite-dimensional space of

$$\min \left\{ TV(\gamma; [0, 1]) + \int_0^1 F(t, \gamma(t)) dt + \psi_0(\gamma(0)) + \psi_1(\gamma(1)) : \gamma \in BV([0, 1]; \mathbb{R}^d) \right\}. \quad (6)$$

By approaching this problem with

$$\min \left\{ \int_0^1 (L_\varepsilon(\gamma'(t)) + F(t, \gamma(t))) dt + \alpha_\varepsilon(\psi_0(\gamma(0)) + \psi_1(\gamma(1))) : \gamma \in H^1([0, 1]) \right\},$$

using the Euler-Lagrange system and maximizing an appropriate function, we obtain a uniform bound on the derivative of the solution

$$|\gamma'_\varepsilon| \leq \max \left\{ \frac{C_0}{c_0}, \frac{\alpha_\varepsilon \varepsilon}{\sqrt{1 - \alpha_\varepsilon^2}} \right\}. \quad (7)$$

By applying the Arzelà-Ascoli theorem and the semi-continuity of the total variation TV , the sequence γ_ε converges up to a subsequence towards the solution of (6). Moreover, by passing to the limit in inequality (7), the solution satisfies

$$|\gamma'| \leq C,$$

which is another way of saying that the solution is Lipschitz in time. Furthermore, when $d = 1$, the solution is either constant or a critical point of F . This gives us an idea of the shape of the curve. Next, we show a similar regularity for the solution of the problem

$$\min_{\substack{\rho: [0, T] \rightarrow L^1(\Omega) \\ \rho \geq 0 \\ \forall t \in [0, T], \int_\Omega \rho(t, x) dx = 1}} \int_0^T \int_\Omega (|\dot{\rho}(t, x)| + V(t, x)\rho(t, x) + f(\rho(t, x))) dx dt + \psi_0(\rho(0)) + \psi_T(\rho(T)), \quad (8)$$

where this time ρ is a curve valued in an infinite-dimensional space. We approach ρ by a curve in a finite-dimensional space $\rho_n: [0, T] \rightarrow \mathbb{R}^n$, and then use similar techniques to obtain

$$\sup_{t \in [0, T]} \int_\Omega |\dot{\rho}(t, x)|^2 dx \leq C,$$

in other words, $\rho:]0, T[\rightarrow L^2(\Omega)$ is Lipschitz. With additional adjustments, we also have that the solution of

$$\min_{\substack{\rho: [0, +\infty[\rightarrow L^1(\Omega) \\ \rho \geq 0 \\ \forall t \in \mathbb{R}_+, \int_\Omega \rho(t, x) dx = 1}} \int_0^\infty e^{-rt} \int_\Omega (|\dot{\rho}(t, x)| + V(t, x)\rho(t, x) + f(\rho(t, x))) dx dt + \psi_0(\rho(0))$$

is Lipschitz in time with values in $L^2(\Omega)$. When the problem has Dirichlet boundary conditions in time, it is possible to show that the solution ρ has the same regularity as the data (i.e., the boundary conditions and the function V) in space almost everywhere.

Numerical simulations of the solution to (8) are compiled in Chapter 5. Firstly, we introduce various existing optimization algorithms, concluding with the algorithm we employ, namely the fast dual proximal gradient method. It is well-known that the sequence generated by this algorithm converges at a rate of $O(1/k)$ to the solution of the problem (cf. [BT14]). For different examples of V and various boundary conditions on $[0, T]$, we perform simulations of the solution to (8). These numerical results are detailed in [DS23, Section 6]. In the case of the mixed problem, where the solution $\bar{\rho}$ minimizes

$$\min_{\substack{\rho \in L^2([0, T] \times \Omega) \\ \rho \geq 0 \\ \forall t \in [0, T], \int_{\Omega} \rho(t, x) dx = 1}} \mathcal{V}_{\bar{\rho}}(\rho) = \int_0^T \int_{\Omega} \left(\lambda |\dot{\rho}(t, x)| + V(\bar{\rho})(t, x) \rho(t, x) + c_0 \frac{\rho(t, x)^2}{2} \right) dx dt + \psi_0(\rho(0)) + \psi_T(\rho(T))$$

being a fixed point of an operator, it is possible to conduct numerical simulations of the latter using the Banach's fixed point theorem, which provides a means to compute it. Note that we select examples of $V(\bar{\rho})$ that are non-variational and depend on $\bar{\rho}$ in a non-autonomous manner (to ensure the solution is not constant over time). We also consider an example dependent on the entire history, i.e., $V(\rho)(t, x) \equiv \int_0^t \rho(s, x) dx$, which may be of interest for modeling purposes. These examples are presented in [DS24, Section 6].

The final Chapter 6 of this thesis presents two applications of the deterministic Mean Field Games model with jumps in the case of multiple populations. The first example considers a model with two populations where individuals from one group are less inclined to live near individuals from the other group. Using the mixed formulation of the problem,

$$\mathcal{G}: L^2([0, T] \times \Omega) \times L^2([0, T] \times \Omega) \rightarrow L^2([0, T] \times \Omega) \times L^2([0, T] \times \Omega)$$

$$(\bar{\rho}_1, \bar{\rho}_2) \mapsto \left(\underset{\substack{\rho_1 \geq 0 \\ \forall t, \int \rho_1(t) = \frac{1}{2}}}{\operatorname{argmin}} \mathcal{V}_{\bar{\rho}_1, \bar{\rho}_2}^1(\rho_1); \underset{\substack{\rho_2 \geq 0 \\ \forall t, \int \rho_2(t) = \frac{1}{2}}}{\operatorname{argmin}} \mathcal{V}_{\bar{\rho}_1, \bar{\rho}_2}^2(\rho_2) \right)$$

with

$$\mathcal{V}_{\bar{\rho}}^1(\rho) := \int_0^T \int_{\Omega} \lambda |\dot{\rho}| + V_1(\bar{\rho}) \rho + f(\rho) dx dt + \psi_0(\rho(0)) + \psi_T(\rho(T))$$

$$\text{and } \mathcal{V}_{\bar{\rho}}^2(\rho) := \int_0^T \int_{\Omega} \lambda |\dot{\rho}| + V_2(\bar{\rho}) \rho + f(\rho) dx dt + \psi_0(\rho(0)) + \psi_T(\rho(T)),$$

we obtain numerical simulations of the solution using the Banach's fixed point theorem. The second example is a model inspired by Carlier and Ekeland [CE07] on urban planning involving workers and firms. In their work, they interpret worker wages φ and firm payments ψ as Kantorovich potentials, where transportation costs correspond to commuting expenses. In other words, the transportation problem is formulated as

$$\inf_{\gamma \in \Pi(\mu, \nu)} \int_{\bar{\Omega} \times \bar{\Omega}} c(x, y) d\gamma(x, y),$$

with μ representing worker density and ν representing firm density, which is equivalent to

$$\sup_{\varphi, \psi} \left\{ \int_{\bar{\Omega}} \psi d\nu - \int_{\bar{\Omega}} \varphi d\mu ; \psi(y) - \varphi(x) \leq c(x, y), \forall (x, y) \in \bar{\Omega} \times \bar{\Omega} \right\}.$$

Each worker aims to minimize

$$\min_{\gamma(0)=x_0} 2\lambda_1 S(\gamma) + \int_0^T [-\varphi_{\mu, \nu}(\gamma(t)) + V_{0,1}(t, \gamma(t)) + f'(\mu)(t, \gamma(t))] dt + \phi_T(\gamma(T))$$

and each firm aims to minimize

$$\min_{\gamma(0)=x_0} 2\lambda_2 S(\gamma) + \int_0^T [\psi_{\mu, \nu}(\gamma(t)) + V_{0,2}(t, \gamma(t)) + f'(\nu)(t, \gamma(t))] dt + \phi_T(\gamma(T)).$$

A Nash equilibrium (μ, ν) satisfies the following Eulerian problem

$$\min_{\substack{\mu, \nu \geq 0 \\ \int \mu_t = \int \nu_t = \frac{1}{2}}} \int_0^T \int_{\Omega} \lambda_1 |\dot{\mu}| + \lambda_2 |\dot{\nu}| + V_{0,1} \mu + V_{0,2} \nu + f(\nu) + f(\mu) dx dt + \int_0^T W_c(\mu_t, \nu_t) dt$$

$$+ \Phi_0(\mu(0)) + \Phi_0(\nu(0)) + \Phi_T(\mu(T)) + \Phi_T(\nu(T)),$$

where $W_c(\mu, \nu) = \inf_{\gamma \in \Pi(\mu, \nu)} \int_{\bar{\Omega} \times \bar{\Omega}} c(x, y) d\gamma(x, y)$. To compute Kantorovich potentials effectively, the Sinkhorn algorithm is used to regularize the Wasserstein distance

$$\begin{aligned} W_{c, \varepsilon}(\mu, \nu) &= \inf_{\gamma \in \Pi(\mu, \nu)} \int_{\Omega \times \Omega} c(x, y) d\gamma(x, y) - \varepsilon \int \gamma(\log \gamma - 1) \\ &= \sup_{\varphi, \psi} \int_{\Omega} \psi(y) d\nu(y) - \int_{\Omega} \varphi(x) d\mu(x) - \varepsilon \int_{\Omega \times \Omega} e^{\frac{\psi(y) - \varphi(x) - c(x, y)}{\varepsilon}} d(\mu \otimes \nu)(x, y), \end{aligned}$$

which also relaxes the constraint $\psi(y) - \varphi(x) \leq c(x, y)$. We propose three different numerical approaches to solving this problem. The simplest approach ignores its variational structure and applies a fixed-point theorem similar to the mixed problem with the function \mathcal{G} , defining

$$\begin{aligned} \mathcal{V}_{\bar{\mu}, \bar{\nu}}^1(\mu) &:= \int_0^T \int_{\Omega} \lambda_1 |\dot{\mu}| - \varphi_{\bar{\mu}, \bar{\nu}} \mu + f(\mu) \, dx dt + \Phi_0(\mu(0)) + \Phi_T(\mu(T)) \\ \text{and } \mathcal{V}_{\bar{\mu}, \bar{\nu}}^2(\nu) &:= \int_0^T \int_{\Omega} \lambda_2 |\dot{\nu}| + \psi_{\bar{\mu}, \bar{\nu}} \nu + f(\nu) \, dx dt + \Phi_0(\nu(0)) + \Phi_T(\nu(T)). \end{aligned}$$

Attempts at numerical simulations have been made, but the results are not yet conclusive and are not included in this thesis. The second approach uses the variational structure to apply the dual proximal gradient method in two different scenarios: either $W_{c, \varepsilon}$ is seen as a perturbation of the uniformly convex function or as a perturbation of the non-smooth convex function. The third approach also utilizes the variational formulation by writing it as

$$\inf_{\mu, \nu} \sup_{\varphi, \psi} \mathcal{L}(\mu, \nu, \varphi, \psi),$$

which allows the application of the Uzawa algorithm, thereby avoiding the computation of Kantorovich potentials with the Sinkhorn algorithm. Due to time constraints, these paths have not yet been fully explored, but they deserve further investigation.

Chapter 1

Mean Field Games

1.1 Classical Mean Field Games

The theory of Mean Field Games (MFG for short) has been simultaneously introduced by Lasry and Lions in [LL06a], [LL06b] and [LL07] and by Huang, Malhamé and Caines [HMC06] in 2006.

In a game with N players, we consider that each player is rational, which means that they choose the best strategy to maximize or minimize a certain cost that depends on the strategy chosen by the other players. Hence, they try to anticipate the choices of the other players. The strategy of each player i can be described by a trajectory $\gamma_i: [0, T] \rightarrow \mathbb{R}^n$ such that $\gamma_i'(t) = \alpha_i(t)$ and $\gamma_i(0) = x_0$, where α_i is a control chosen by the player in order to minimize a cost

$$J^i(\alpha_1, \dots, \alpha_N) = \int_0^T L(t, \gamma_i(t), \alpha_i(t)) + f(\rho_N(t)) dt$$

which depends on $\rho_N(t) = \frac{1}{N-1} \sum_{j \neq i} \delta_{\gamma_j(t)}$, the distribution of the other agents produced by the controls α_j with $j \neq i$.

A *Nash equilibrium* is a N -tuple $(\bar{\alpha}^1, \dots, \bar{\alpha}^N)$ such that for all i ,

$$\forall \alpha^i, J^i(\bar{\alpha}^1, \dots, \bar{\alpha}^N) \leq J^i(\bar{\alpha}^1, \dots, \bar{\alpha}^{i-1}, \alpha^i, \bar{\alpha}^{i+1}, \dots, \bar{\alpha}^N).$$

This notion was introduced by John Nash [Nas51] in 1951, and now we know that such an equilibrium always exists in a N -player game with mixed strategies, by using the Brouwer fixed point theorem.

The idea of Lasry and Lions was to extend the Nash equilibrium in a game with an infinite number of players by taking $N \rightarrow \infty$. This means that ρ_N should converge to a probability measure ρ representing an infinite number of players. This density ρ coupled with a function u should verify a system containing the Hamilton-Jacobi-Bellman equation coupled with the continuity equation. The function u is the value function of a control problem corresponding to the minimum of J . The MFG system is as follows:

$$\begin{cases} -\partial_t u(t, x) + H(t, x, -\nabla_x u) & = f[\rho(t)](x) \\ \partial_t \rho(t, x) + \operatorname{div}(\rho \nabla_p H(t, x, -\nabla_x u)) & = 0 \\ \rho(0) & = \rho_0 \\ u(T) & = u_T \end{cases} \quad (1.1)$$

where $H(t, x, p) = \sup_q -L(t, x, q) + q \cdot p$ in some cases. Thanks to the Schauder fixed point theorem, the solution (u, ρ) to (1.1) exists (see Section 4 in the notes by Cardaliaguet [Car12]).

Some noise can be added in the trajectories γ_i by using a Brownian motion. The agents would then minimize

$$J^i(\alpha_1, \dots, \alpha_N) = \mathbb{E} \int_0^T L(t, \gamma_i(t), \alpha_i(t)) + f(\rho_N(t)) dt$$

and the MFG system (1.1) would be of second order by Ito's formula. This will be referred to as the *stochastic case* while here I introduced the *deterministic case* which is of interest in this thesis. Originally, Mean Field Games were first introduced in the stochastic case by Lions and then were studied in the deterministic case using vanishing viscosity.

A difficult question is whether the limit of such a N -player game is always of the form of a MFG system (1.1), and conversely whether the solution to the system (1.1) can provide an approximation of a N -player game. A breakthrough has been done to prove the mean field limit by Cardaliaguet, Delarue, Lasry and Lions [CDLL19] by using the Master Equation in the case where the players undergo a common noise. Carmona and Delarue [CD18, Chapter 6] explore also both of these questions in their book.

To explain the system (1.1), we need a presentation of the main ideas in optimal control.

Let us denote $y_{x,\alpha}$ the solution to the ODE

$$\begin{cases} y'(t) &= g(y(t), \alpha(t)) \\ y(0) &= x. \end{cases}$$

Usually, the function α is called the *control*. We denote $\phi(x)$ the value of the infimum

$$\phi(x) := \inf_{\alpha} \int_0^T L(s, y_{x,\alpha}(s), \alpha(s)) ds$$

and the *value function*

$$\phi(t, x) := \inf_{\alpha} \left\{ \int_t^T L(s, y_{x,\alpha}(s), \alpha(s)) ds \mid y_{x,\alpha}(t) = x \right\}.$$

Then, the *Dynamic Programming Principle* claims that if α^* is optimal for $\phi(x)$ on $[0, T]$, then it is also optimal on $[t, T]$, i.e. α^* is optimal for the problem

$$\phi(t, y_{x,\alpha^*}(t)) = \inf_{\alpha} \left\{ \int_t^T L(s, y_{x,\alpha}(s), \alpha(s)) ds \mid y_{x,\alpha}(t) = y_{x,\alpha^*}(t) \right\}.$$

As a consequence, $\phi(x)$ can be rewritten as a solution to a Bellman equation

$$\phi(x) = \inf_{\alpha} \left\{ \int_0^t L(s, y_{x,\alpha}(s), \alpha(s)) ds + \phi(t, y_{x,\alpha}(t)) \right\}.$$

Further computations can prove that $\phi(t, x)$ is a viscosity solution of the Hamilton-Jacobi-Bellman equation

$$-\partial_t \phi(t, x) + H(t, x, -\nabla_x \phi(t, x)) = 0, \quad (1.2)$$

where H is the Hamiltonian defined by

$$H(t, x, p) = \sup_{q \in \mathbb{R}^n} -L(t, x, q) + p \cdot g(x, q).$$

I refer to [FR75] for rigorous proofs about optimal control. The lecture notes by Guillaume Carlier [Car07] with heuristic computations were a good help to understand dynamic programming principle. A very short introduction to optimal control has also been written in the second book by Filippo Santambrogio [San23, Section 1.7].

The heuristic proof of (1.2) consists in showing it by double inequality. The direction \leq is established for all $q \in \mathbb{R}^n$ in $-L(t, x, q) + p \cdot g(x, q)$ with $p = -\nabla_x \phi$ and then one takes the supremum over q . For the other direction \geq , one has to take the optimal control α for the problem ϕ . The inequality is proved and it happens that at the same time this optimal control α is a maximiser of H , i.e. $\alpha \in \operatorname{argmax}_q -L(t, x, q) + p \cdot g(x, q)$ with $p = -\nabla_x \phi$. *A posteriori*, this furnishes a way to compute the optimal control. The connection between the optimiser of H and the optimal control is established in [BC97, Corollary 2.18].

Now, we can link optimal control to MFG. The only difference is that the MFG system (1.1) is coupled with the continuity equation in a way that the right-hand side of the Hamilton-Jacobi equation $f(x, \rho(t, x))$ depends on the density. If ρ is given, then u is the value function

$$u(t, x) = \inf_{\alpha} \int_t^T L(s, y_{x,\alpha}(s), \alpha(s)) + f(y_{x,\alpha}(s), \rho(s, y_{x,\alpha}(s))) ds \quad (1.3)$$

and by the previous discussion, u verifies the first equation in the MFG system. To recover $g(x, \alpha)$, we use the fact that the optimal control α is optimal for H . Then we use the *envelope theorem* (reminded in Box 1.11 [San23]) which claims that if $(h_a)_a$ is a family of functions such that for all p , $h(p) = \sup_a h_a(p)$ and if h is differentiable at p_0 such that $h(p_0) = h_{a_0}(p_0)$ for some a_0 , then $\nabla_p h(p_0) = \nabla_p h_{a_0}(p_0)$. Thanks to this and to the fact that α is optimal for H , we obtain that

$$\nabla_p H(t, x, p) = \nabla_p [-L(t, x, \alpha) + p \cdot g(x, \alpha)] = g(x, \alpha),$$

with $p = -\nabla_x u$. Hence, the optimal trajectory $y_{x,\alpha}$ verifies the ODE

$$\begin{cases} y'(t) &= \nabla_p H(t, y(t), -\nabla_x u(t, y(t))) \\ y(0) &= x. \end{cases}$$

If we denote $v(t, x) = \nabla_p H(t, x, -\nabla_x u(t, x))$, then v can be seen as the velocity field of the particles $y(t)$. Therefore, by the *mass conservation equation* which arises in fluid mechanics, we obtain that the density ρ solves the continuity equation in the sense of distributions

$$\partial_t \rho + \operatorname{div}(\rho v) = 0 \text{ with } v = \nabla_p H(t, x, -\nabla_x u),$$

which is exactly the second equation in the MFG system (1.1).

The existence of the solution to (1.1) has been established when the coupling $f[\rho(t)](x)$ is *nonlocal* (see [LL06b] and [Car12]), which means that $f[\rho(t)](x) = f[\rho(t, \cdot)](x)$ can be for example a convolution of a function with $\rho(t, \cdot)$ which has a regularization effect. If one considers the application that with ρ associates the solution u to the Hamilton-Jacobi equation and the application that with u associates the solution $\bar{\rho}$ to the continuity, then by Schauder fixed point theorem there exists a fixed point to the composition of these two applications. This is possible thanks to the nonlocal property of f which allows the composition to be a continuous application.

If the coupling is *local*, namely f directly depends on $\rho(t, x)$ such that $f[\rho(t)](x) = f(x, \rho(t, x))$, then it was not clear whether the solution exists. This problem has been overcome later by Cardaliaguet [CG15] in 2015 using a variational formulation that will be introduced in the next section.

1.2 Variational Mean Field Games

In this section, we pass to the case where the coupling f is local, i.e. $f[\rho(t)](x) := f(x, \rho(t, x))$.

In [Car15] Cardaliaguet proves the existence of a weak solution to (1.1) by describing it as the minimizer of two variational problems that are in duality.

The first one essentially comes from the weak formulation of the continuity equation and is

$$\inf_{\phi} \mathcal{A}(\phi) = \int_0^T \int_{\mathbb{T}^d} F^*(x, -\partial_t \phi(t, x) + H(x, -\nabla_x \phi(t, x))) dx dt - \int_{\mathbb{T}^d} \phi(0, x) dm_0(x)$$

where F is the antiderivation of f in the sense that $dF/dm = f$ and ϕ should play the role of the value function. The function F^* is the Fenchel conjugate of F .

The second one stems from the minimization problem $u(0, x)$ in (1.3) and is

$$\inf_{\substack{(\rho, v) \\ \partial_t \rho + \operatorname{div}(\rho v) = 0}} \mathcal{B}(\rho, v) = \int_0^T \int_{\mathbb{T}^d} \rho(t, x) H^*(x, -v(t, x)) + F(x, \rho(t, x)) dx dt + \int_{\mathbb{T}^d} \phi_T(x) \rho(T, x) dx \quad (1.4)$$

where ρ is the density and v the velocity. The function H^* is also the Fenchel conjugate of H .

By the Fenchel-Rockafellar duality theorem (see [ET99]), we have that

$$\inf_{\phi} \mathcal{A}(\phi) = - \min_{\substack{(\rho, w) \\ \partial_t \rho + \operatorname{div}(w) = 0}} \mathcal{B}(\rho, w), \quad (1.5)$$

where we use the change of variable $w := \rho v$ to make \mathcal{B} convex in its second variable.

Let us note that this formulation of variational MFG was the first to occur in the literature. From a modeling point of view the problem $\min \mathcal{B}$ is the most meaningful, but since it is set in the setting of measures it is itself the dual of $\min \mathcal{A}$, and not vice versa. A similar game with potential has also been studied in the case of Wardrop equilibria [CJS08] which was introduced in the 1950s by Wardrop [War52] to study a large number of vehicles.

In [BCS17], the authors explain formally how we pass from the MFG system to these dual problems in the simple case where $L(t, \gamma(t), \dot{\gamma}(t)) = |\dot{\gamma}(t)|^2/2$ and the control is $g(x, \alpha) = \alpha$. By using the formulation (1.5) of the problem, existence and uniqueness of weak solutions of the MFG system has been rigorously proved in [Car15] when H and f has a superlinear growth, in the case of a first order MFG with a local coupling. Under slightly less restrictions on H and f , existence and uniqueness are proved in [CG15]. Thanks to this variational approach, there is also existence and uniqueness of weak solutions to the second order MFG system with degenerate diffusion and local coupling in [CGPT15].

A particular case of the problem (1.4) has been reformulated by Benamou and Brenier in [BB00] into an optimal transport problem at the beginning of the 2000s before the introduction of Mean Field Games. The authors showed that the dynamical formulation of the optimal transport problem is actually the minimizer of some kinetic energy subjected to the continuity equation:

$$W_2(\rho_0, \rho_T)^2 = \inf_{\substack{(\rho, v) \\ \partial_t \rho + \nabla \cdot (\rho v) = 0 \\ \rho(0) = \rho_0, \rho(T) = \rho_T}} T \int_{\mathbb{R}^d} \int_0^T \rho(t, x) |v(t, x)|^2 dx dt. \quad (1.6)$$

Here the density is not coupled with the Hamilton-Jacobi equation since $f = 0$, so the function f does not really depend on the density in (1.1), so this is not a true MFG problem. Nevertheless, this result allows to use dynamical formulation of optimal transport in MFG. This link is described in [BCS17]. The result (1.6) has been extended to $p > 1$ instead of $p = 2$ in [Bre03]. In [BJO09], Buttazzo, Jimenez and Oudet generalize the problem in the right-hand side of (1.6) to a more general cost function which enables to add a congestion term to the kinetic energy $\iint \rho|v|^2$. Without knowing it, the authors studied a variational Mean Field Game whose formulation is described later in (1.8).

Without being exhaustive, several studies have been done thanks to this approach in Mean Field Games. For example, the regularity of the solution is studied via the variational structure of the problem. In the paper by Lavenant and Santambrogio [LS18], the authors prove the L^∞ regularity of the solution thanks to the JKO scheme. For a particular case of the Hamiltonian, i.e. $H(x, \xi) = \frac{1}{2}|\xi|^2$, Prosinski and Santambrogio [PS17] show Sobolev regularity of the solution by using the duality of the problem. This result has been improved by Graber and Mészáros [GM18] for more general H .

If H is non-standard and if f , whose increasing property usually describes the congestion in the models, is decreasing, then Cirant and Nurbekyan [CN18] showed the existence of the solution to viscous MFG by using a suitable variational formulation of the problem.

In order to perform numerical simulations of variational MFG problems, the Alternating Direction Method of Multipliers (ADMM) can be used. It was introduced in 1983 by Fortin and Glowinski [FG83]. They give two different versions which they call ALG1 and ALG2. Since ALG2 is cheaper than the second, it is applied in [BC15], [AL16], [BCS17] and [And17] to model Variational Mean Field Games. The authors Briceños-Arias, Kalise and Silva compare different proximal methods in [BKS18] such as ADMM, Chambolle-Pock, Predictor-Corrector Proximal Multiplier and Monotone+Skew to model stationary MFG.

Up to now, the variational formulation has been presented in an *Eulerian* point of view where we consider the density $\rho(t)$ as a curve of measures. Now we introduce the *Lagrangian* formulation where the density is represented by a probability measure $Q \in \mathcal{P}(\mathcal{C})$ supported on the set of curves taken by each agent.

Formally, if we think about dividing Eq. (1.6) by T^2 and letting T tend to 0 and if we define the *metric derivative* by

$$|\rho'| (t) = \lim_{s \rightarrow t} \frac{W_p(\rho(s), \rho(t))}{|s - t|}, \quad (1.7)$$

then we have the following characterization of absolutely continuous curves in the set of probability measures equipped with the Wasserstein distance W_p whose statement and proof can be found in [AGS05, Theorem 8.3.1] or in [San15, Theorem 5.14]:

Theorem 1.2.1. *Let $p > 1$, $\mathcal{P}_p(\Omega)$ be a probability space equipped with the Wasserstein distance W_p and I be an open interval of \mathbb{R} .*

If $\rho: I \rightarrow \mathcal{P}_p(\Omega)$ is an absolutely continuous curve, then there exists a vector field v such that $\partial_t \rho + \nabla \cdot (\rho v) = 0$ in the sense of distributions and

$$\|v(t)\|_{L^p(\rho_t)} \leq |\rho'| (t), \text{ a.e. } t.$$

Conversely, if ρ satisfies $\partial_t \rho + \nabla \cdot (\rho v) = 0$ in the sense of distributions for some v , then ρ is an absolutely continuous curve and

$$|\rho'| (t) \leq \|v(t)\|_{L^p(\rho_t)}, \text{ a.e. } t.$$

Theorem 1.2.1 stems from a generalisation of (1.6) for $p > 1$. Thanks to the next Proposition, we will be able to write a Lagrangian version of Theorem 1.2.1.

Proposition 1.2.2 ([San15], Proposition 5.31). *For every Lipschitz curve $(\rho_t)_t$ in $\mathcal{P}_p(\Omega)$ endowed with the Wasserstein distance W_p , there exists a measure $Q \in \mathcal{P}(\mathcal{C})$ such that $\rho_t = e_t \# Q$ and*

$$\int_{\mathcal{C}} \int_0^1 |\gamma'(t)|^p dt dQ(\gamma) \leq \int_0^1 |\rho'| (t)^p dt.$$

Conversely, if $Q \in \mathcal{P}(\mathcal{C})$ is such that $\rho(t) = e_t \# Q$, then

$$\int_0^1 |\rho'| (t)^p dt \leq \int_{\mathcal{C}} \int_0^1 |\gamma'(t)|^p dt dQ(\gamma).$$

Proposition 1.2.2 allows us to rewrite Theorem 1.2.1 in a Lagrangian way:

Proposition 1.2.3 ([BCS17], Proposition 3.1). *If $Q \in \mathcal{P}(\mathcal{C})$ is a probability measure such that $\rho_t = e_t \# Q$ and $\int_{\mathcal{C}} \int_0^T \frac{|\gamma'(t)|^p}{p} dt dQ(\gamma) < \infty$, then $\rho: [0, T] \rightarrow \mathcal{P}_p(\Omega)$ is absolutely continuous and there exists a vector field v such that the continuity equation is satisfied and*

$$\frac{1}{p} \int_0^T \int_{\Omega} \rho |v|^p \leq \int_{\mathcal{C}} \int_0^T \frac{|\gamma'(t)|^p}{p} dt dQ(\gamma).$$

Conversely, if ρ is absolutely continuous, (ρ, v) satisfies the continuity equation and $\int_0^T \int_{\Omega} \rho |v|^p < \infty$, then $|\rho'(t)| \leq \|v(t)\|_{L^p(\rho_t)}$ for almost every t and there exists $Q \in \mathcal{P}(\mathcal{C})$ such that $\rho_t = e_t \# Q$ and

$$\int_{\mathcal{C}} \int_0^T \frac{|\gamma'(t)|^p}{p} dt dQ(\gamma) \leq \frac{1}{p} \int_0^T \int_{\Omega} \rho |v|^p.$$

Therefore, the minimization problem

$$\min_{\substack{(\rho, v) \\ \partial_t \rho + \nabla \cdot (\rho v) = 0}} \left\{ \int_0^T \int_{\Omega} \left(\frac{1}{p} \rho |v|^p + G(x, \rho_t) \right) dx dt + \int \Psi d\rho_T \right\} \quad (1.8)$$

can be written as:

$$\min_{Q \in \mathcal{P}(\mathcal{C}), e_0 \# Q = \rho_0} \left\{ \int_{\mathcal{C}} \int_0^T \frac{|\gamma'(t)|^p}{p} dt dQ(\gamma) + \int_0^T \mathcal{G}(e_t \# Q) dt + \int_{\Omega} \Psi d(e_T \# Q) \right\}, \quad (1.9)$$

where $\mathcal{G}: \mathcal{P}(\Omega) \rightarrow \bar{\mathbb{R}}$ is such that

$$\mathcal{G}(\rho) = \begin{cases} \int G(x, \rho(x)) dx, & \text{if } \rho \ll \mathcal{L}^d, \\ +\infty & \text{otherwise.} \end{cases}$$

The link between the problem (1.9) and MFG is that if \bar{Q} minimizes (1.9), then we have for all Q such that $\int \mathcal{G}(e_t \# Q) dt < +\infty$,

$$\begin{aligned} & \int_{\mathcal{C}} \left(\int_0^T \frac{|\gamma'(t)|^p}{p} dt + \int_0^T G(\gamma(t), e_t \# \bar{Q}) dt + \Psi(\gamma(T)) \right) d\bar{Q}(\gamma) \\ & \leq \int_{\mathcal{C}} \left(\int_0^T \frac{|\gamma'(t)|^p}{p} dt + \int_0^T G(\gamma(t), e_t \# Q) dt + \Psi(\gamma(T)) \right) dQ(\gamma). \end{aligned}$$

Further work shows that this inequality is true for all Q , thus proving that \bar{Q} is supported on the optimal trajectories γ . I refer to [BCS17] and [San20] for a more rigorous presentation.

The Lagrangian point of view was used in a MFG model with density constraints studied by Cardaliaguet, Mészáros and Santambrogio [CMS16], where the agents going through a congested area have to pay an extra price. They showed that the solution the MFG system generates a probability measure that is supported on the optimal trajectories. By using the Eulerian formulation, they proved $L^2_{t,loc} BV_x$ regularity of the pressure. Under more restrictive hypotheses on H , Lavenant and Santambrogio [LS19] proved better regularity that is $L^\infty_t H^1_x$.

The problem where the conditions at time 0 and at time T are prescribed (i.e we impose the density to verify $\rho(0) = \rho_0$ and $\rho(T) = \rho_T$) is called the *planning problem*. In [OPS19], the Eulerian variational structure is used to prove existence and uniqueness of a weak solution for the first order MFG where H is convex with a quadratic growth. In addition, they show that there exists a probability measure supported on the optimal paths. Still in the Eulerian framework, Graber, Mészáros, Silva and Tonon [GMST19] improve this existence and uniqueness for more general H . Moreover, they obtain Sobolev regularity of the solution under some restrictions. In [BFG22], the authors propose a different variational approach of the planning problem in MFG which eliminates the continuity equation in the variational problem. They also prove existence and uniqueness of the solution.

To modelize crowd motion, Mazanti and Santambrogio [MS19] use a minimal-time control problem where agents seek to exit a bounded domain as fast as possible, but are subjected to bounded maximal speed in crowded areas. By using the Lagrangian formulation of the problem, they prove existence of an equilibria thanks to Kakutani's fixed point theorem.

The Lagrangian setting can also be used to modelize crowd motion where the velocity of an agent depends on the velocity of the other's. One can think of a group of birds which moves by following each other. The existence and the regularity properties of such a model has been investigated by Santambrogio

and Shim [SS21]. The problem is said to be *Cucker-Smale inspired*. The variational problem is the minimization of a cost that depends on the kinetic energy and a convolution term that depends on an interaction kernel. Sadeghi Arjmand in his thesis [Sad22] improves this result in the free final time case with a general running cost L and a more general interaction cost.

As far as numerics are concerned, we note that in [Sar22], Sarrazin uses the Lagrangian formulation to approximate the solution of a variational MFG with a variational MFG of finite number of players.

1.3 Mean Field Games with jumps

In this thesis, we concentrate on deterministic Mean Field Games with jumps, which means that we allow the trajectory of an agent γ to be discontinuous. More precisely, we are motivated by the modeling of the residents in a city. Each inhabitant will choose a place to live according to their preferences (e.g. city center, suburbs, ...) and to the density of the population. The move of a person will count as a jump. Between one jump and the next one, the trajectory is constant. The model we introduce is only a toy model and our aim is not to fit any real experimental data, but to study the basic properties of the model such as existence and uniqueness of solutions, regularity properties, numerical simulations, etc.

Mean Field Games with jumps have already been studied by Bertucci in his thesis in 2018 [Ber18] in the stochastic case where the agent's trajectories are controlled by a Brownian motion and are allowed to make jumps. In particular, he extends impulse control problems to Mean Field Games by making the value function depend on the density. Impulse control theory has been brought to the community's attention in the 1970s by Bensoussan and J.-L. Lions in [BL75] who introduced new methods such as Quasi-Variational Inequalities (QVI) to solve these kind of problems.

The motivations for the study of impulse control problems is for example stock management. To present an impulse control problem, I borrow the notations from [BL75]:

- $y(t) \in \mathbb{R}^n$ is the state of the stock at time $t \in [0, T]$ of n different types of goods,
- $D(s, t) = (t - s)\mu + \sigma(W(t) - W(s))$ is the demand between times s and t where $\mu \in \mathbb{R}^n$ and $W(t) \in \mathbb{R}^m$ is a standard Wiener process and σ is a linear application from \mathbb{R}^m to \mathbb{R}^n ,
- A is a matrix such that $A_{i,j} = (\sigma\sigma^*)_{i,j}$,
- The set of random variables $v = \{\theta^1, w^1, \theta^2, w^2, \dots\}$ is the policy of commands to replenish the stock when knowing the state $y_{t,x}$, where $0 \leq \theta^1 \leq \dots \leq \theta^i \leq \theta^{i+1} \leq T$ are the times at which the command is performed and w^1, w^2, \dots in \mathbb{R}^n are the quantity of goods that are ordered at respectively times $\theta^1, \theta^2, \dots$

When there is no jump, the stock y evolves simply according to the demand, i.e. for all $s \in [t, \theta^1[$,

$$y(0) = x \in \mathbb{R}^n \text{ and } y(s) = x - D(0, s).$$

Then, when a jump occurs, we add new stock to y :

$$y(\theta^1) = x - D(t, \theta^1) + w^1.$$

Knowing $y(\theta^i)$, the next part of the trajectory is defined as follows

$$\begin{aligned} y(s) &= y(\theta^i) - D(\theta^i, s), \forall s \in [\theta^i, \theta^{i+1}[\\ y(\theta^{i+1}) &= y(\theta^i) - D(\theta^i, \theta^{i+1}) + w^i. \end{aligned}$$

To choose the best control $v = \{\theta^1, w^1, \theta^2, w^2, \dots\}$, one would like to minimize a cost that depends on the number of jumps and on the state of stock. We define a function $f(t, x)$ that describes the cost paid for storing the goods or for stock shortage. The integer N_t describes the number of instants θ_i starting from t , i.e. the number of jumps. The problem is then

$$u(t, x) = \inf_{\substack{v = \{\theta^1, w^1, \dots\} \\ y(t) = x}} J(v) := \mathbb{E} \left(N_t + \int_t^T f(s, y(s)) ds \right). \quad (1.10)$$

The dynamic programming principle (already introduced in the first section of this chapter) provides another formulation of u . At time t , if it is optimal to place an order $\xi \in \mathbb{R}^n$, then

$$u(t, x) = 1 + u(x + \xi, t)$$

and if it exists, one can chose ξ such that it optimizes

$$1 + u(x + \xi, t) = 1 + \inf_w u(x + w, t).$$

When it is optimal not to jump and to let the demand reduce the stock during a period of time δ , then

$$u(t, x) = \mathbb{E} \left(\int_t^{t+\delta} f(x - D(t, s), s) ds + u(t + \delta, x - D(t, t + \delta)) \right). \quad (1.11)$$

Therefore, the new formulation of u is

$$u(t, x) = \min \left\{ 1 + \inf_w u(x + w, t), \mathbb{E} \left(\int_t^{t+\delta} f(x - D(t, s), s) ds + u(t + \delta, x - D(t, t + \delta)) \right) \right\}. \quad (1.12)$$

By writing the Taylor expansion of first order of the right-hand side of (1.11), we obtain

$$\delta f(t, x) + u(t, x) + \delta \left(- \sum_i \mu_i \frac{\partial u}{\partial x_i} + \frac{1}{2} \sum_{i,j} A_{i,j} \frac{\partial^2 u}{\partial x_i \partial x_j} + \partial_t u(t, x) \right) + o(\delta).$$

By denoting \tilde{A} the operator such that

$$\tilde{A}u = \sum_i \mu_i \frac{\partial u}{\partial x_i} - \frac{1}{2} \sum_{i,j} A_{i,j} \frac{\partial^2 u}{\partial x_i \partial x_j},$$

the problem (1.12) is equivalent to the system of inequalities

$$\begin{cases} u(t, x) \leq 1 + \inf_w u(t, x + w), \\ -\partial_t u + \tilde{A}u - f \leq 0, \\ \left[\partial_t u(t, x) - \tilde{A}u(t, x) + f(t, x) \right] [u(t, x) - 1 - \inf_w u(t, x + w)] = 0. \end{cases} \quad (1.13)$$

The system (1.13) can be written in a weaker way in a system of quasi-variational inequalities (QVI): under some assumptions on f and u , we get

$$\begin{cases} u(t, x) \leq 1 + \inf_w u(t, x + w) \text{ a.e.}, \\ -\langle \partial_t u | v - u \rangle + a(u, v - u) - \langle f | v - u \rangle \geq 0, \text{ for all } v \text{ such that } v(t, x) \leq 1 + \inf_w u(t, x + w), \end{cases} \quad (1.14)$$

where a is a bilinear application linked to \tilde{A} .

It is possible to generalize the cost of jumps. Instead of 1 in the formula $1 + \inf_w u(t, x + w)$ we can put a cost $k(w)$ that depends on the place where it will jump so that we obtain

$$Mu(t, x) := \inf_w k(w) + u(t, x + w).$$

M is a frequent notation for this operator.

The QVI (1.14) can be extended to more general cost functions f that depend on the control α on the trajectories y such that

$$J(v, \alpha) = \mathbb{E} \left(N_0 + \int_0^T L(t, y(t), \alpha(t)) dt \right).$$

If we define $H(t, x, p) = \min_q L(t, x, q) - p \cdot q$, the QVI would be roughly of the form

$$\begin{cases} u \leq Mu \text{ a.e.}, \\ -\langle \partial_t u | v - u \rangle + a(u, v - u) - \langle H(t, x, \nabla u) | v - u \rangle \geq 0, \text{ for all } v \leq Mu. \end{cases} \quad (1.15)$$

This extension has been made for the first time in [BL77] by Bensoussan and J.-L. Lions and the case when H has a quadratic growth has been studied in [BFM82].

Several studies have been carried out since the 1980s on impulse control problems. To cite a few, Frehse and Mosco in [FM82] study the problem when there are irregular obstacles. The idea of obstacle is contained in the operator Mu since the jumps occurs when the trajectory $y(t)$ reaches the set of points defined by $\{x ; u(t, x) = Mu(t, x)\}$. Further investigation on the obstacle and the regularity of the solution u has been done by Perthame in [Per85].

All the cited works so far have been performed in the stochastic case, which means that the trajectories are subjected to a Brownian motion and the jumps occurs in a random way. It is possible to pass from the stochastic case to the deterministic case by defining an operator

$$\tilde{A}_\varepsilon = c + \sum_i \mu_i \frac{\partial u}{\partial x_i} - \varepsilon \frac{1}{2} \sum_{i,j} a_{i,j} \frac{\partial^2 u}{\partial x_i \partial x_j},$$

and letting $\varepsilon \rightarrow 0$. In [Men81], Menaldi showed that the solution u_ε of the following QVI

$$u_\varepsilon \leq Mu_\varepsilon \text{ and } \forall v \leq Mu_\varepsilon \text{ a.e. } a_\varepsilon(u_\varepsilon, v - u_\varepsilon) \geq \langle f | v - u_\varepsilon \rangle$$

converges towards the solution u of

$$u \leq Mu \text{ and } \tilde{A}_0 u \leq f. \quad (1.16)$$

In [Men82], the author describes the solution of (1.16) as the value function of an impulse control problem

$$u(x) = \inf_{\substack{v=(\theta_1, \xi_1, \dots) \\ y(0)=x}} J(v) := \sum_{i=0}^{\infty} k(\xi_i) e^{-\beta \theta_i} + \int_0^\tau f(y(s)) e^{-\beta s} ds$$

where τ is some optimal stopping time and $\beta > 0$ a sufficiently large coefficient.

Barles in [Bar85] extends this result to a more general cost function

$$u(x) = \inf_{\substack{v=(\theta_1, \xi_1, \dots), \alpha \\ y(0)=x}} J(v, \alpha) := \sum_{i=0}^{\infty} k(\xi_i) e^{-\beta \theta_i} + \int_0^{+\infty} f(y(s), \alpha(t)) e^{-\beta s} ds$$

which depends on the control α where the continuous part of the trajectory verifies $y'(t) = b(y(t), \alpha(t))$.

Another approach is to formulate the optimality condition as a maximum principle. Blaquière introduces this method in [Bla77] for finite time horizon and in [Bla85] for infinite time horizon. Rempala and Zabczyk [RZ88] suggest a simpler proof for the maximum principle.

Deterministic impulse control problems are also the subject of more recent research for example in [EBB10], where El Farouq, Barles and Bernhard study the uniqueness of the viscosity solution of an Isaacs QVI which comes from an impulse control minimax problem.

The impulse control problems introduced in this subsection only deal with individual trajectories while in Mean Field Games one would like to study the behavior of an infinite number of players. This is the problem Bertucci investigated in his thesis [Ber18].

First, Bertucci presents what would be the density if the trajectories were subject to a Brownian motion and the value function of the problem were

$$u(t, x) = \inf_{v=(\theta_1, \xi_1, \dots)} \mathbb{E} \left(\int_t^T f(s, y(s)) ds + \sum_{i=1}^{\#(\theta_i)_i} k(\xi_i) \right).$$

For any $\varepsilon > 0$, he defines the PDE

$$\partial_t m_\varepsilon(t, x) - \nu \Delta m_\varepsilon(t, x) + \frac{1}{\varepsilon} (\mathbb{1}_A m_\varepsilon)(t, x) - \frac{1}{\varepsilon} (\mathbb{1}_A m_\varepsilon)(t, x - \xi) = 0 \quad (1.17)$$

where $A \subset \mathbb{R}^n$ is the set at which the particle jump, the term $\frac{1}{\varepsilon} \mathbb{1}_A m_\varepsilon(t, x)$ represents the rate at which the particles jumps and $\frac{1}{\varepsilon} \mathbb{1}_A m_\varepsilon(t, x - \xi)$ stands for the rate of particles that arrive. For simplicity, $\xi \in \mathbb{R}^n$ is the unique possible jump, i.e. we jump from x to $x + \xi$. By using appropriate estimates on m_ε , he extracts a sequence which converges towards a solution m such that for all $v \leq Mv$ on A ,

$$\int_0^T \langle \partial_t u - \nu \Delta u | m \rangle - \int_{\mathbb{T}^n} u(0) m_0 \leq \int_0^T \langle \partial_t v - \nu \Delta v | m \rangle - \int_{\mathbb{T}^n} v(0) m_0$$

for any u such that $u = Mu$ on A . This inequality is called by the author the *Fokker-Planck equation of particles* which can be seen as the dual of a QVI. Then, he extends this result to a finite number of possible jumps. Nevertheless, it is possible to apply this to any possible jump.

Second, Bertucci proves the existence of a solution to the MFG of an impulse control problem. Now the running cost $f(m)$ depends on the density and the MFG system is of the form

$$\begin{cases} \max(-\partial_t u - \nu \Delta u - f(m), u - Mu) = 0, \\ \forall v \leq Mv, \int_0^T \langle -\partial_t(v - u) - \nu \Delta(v - u) | m \rangle - \int_{\mathbb{T}^n} (v - u)(0) m_0 \geq 0, \\ \int_0^T \int_{\mathbb{T}^n} (-\partial_t u - \nu \Delta u - f(m)) m = 0. \end{cases}$$

The last equation can be viewed as a relaxation of the fact that $m = 0$ on the set $\{u = Mu\}$, which is the set at which it is optimal to jump. Yet, one may have at the same time $-\partial_t u - \nu \Delta u = f(m)$ where it is optimal not to jump. Therefore, the last equation allows the density m not to vanish in that situation which makes the problem more convex.

To prove existence of an equilibrium (u, m) , Bertucci approximates the second inequality of the system as in (1.17). Then, he uses Kakutani's fixed point theorem to show the existence of a solution $(u_\varepsilon, m_\varepsilon)$ of the approximated system. Finally, he passes to the limit $\varepsilon \rightarrow 0$ up to a subsequence by using suitable estimates on m_ε and u_ε . Next, the author discusses the uniqueness of the solution which occurs under some monotonicity conditions on f .

As pointed out in his work, the existence of such an equilibrium strongly depends on an estimate on m_ε which comes from the second order property of the problem that is $\nu > 0$. Thus, the MFG with jumps can only be treated in the stochastic case with his approach.

To our knowledge, the study of MFG with jumps in the deterministic case has not yet been tackled in the literature. This will be the object of this thesis, and our strategy will be different from the one used in impulse control problems which are dealt with QVIs. In fact, we will use the variational formulations of the MFG (both Eulerian and Lagrangian) for which we will prove existence and some regularity properties in the Eulerian case.

The cost function considered by each agent is similar to the one used in impulse control theory:

$$\inf_{\gamma: [0, T] \rightarrow \mathbb{R}^n} A(\gamma) + \int_0^T \frac{dI}{dm}(e_t \# \bar{Q})(\gamma(t)) dt + \int_0^T F(t, \gamma(t), e_t \# \bar{Q}) dt + \psi_T(\gamma(T)).$$

We consider a general framework where A can either represent the number of jumps of γ or the kinetic energy $\int_0^T \frac{|\dot{\gamma}(t)|^p}{p} dt$. Besides the fact that this problem is deterministic in order to be more general, the novelty of the model is that the running function is the sum of a variational term dI/dm and a continuous (not necessarily variational) term F .

1.4 Mixed Mean Field Games

In this section, I present a class of variational Mean Field Games that we call *mixed*, because the term \mathcal{G} in the potential (1.9) is the sum of two terms: one that is variational and one that is not. The problem when \mathcal{G} is variational, i.e G is the anti-derivative of a certain function, has been broached in a lot of studies that I mentioned in Section 1.2. A presentation of this class of problems can be found in [BCS17]. The case where \mathcal{G} is not the anti-derivative of a function and is *only* continuous has been studied in the paper by Cannarsa and Capuani [CC18], where the existence of a solution is established using Kakutani's fixed point theorem. Our strategy to prove existence of a solution Q to the mixed MFG is highly inspired by the latter.

The problem is as follows: given \bar{Q} , we solve

$$\min_{\substack{Q \in \mathcal{P}(\mathcal{C}) \\ e_0 \# Q = \rho_0}} \left\{ \int_{\mathcal{C}} A(\gamma) dQ(\gamma) + \int_0^T I(e_t \# Q) dt + \int_{\mathcal{C}} \int_0^T F(t, \gamma(t), e_t \# \bar{Q}) dt dQ(\gamma) + \int_{\Omega} \psi_T(x) d(e_T \# Q)(x) \right\}.$$

We will prove the existence of an equilibrium measure \bar{Q} such that \bar{Q} solve the problem for $\bar{Q} = \bar{Q}$, characterized as a fixed point of a multi-valued function.

First, let us introduce the notations and the problem solved by each agent. The metric space (\mathcal{C}, d) will designate the set of curves $\gamma: [0, T] \rightarrow \Omega$ followed by an agent in the space $\Omega \subset \mathbb{R}^d$ with $d \geq 1$ (it would also be possible to consider other domains such as the torus). Since the dimension $d \geq 1$ is a number and the distance $d(x, y)$ takes two variables, the context will separate the meaning of the notations. The set Ω is supposed to be compact. On the space \mathcal{C} we will define an action $A: \mathcal{C} \rightarrow [0, +\infty]$ which represents the cost for moving. The choice of \mathcal{C} is not very important since it is more important to know the set of curves with finite action. In particular, we require $\{\gamma \in \mathcal{C} : A(\gamma) < +\infty\} \subset BV([0, T])$. We recall that we have

$$BV([0, T], \Omega) = \left\{ \gamma: [0, T] \rightarrow \bar{\Omega} : TV(\gamma, [0, T]) := \sup_{0=t_0 < \dots < t_n=T} \sum_{i=1}^n |\gamma(t_i) - \gamma(t_{i-1})| < \infty \right\}.$$

I refer to [Loj88] for an introduction to bounded variation functions with this definition.

To be explicit, the main setting that we consider is the one where we take for \mathcal{C} the set of measurable curves defined on $[-1, T]$, valued into Ω , and constant on $[-1, 0]$. These curves are identified by a.e. equivalence, and we endow it with the L^1 distance. This space is a complete metric Polish space. We spend few words on this choice of modeling:

- Even if curves with finite action are all BV we do not want to choose $\mathcal{C} = BV$ since this would suggest to endow it with the BV norm, which would make our compactness results fail.
- Neither do we want to choose $\mathcal{C} = BV$ and endow it with the L^1 norm since this would not make it a complete metric space and some measure-theoretic tools that we use require to be in a Polish space.
- The choice of the left-continuous representative together with the choice of extending the curves to $[-1, T]$ instead of only $[0, T]$ allows to give a precise meaning of the initial point $\gamma(0)$, but also allows for immediate jumps at $t = 0^+$. Another possibility would have been to require curves to be right-continuous at $t = 0$ and, instead of imposing initial data of the form $\gamma(0) = x_0$ penalizing the difference $|\gamma(0) - x_0|$ as done in [DS23], but this would have made the presentation much heavier. On the other hand, we never impose Dirichlet boundary conditions at $t = T$, so there is no need to do the same for the final point $\gamma(T)$ and we can just think that the final penalization ψ includes the possibility to jump at $t = T$.
- Topologically, the only effect of extending curves in a constant way to $[-1, 0]$ is that the convergence in \mathcal{C} means convergence for a.e. $t \in (0, T]$ and at $t = 0$.
- When only continuous curves have finite action (we will see later that we consider two cases: the jump case where the BV setting is the adapted one, and the kinetic case where curves with finite action are continuous) it would also be possible to choose \mathcal{C} to be the set of continuous curves valued into Ω , endowed with the uniform convergence instead of the L^1 convergence, and many difficulties would disappear.

For all $t \in [0, T]$, the evaluation map e_t is defined as follows:

$$\begin{aligned} e_t: \mathcal{C} &\rightarrow \Omega \\ \gamma &\mapsto \gamma(t). \end{aligned}$$

Let $m_0 \in \mathcal{P}(\Omega)$ be a probability measure over the space Ω . The set $\mathcal{P}_{m_0}(\mathcal{C})$ will be the set of probability measures Q over the set of curves \mathcal{C} such that $\int AdQ < +\infty$ and such that the push-forward measure by the evaluation map satisfies $e_0\#Q = m_0$. We recall that for all $t \in [0, T]$, the *push-forward measure* $e_t\#Q$ is characterized by

$$\forall \varphi \in C_b(\Omega), \quad \int_{\Omega} \varphi(x) d(e_t\#Q)(x) = \int_{\mathcal{C}} \varphi(\gamma(t)) dQ(\gamma),$$

where $C_b(\Omega)$ is the set of continuous and bounded functions on Ω . The measures $e_t\#Q$ are well-defined whenever $\int AdQ < +\infty$ since in this case Q is concentrated on left-continuous curves.

Let $\bar{Q} \in \mathcal{P}_{m_0}(\mathcal{C})$. Each agent whose starting point is x_0 seeks to solve the following optimization problem:

$$\min_{\substack{\gamma \in \mathcal{C} \\ \gamma(0) = x_0}} J_{\bar{Q}}(\gamma) := A(\gamma) + \int_0^T \frac{dI}{dm}(e_t\#\bar{Q})(\gamma(t)) dt + \int_0^T F(t, \gamma(t), e_t\#\bar{Q}) dt + \psi_T(\gamma(T)). \quad (1.18)$$

In this section, we make the following hypotheses:

- (H_1) $A: \mathcal{C} \rightarrow \mathbb{R}$ is lower semi-continuous and *coercive* (its sublevel sets are compact in \mathcal{C}), and $A(\gamma) = 0$ for any constant curve γ (this simplifying assumption is the reason not to include $\psi_T(\gamma(T))$ in the action A); the function $\psi_T: \Omega \rightarrow \mathbb{R}$ is also lower semi-continuous and we suppose $\psi_T \geq 0$;
- (H_2) $I: \mathcal{P}(\Omega) \rightarrow [0, +\infty]$ is convex, l.s.c. for the weak convergence, and admits a first variation dI/dm ; we also assume $I(m_0) < +\infty$;
- (H_3) $F: [0, T] \times \Omega \times \mathcal{P}(\Omega) \rightarrow \mathbb{R}$ is continuous with respect to its three variables, and we suppose $F \geq 0$;
- (H_4) the space \mathcal{C} is such that if $Q_n \xrightarrow{*} Q$ in $\mathcal{P}_{m_0}(\mathcal{C})$ with $\int AdQ_n \leq C$, then up to a subsequence, we have $e_t\#Q_n \xrightarrow[n]{*} e_t\#Q$, for a.e. t . The notation $\xrightarrow{*}$ represents the *narrow convergence* of a sequence of measures, i.e.

$$\forall \varphi \in C_b(\mathcal{C}), \quad \int_{\mathcal{C}} \varphi(\gamma) dQ_n(\gamma) \xrightarrow[n]{*} \int_{\mathcal{C}} \varphi(\gamma) dQ(\gamma).$$

Note that the assumption on the coercivity of A could equivalently be replaced with the coercivity of $\gamma \mapsto A(\gamma) + \phi(\gamma(0))$ for a suitable coercive function $\phi : \Omega \rightarrow \mathbb{R}_+$ (and that ψ could be included in A). Indeed, due to the fact that we prescribe the initial point in the variational problem of each agent and we prescribe the initial distribution of the agents, adding a penalization on their initial point has no effect, and we can use an arbitrary coercive function ϕ such that $\int \phi dm_0 < +\infty$. In particular, very often the action A penalizes the derivative of the curve γ and for compactness we also need to bound the curve inside a compact domain, which can be obtained enforcing a bound on $\gamma(0)$.

For the sake of clarity, I recall the definitions of *lower semi-continuity* and *first variation* below. Some examples of standard first variation functions are presented in [San15, Section 7.2].

Definition 1.4.1. *Let X be a topological space which verifies the first axiom of countability. A function $f : X \rightarrow \mathbb{R}$ is lower semi-continuous (l.s.c.) in $x_0 \in X$, if f verifies one of these equivalent conditions:*

- (i) *for all $t < f(x_0)$, there exists a neighborhood U of x_0 such that for all $x \in U \cap X$, $t < f(x)$;*
- (ii) *for all $\alpha \in \mathbb{R}$, the set $\{x ; f(x) \leq \alpha\}$ is closed;*
- (iii) *$f(x_0) \leq \liminf_{x \rightarrow x_0} f(x)$.*

Definition 1.4.2. *Let $f : \mathcal{P}(\bar{\Omega}) \rightarrow [0, +\infty]$ be a functional. Let $m \in \mathcal{P}(\bar{\Omega})$ be such that for all $\varepsilon \in [0, 1]$ and all $\tilde{m} \in \mathcal{P}(\bar{\Omega}) \cap L_c^\infty(\bar{\Omega})$, $f((1-\varepsilon)m + \varepsilon \tilde{m}) < +\infty$. If it exists, we call $\frac{df}{dm}(m)$ a first variation of f such that for every perturbation $\chi := \tilde{m} - m$ with $\tilde{m} \in \mathcal{P}(\bar{\Omega}) \cap L_c^\infty(\bar{\Omega})$,*

$$\frac{d}{d\varepsilon} f(m + \varepsilon \chi)|_{\varepsilon=0} = \int_{\Omega} \frac{df}{dm}(m)(x) d\chi(x).$$

We also say that f is a first integral of $\frac{df}{dm}(m)$.

The goal is to find a measure $\bar{Q} \in \mathcal{P}_{m_0}(\mathcal{C})$ whose support is included in the set of optimal trajectories of $J_{\bar{Q}}$. This is equivalent to satisfying the following inequality:

$$\forall Q \in \mathcal{P}_{m_0}(\mathcal{C}), \int_{\mathcal{C}} J_{\bar{Q}}(\gamma) d\bar{Q}(\gamma) \leq \int_{\mathcal{C}} J_{\bar{Q}}(\gamma) dQ(\gamma). \quad (1.19)$$

One can see that (1.19) is the optimality condition of an optimization problem. In particular, if \bar{Q} minimizes

$$Q \mapsto \mathcal{U}_{\bar{Q}}(Q) = \int_{\mathcal{C}} A(\gamma) dQ(\gamma) + \int_0^T \int_{\Omega} I(e_t \# Q)(x) dx dt + \int_{\mathcal{C}} \int_0^T F(t, \gamma(t), e_t \# \bar{Q}) dt dQ(\gamma) + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ(\gamma),$$

then it formally verifies (1.19). A rigorous statement will be proved in Chapter 3, but here it motivates the study of the problem $\min_Q \mathcal{U}_{\bar{Q}}(Q)$. The goal then becomes to find \bar{Q} such that \bar{Q} is a minimizer of $\mathcal{U}_{\bar{Q}}$.

To find such a measure \bar{Q} , we will define an application H that associates with any Q_0 the set of minimizers of $\min_Q \mathcal{U}_{Q_0}(Q)$, and we will prove that it admits a fixed point \bar{Q} thanks to Kakutani's fixed point theorem [Kak41] which is reminded below.

Theorem 1.4.3. (*Kakutani's fixed point theorem*) *Let X be a nonempty convex and compact set. Let $f : X \rightarrow P(X)$ be a multi-valued function, i.e. $f(x)$ is a subset of X , which in this case will be denoted $f : X \rightrightarrows X$.*

If for all $x \in X$, $f(x)$ is nonempty and convex, and if the graph of f is closed, then f admits a fixed point, i.e. there exists $\bar{x} \in X$ such that $\bar{x} \in f(\bar{x})$.

The following Lemma is a useful tool when a sequence $(Q_N)_N$ narrowly converges to Q , and we state it in its most general form for possible use.

Lemma 1.4.4. *Let (\mathcal{C}, d) be a metric space and $f : \mathcal{C} \rightarrow \mathbb{R}$ a function bounded from below and $(Q_N)_N \subset \mathcal{P}(\mathcal{C})$ a sequence of probability measures narrowly converging towards Q . Set $\text{ldisc}(f) := \{x \in \mathcal{C} : \exists x_n \rightarrow x : \liminf_n f(x_n) < f(x)\}$, i.e. $\text{ldisc}(f)$ is the set of points where f is not l.s.c. Then, if $Q(\text{ldisc}(f)) = 0$, we have*

$$\int_{\mathcal{C}} f dQ \leq \liminf_N \int_{\mathcal{C}} f dQ_N.$$

Proof. For all $k \in \mathbb{N}$, we define $f_k : \mathcal{C} \rightarrow \mathbb{R}$ such that

$$\forall k \in \mathbb{N}, f_k(x) = \min \left(\inf_{y \in \mathcal{C}} f(y) + kd(x, y), k \right).$$

Then, for all $k \in \mathbb{N}$, f_k is k -Lipschitz and bounded, so

$$\int_{\mathcal{C}} f_k dQ_N \xrightarrow{N} \int_{\mathcal{C}} f_k dQ.$$

In particular, using $f \geq f_k$, we obtain

$$\liminf_N \int_{\mathcal{C}} f dQ_N \geq \int_{\mathcal{C}} f_k dQ.$$

Moreover, $f(x) = \sup_k f_k(x)$ for all $x \in \mathcal{C} \setminus \text{ldisc}(f)$. Indeed, for a fixed $x \in \mathcal{C}$, let $k_0 \in \mathbb{N}$ be such that $f(x) \leq k_0$. Then, for all $k \geq k_0$, we have $f_k(x) = \inf_{y \in \mathcal{C}} f(y) + kd(x, y)$. Therefore, for all $k \geq k_0$, there exists $y_k \in \mathcal{C}$ such that

$$f_k(x) \leq f(y_k) + k d(x, y_k) \leq f_k(x) + 1/k. \quad (1.20)$$

Dividing Ineq. (1.20) by $k > 0$, and letting $k \rightarrow +\infty$, we obtain

$$\lim_{k \rightarrow \infty} \frac{f(y_k)}{k} + d(x, y_k) = 0.$$

Since f is bounded from below, we have $\lim_{k \rightarrow \infty} d(x, y_k) = 0$, so

$$y_k \xrightarrow[k \rightarrow \infty]{} x.$$

Then, by taking the lim inf limit in (1.20), we have

$$\liminf_{k \rightarrow \infty} f(y_k) + k d(x, y_k) \leq \sup_k f_k(x). \quad (1.21)$$

Next, by definition of f_k and using the l.s.c. behavior of f at points in $\mathcal{C} \setminus \text{ldisc}(f)$ we obtain

$$\sup_k f_k(x) \leq f(x) \leq \liminf_{k \rightarrow \infty} f(y_k) + k d(x, y_k). \quad (1.22)$$

Ineq (1.21) and (1.22) imply

$$\sup_k f_k(x) \leq f(x) \leq \sup_k f_k(x),$$

so $f(x) = \sup_k f_k(x)$.

Using the assumption $Q(\text{ldisc}(f)) = 0$, by the monotone convergence theorem we have

$$\int_{\mathcal{C}} f dQ = \lim_{k \rightarrow \infty} \int_{\mathcal{C}} f_k dQ,$$

hence the claim. □

In order to apply the previous Lemma, we will need A to be bounded from below.

Lemma 1.4.5. *If $A: \mathcal{C} \rightarrow \mathbb{R}$ is l.s.c. and coercive, then A is bounded from below.*

Proof. One could already see it with the existence of a minimizer of a non-empty compact sublevel set by the lower semi-continuity of A , but here I suggest a proof with the definition (i) of the lower semi-continuity.

Let $\alpha > 0$ be such that $\mathfrak{C} := \{\gamma \in \mathcal{C}; A(\gamma) \leq \alpha\}$ is nonempty. Since A is l.s.c., for all $\gamma_0 \in \mathfrak{C}$, we have $\forall t < A(\gamma_0)$, there exists a neighborhood $U_{\gamma_0} \subset \mathcal{C}$ of γ_0 , such that

$$\forall \gamma \in U_{\gamma_0} \cap \mathfrak{C}, t < A(\gamma).$$

Thus, $\bigcup_{\gamma_0 \in \mathfrak{C}} U_{\gamma_0}$ is a subcover of \mathfrak{C} . By compactness of \mathfrak{C} , there exists a finite number of neighborhoods that covers \mathfrak{C} . Let us denote U_i each neighborhood of a point γ_i for every $i = 1, \dots, n$. For each i , we chose t_i such that $t_i < A(\gamma_i)$ (e.g. $t_i = A(\gamma_i) - 1$). The smallest t_{i_0} for some $i_0 \in \{1, \dots, n\}$ will verify

$$\forall \gamma \in \mathfrak{C}, t_{i_0} < A(\gamma).$$

A fortiori, for all $\gamma \in \mathcal{C}$, we have $t_{i_0} \leq A(\gamma)$. Hence A is bounded from below. □

The next Lemma shows that the coercivity of A implies the coercivity of $Q \mapsto \int A(\gamma) dQ(\gamma)$, so that we will be able to define compact subsets of $\mathcal{P}(\mathcal{C})$.

Lemma 1.4.6. *If $A: \mathcal{C} \rightarrow \mathbb{R}$ is coercive, then $Q \mapsto \int_{\mathcal{C}} A(\gamma) dQ(\gamma)$ is coercive.*

Proof. We aim to prove that for all $\alpha \in \mathbb{R}$, the set $\Gamma_\alpha = \{Q \in \mathcal{P}(\mathcal{C}); \int_{\mathcal{C}} A(\gamma) dQ(\gamma) \leq \alpha\}$ is compact.

If $\alpha > 0$, then by the Markov inequality, we have for all $N \in \mathbb{N}^*$,

$$\forall Q \in \Gamma_\alpha, Q(\{\gamma \in \mathcal{C} ; A(\gamma) > N\}) \leq \frac{\int_{\mathcal{C}} A(\gamma) dQ(\gamma)}{N} \leq \frac{\alpha}{N}.$$

In particular, for all $\varepsilon > 0$, there exists $N \in \mathbb{N}^*$ sufficiently large such that

$$\forall Q \in \Gamma_\alpha, Q(\{\gamma ; A(\gamma) \leq N\}^c) \leq \varepsilon.$$

Since $\{\gamma ; A(\gamma) \leq N\}$ is a compact set (by the coercivity of A), the family Γ_α is tight. Consequently, by Prokhorov's theorem (see e.g. [Bil99, Chapter 1, Theorems 1.4 and 6.1]), the set Γ_α is relatively compact.

There remains to show that Γ_α is closed. Let $(Q_n)_n$ be a sequence in Γ_α narrowly converging towards $Q \in \mathcal{P}(\mathcal{C})$. Let us show that $Q \in \Gamma_\alpha$.

Since A is l.s.c. by Hypothesis (H_1) and bounded from below (see Lemma 1.4.5), we have by Lemma 1.4.4 that

$$\int_{\mathcal{C}} A(\gamma) dQ(\gamma) \leq \liminf_n \int_{\mathcal{C}} A(\gamma) dQ_n(\gamma) \leq \alpha,$$

hence $Q \in \Gamma_\alpha$.

In conclusion, Γ_α is relatively compact and closed, so it is compact.

If $\alpha \leq 0$, then Γ_α is a closed subset of Γ_β for any $\beta > 0$ which is compact, so Γ_α is compact. \square

The following Theorem states the existence of an equilibrium measure \bar{Q} , as introduced at the beginning of this section.

Theorem 1.4.7. *Let $\Gamma := \Gamma_C = \{Q \in \mathcal{P}_{m_0}(\mathcal{C}) : \int_{\mathcal{C}} A(\gamma) dQ(\gamma) \leq C\}$ with C defined in Remark 1.4.8. Under the assumptions $(H1-4)$, the multifunction*

$$H: \Gamma \rightrightarrows \Gamma$$

$$\begin{aligned} \tilde{Q} &\longmapsto \operatorname{argmin}_{Q, e_0 \# Q = m_0} \mathcal{U}_{\tilde{Q}}(Q) \\ &= \operatorname{argmin}_{Q, e_0 \# Q = m_0} \left\{ \int_{\mathcal{C}} A(\gamma) dQ(\gamma) + \int_0^T I(e_t \# Q) dt + \int_{\mathcal{C}} \int_0^T F(t, \gamma(t), e_t \# \tilde{Q}) dt dQ(\gamma) + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ(\gamma) \right\} \end{aligned}$$

admits a fixed point.

Remark 1.4.8. *The application H is well-defined if we determine a suitable C such that Γ is nonempty and $\mathcal{U}_{\tilde{Q}}(Q) \subset \Gamma$. For example, we take a measure $Q_{still} \in \mathcal{P}_{m_0}(\mathcal{C})$ concentrated on constant curves, i.e. the image of m_0 through the map associating with every $x \in \Omega$ the constant curve equal to x . Then, $\int_{\mathcal{C}} A(\gamma) dQ_{still}(\gamma) = 0$ and $e_t \# Q_{still} = m_0$ for all t . If Q minimizes $\mathcal{U}_{\tilde{Q}}$, then*

$$\begin{aligned} \mathcal{U}_{\tilde{Q}}(Q) &\leq 0 + \int_0^T I(m_0) dt + \int_{\mathcal{C}} \int_0^T F(t, \gamma(0), e_t \# \bar{Q}) dt dQ_{still}(\gamma) + \int_{\mathcal{C}} \psi_T(\gamma(0)) dQ_{still}(\gamma) \\ &= TI(m_0) + \int_{\Omega} \int_0^T F(t, x, e_t \# \bar{Q}) dt dm_0(x) + \int_{\Omega} \psi_T dm_0 \\ &\leq TI(m_0) + \int_{\Omega} \int_0^T \sup_{\mathcal{P}(\bar{\Omega})} F(t, x, \cdot) dt dm_0(x) + \int_{\Omega} \psi_T dm_0 := C. \end{aligned}$$

and in particular we have $\int_{\mathcal{C}} A(\gamma) dQ(\gamma) \leq C$.

Proof. To prove the theorem, we will apply Kakutani's fixed point theorem, hence we need to verify the three following conditions:

(i) *For each measure $\tilde{Q} \in \Gamma$, the set $H(\tilde{Q})$ is convex.*

This is a consequence of the fact that the functional which is minimized over Q is convex in Q , and it is actually the only reason to require the convexity of I .

Let Q_1 and Q_2 be in $H(\tilde{Q})$ and $\lambda \in [0, 1]$. We denote by M the minimum value $M := \mathcal{U}_{\tilde{Q}}(Q_1) = \mathcal{U}_{\tilde{Q}}(Q_2)$. By the linearity of e_t and the convexity of I , we have

$$\begin{aligned}
\mathcal{U}_{\tilde{Q}}(\lambda Q_1 + (1 - \lambda)Q_2) &= \int_{\mathcal{C}} A(\gamma) d(\lambda Q_1 + (1 - \lambda)Q_2)(\gamma) \\
&\quad + \int_0^T I(e_t \# (\lambda Q_1 + (1 - \lambda)Q_2)) dt \\
&\quad + \int_{\mathcal{C}} \int_0^T F(t, \gamma(t), e_t \# \tilde{Q}) dt d(\lambda Q_1 + (1 - \lambda)Q_2)(\gamma) \\
&\quad + \int_{\mathcal{C}} \psi_T(\gamma(T)) d(\lambda Q_1 + (1 - \lambda)Q_2)(\gamma) \\
&= \lambda \int_{\mathcal{C}} A(\gamma) dQ_1(\gamma) + \int_0^T I(e_t \# (\lambda Q_1 + (1 - \lambda)Q_2)) dt \\
&\quad + \lambda \int_{\mathcal{C}} \int_0^T F(t, \gamma(t), e_t \# \tilde{Q}) dt dQ_1(\gamma) + \lambda \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ_1(\gamma) \\
&\quad + (1 - \lambda) \int_{\mathcal{C}} A(\gamma) dQ_2(\gamma) + (1 - \lambda) \int_{\mathcal{C}} \int_0^T F(t, \gamma(t), e_t \# \tilde{Q}) dt dQ_2(\gamma) \\
&\quad + (1 - \lambda) \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ_2(\gamma) \\
&\leq \lambda \mathcal{U}_{\tilde{Q}}(Q_1) + (1 - \lambda) \mathcal{U}_{\tilde{Q}}(Q_2) = M,
\end{aligned}$$

which proves that $\lambda Q_1 + (1 - \lambda)Q_2 \in H(\tilde{Q})$, and consequently, that $H(\tilde{Q})$ is a convex set.

(ii) For each $\tilde{Q} \in \Gamma$, the set $H(\tilde{Q})$ is non-empty.

We have to prove the existence of a minimizer in the problem defining $H(\tilde{Q})$. The coercivity of A implies that the set Γ is compact for the narrow convergence of probability measures on \mathcal{C} (see Lemma 1.4.6). Hence, there exists a subsequence $(Q_{n_k})_k$ narrowly converging towards $Q \in \Gamma$. We re-extract a subsequence in order to apply, later on (H_4) .

Let us estimate $\liminf_k \mathcal{U}_{\tilde{Q}}(Q_{n_k})$ as $k \rightarrow \infty$.

Since A and ψ_T are l.s.c. and bounded from below (see Lemma 1.4.5), we have by Lemma 1.4.4,

$$\int_{\mathcal{C}} A(\gamma) dQ(\gamma) \leq \liminf_k \int_{\mathcal{C}} A(\gamma) dQ_{n_k}(\gamma) \quad (1.23)$$

$$\text{and } \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ(\gamma) \leq \liminf_k \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ_{n_k}(\gamma). \quad (1.24)$$

Using (H_4) and after extracting a subsequence, we have $e_t \# Q_{n_k} \xrightarrow{*} e_t \# Q$ for a.e. t . By the lower semi-continuity of I and Fatou's Lemma, we obtain

$$\int_0^T I(e_t \# Q) dt \leq \int_0^T \liminf_k I(e_t \# Q_{n_k}) dt \leq \liminf_k \int_0^T I(e_t \# Q_{n_k}) dt, \quad (1.25)$$

and since F is continuous in its variables, we have

$$\int_{\mathcal{C}} \int_0^T F(t, \gamma(t), e_t \# \tilde{Q}) dt dQ_{n_k}(\gamma) \xrightarrow{k \rightarrow \infty} \int_{\mathcal{C}} \int_0^T F(t, \gamma(t), e_t \# \tilde{Q}) dt dQ(\gamma). \quad (1.26)$$

Gathering the results (1.23), (1.25), (1.26) and (1.24) gives

$$\mathcal{U}_{\tilde{Q}}(Q) \leq \liminf_k \mathcal{U}_{\tilde{Q}}(Q_{n_k}) \leq \inf_{Q \in \Gamma} \mathcal{U}_{\tilde{Q}}(Q),$$

which proves that Q is a minimizer of $\mathcal{U}_{\tilde{Q}}$ and consequently, $H(\tilde{Q})$ is non-empty.

(iii) The graph of H is closed.

Let $(\tilde{Q}_n)_n$ be a sequence narrowly converging towards \tilde{Q}_∞ and $Q_n \in H(\tilde{Q}_n)$ a sequence converging towards Q_∞ . We will prove $Q_\infty \in H(\tilde{Q}_\infty)$.

Since $Q_n \in H(\tilde{Q}_n)$, we have for all $Q \in \Gamma$,

$$\begin{aligned} & \int_{\mathcal{C}} A(\gamma) dQ_n(\gamma) + \int_0^T I(e_t \# Q_n) dt + \int_{\mathcal{C}} \int_0^T F(t, \gamma(t), e_t \# \tilde{Q}_n) dt dQ_n(\gamma) + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ_n(\gamma) \quad (1.27) \\ & \leq \int_{\mathcal{C}} A(\gamma) dQ(\gamma) + \int_0^T I(e_t \# Q) dt + \int_{\mathcal{C}} \int_0^T F(t, \gamma(t), e_t \# \tilde{Q}_n) dt dQ(\gamma) + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ(\gamma). \end{aligned}$$

For each term in (1.27), we consider its limit when $n \rightarrow +\infty$.

Let us begin with $\int A(\gamma) dQ_n(\gamma)$ and $\int \psi_T(\gamma(T)) dQ_n(\gamma)$. The functions A and ψ_T are l.s.c. and bounded from below, so by Lemma 1.4.4, we have

$$\begin{aligned} & \int_{\mathcal{C}} A(\gamma) dQ_{\infty}(\gamma) \leq \liminf_n \int_{\mathcal{C}} A(\gamma) dQ_n(\gamma) \\ & \text{and } \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ_{\infty}(\gamma) \leq \liminf_n \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ_n(\gamma). \end{aligned}$$

For the second term, by Hypothesis (H_4) (up to extracting a subsequence), the lower semi-continuity of I and Fatou's Lemma, we have the inequality

$$\int_0^T I(e_t \# Q_{\infty}) dt \leq \int_0^T \liminf_n I(e_t \# Q_n) dt \leq \liminf_n \int_0^T I(e_t \# Q_n) dt.$$

When it comes to $\int_{\mathcal{C}} \int_0^T F(t, \gamma(t), e_t \# \tilde{Q}_n) dt dQ_n(\gamma)$, the continuity of F gives

$$\int_{\mathcal{C}} \int_0^T F(t, \gamma(t), e_t \# \tilde{Q}_n) dt dQ_n(\gamma) \xrightarrow{n \rightarrow +\infty} \int_{\mathcal{C}} \int_0^T F(t, \gamma(t), e_t \# \tilde{Q}_{\infty}) dt dQ_{\infty}(\gamma).$$

In the right-hand side of (1.27), only one term depends on n . The dominated convergence theorem and Hypothesis (H_4) yields up to a subsequence

$$\int_{\mathcal{C}} \int_0^T F(t, \gamma(t), e_t \# \tilde{Q}_n) dt dQ(\gamma) \xrightarrow{n} \int_{\mathcal{C}} \int_0^T F(t, \gamma(t), e_t \# \tilde{Q}_{\infty}) dt dQ(\gamma).$$

Finally, passing to the limit \liminf_n in (1.27) gives

$$\begin{aligned} & \int_{\mathcal{C}} A(\gamma) dQ_{\infty}(\gamma) + \int_0^T I(e_t \# Q_{\infty}) dt + \int_{\mathcal{C}} \int_0^T F(t, \gamma(t), e_t \# \tilde{Q}_{\infty}) dt dQ_{\infty}(\gamma) + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ_{\infty}(\gamma) \\ & \leq \int_{\mathcal{C}} A(\gamma) dQ(\gamma) + \int_0^T I(e_t \# Q) dt + \int_{\mathcal{C}} \int_0^T F(t, \gamma(t), e_t \# \tilde{Q}_{\infty}) dt dQ(\gamma) + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ(\gamma). \end{aligned}$$

Since this inequality is true for all $Q \in \Gamma$, Q_{∞} is a minimizer of $\mathcal{U}_{\tilde{Q}_{\infty}}$, namely $Q_{\infty} \in H(\tilde{Q}_{\infty})$. This proves that the graph of H is closed.

Since the conditions (i), (ii), (iii) are verified, by Kakutani's theorem, we can conclude that H admits a fixed point in Γ . \square

1.4.1 The jump case

The theorem 1.4.7 can be applied to the case where $A = S$, where the functional S describes the number of *jumps* of a curve γ . As we said at the beginning of the section, we choose \mathcal{C} to be the space of measurable curves (defined on $[-1, T]$ but constant on $[-1, 0]$) equipped with the L^1 -norm and $S: \mathcal{C} \rightarrow [0, \infty]$ is defined via

$$S(\gamma) = \inf \{ \#\{t ; \tilde{\gamma} \text{ is discontinuous in } t\} ; \tilde{\gamma} = \gamma \text{ a.e. and } \tilde{\gamma} \text{ is piecewise constant} \}. \quad (1.28)$$

Of course, $S(\gamma) = +\infty$ whenever γ does not admit a piecewise constant representative.

We note that we have

$$TV(\gamma; [0, T]) \leq \text{diam}(\Omega) S(\gamma),$$

where $\text{diam}(\Omega)$ is the diameter of the compact domain Ω .

The following proposition shows that the function S is lower semi-continuous and coercive:

Proposition 1.4.9. *$S: \mathcal{C} \rightarrow \mathbb{N} \cup \{+\infty\}$ is l.s.c. and coercive.*

Proof. We need to prove that for all $\alpha \in \mathbb{R}$, the set $E_\alpha := \{\gamma \in \mathcal{C} ; S(\gamma) \leq \alpha\}$ is compact. This proves at the same time lower semi-continuity (for which the closedness of E_α would have been enough) and coercivity.

We assume $\alpha \geq 0$ since otherwise $E_\alpha = \emptyset$. Let $(\gamma_n)_n$ be a sequence in E_α . Take $\tilde{\gamma}_n$ a representative of γ_n which realizes exactly $S(\gamma_n)$ jumps. Up to extracting a subsequence there exists a number N such that $N - 1 \leq \alpha$ and each $\tilde{\gamma}_n$ is of the form $\tilde{\gamma}_n = \sum_{i=0}^{N-1} z_i^n \mathbb{1}_{[a_i^n, a_{i+1}^n)}$. It is then possible to extract converging subsequences from $(z_i^n)_n$ and $(a_i^n)_n$ and prove that $\tilde{\gamma}_n$ converges to a curve γ of the form $\gamma = \sum_{i=0}^{N-1} z_i \mathbb{1}_{[a_i, a_{i+1})}$. This convergence will be very strong: from the convergence of the values z_i^n taken by the curves and of the intervals on which these values are taken we can deduce that we have a.e. and L^1 convergence (but not uniform convergence). Since γ is piecewise constant and defined using N intervals we have $S(\gamma) \leq N - 1 \leq \alpha$ (the inequality in $S(\gamma) \leq N$ could be strict, since we do not know whether the points z_i are distinct). This proves the compactness of E_α since we extracted a converging subsequence, and the limit still belongs to the same set. \square

There remains to verify that Assumption (H_4) holds true:

Proposition 1.4.10. *Let $(Q_n)_n$ be a sequence narrowly converging towards Q in $\mathcal{P}(\mathcal{C})$ with $\int_{\mathcal{C}} S(\gamma) dQ_n(\gamma) \leq C < +\infty$. Then there exists a subsequence Q_{n_k} such that*

$$e_t \# Q_{n_k} \xrightarrow[k]{*} e_t \# Q, \text{ a.e. } t.$$

Proof. For each n , let us define a measure μ_n on $[0, T]$ via

$$\mu_n(I) := \int_{\mathcal{C}} TV(\gamma; I) dQ_n(\gamma)$$

for every open interval I . The total mass of μ_n is equal to

$$\int_{\mathcal{C}} TV(\gamma; [0, T]) dQ_n(\gamma) \leq \text{diam}(\Omega) \int_{\mathcal{C}} S(\gamma) dQ_n(\gamma) \leq C.$$

Since the distribution function defined by $F_{\mu_n}(t) = \inf_{s \geq t} \mu_n(] - 1, s])$ is nondecreasing, bounded and right-continuous, μ_n is well-defined as a positive measure on $[0, T]$. The sequence μ_n is then bounded in the space of positive measures on $[0, T]$ and we can extract a converging subsequence $\mu_{n_k} \xrightarrow[k]{*} \mu$. We will then choose the corresponding subsequence Q_{n_k} .

We now claim that the convergence $e_t \# Q_{n_k} \xrightarrow[k]{*} e_t \# Q$ occurs for each t which is not an atom of μ , i.e. for all but a countable quantity of t .

We fix a test function $\phi \in \text{Lip}(\Omega)$ and we need to consider $\int_{\mathcal{C}} \phi(\gamma(t)) dQ_n(\gamma)$. The function $\mathcal{C} \ni \gamma \mapsto \Phi(\gamma) := \phi(\gamma(t))$ is not continuous, so this term does not pass easily to the limit. For this reason we fix $\varepsilon > 0$ and consider instead the function $\mathcal{C} \ni \gamma \mapsto \Phi_\varepsilon(\gamma) := \int_{t-\varepsilon}^{t+\varepsilon} \phi(\gamma(s)) ds$, which is continuous for the L^1 convergence of γ . In particular, we have $\lim_n \int_{\mathcal{C}} \Phi_\varepsilon(\gamma) dQ_n(\gamma) = \int_{\mathcal{C}} \Phi_\varepsilon(\gamma) dQ(\gamma)$.

Our goal is to prove $\lim_k \int_{\mathcal{C}} \Phi(\gamma) dQ_{n_k}(\gamma) = \int_{\mathcal{C}} \Phi(\gamma) dQ(\gamma)$. By the lower semi-continuity of $\gamma \mapsto TV(\gamma, I)$ on each open interval I with the topology induced by the L^1 -norm, we have

$$\int_{\mathcal{C}} TV(\gamma,]t - \varepsilon, t + \varepsilon]) dQ(\gamma) \leq \liminf_n \int_{\mathcal{C}} TV(\gamma,]t - \varepsilon, t + \varepsilon]) dQ_n(\gamma)$$

and by the Portmanteau Lemma we have

$$\limsup_n \mu_n(]t - \varepsilon, t + \varepsilon]) \leq \mu(]t - \varepsilon, t + \varepsilon]).$$

Both of these inequalities and the definition of μ_n imply

$$\int_{\mathcal{C}} TV(\gamma,]t - \varepsilon, t + \varepsilon]) dQ(\gamma) \leq \mu(]t - \varepsilon, t + \varepsilon]). \quad (1.29)$$

We use $|\Phi(\gamma) - \Phi_\varepsilon(\gamma)| \leq \text{Lip}(\phi) TV(\gamma;]t - \varepsilon, t + \varepsilon])$ and the inequality (1.29) which allows to obtain

$$\left| \int_{\mathcal{C}} \Phi dQ_{n_k} - \int_{\mathcal{C}} \Phi_\varepsilon dQ_{n_k} \right| \leq \text{Lip}(\phi) \mu_{n_k}(]t - \varepsilon, t + \varepsilon]), \quad \left| \int_{\mathcal{C}} \Phi dQ - \int_{\mathcal{C}} \Phi_\varepsilon dQ \right| \leq \text{Lip}(\phi) \mu(]t - \varepsilon, t + \varepsilon]).$$

This implies

$$\left| \int_{\mathcal{C}} \Phi dQ_{n_k} - \int_{\mathcal{C}} \Phi dQ \right| \leq \text{Lip}(\phi) (\mu(]t - \varepsilon, t + \varepsilon]) + \mu_{n_k}(]t - \varepsilon, t + \varepsilon])) + \left| \int_{\mathcal{C}} \Phi_\varepsilon dQ_{n_k} - \int_{\mathcal{C}} \Phi_\varepsilon dQ \right|.$$

Using $\limsup_k \mu_{n_k}(E) \leq \mu(E)$, which holds for every closed set E , we obtain

$$\limsup_k \left| \int_{\mathcal{C}} \Phi dQ_{n_k} - \int_{\mathcal{C}} \Phi dQ \right| \leq 2\text{Lip}\phi \mu([t - \varepsilon, t + \varepsilon]).$$

The arbitrariness of ε shows the desired limit as soon as $\mu(\{t\}) = 0$. □

All the assumptions for Theorem 1.4.7 are verified, so we can deduce that there exists a measure $\bar{Q} \in \Gamma$ which is a fixed point, i.e. it minimizes $\mathcal{U}_{\bar{Q}}$ in the case $A = S$.

1.4.2 The kinetic case

Another application of Theorem 1.4.7 can be obtained by using $A = K$ with the function K defined as follows

$$K(\gamma) = \begin{cases} \int_0^T \frac{|\dot{\gamma}(t)|^p}{p} dt & \text{if } \gamma \in W^{1,p}([0, T]), \\ +\infty & \text{if not,} \end{cases} \quad (1.30)$$

for a given exponent p with $1 < p < \infty$. In this case we can either choose \mathcal{C} to be again the same space as in the jump case, or even choose $\mathcal{C} = C^0([0, T]; \Omega)$, the set of continuous curves endowed with the sup distance, i.e. with the topology of uniform convergence. This second choice simplifies some arguments and make this section independent of the previous one.

Proposition 1.4.11. *The application $K: \mathcal{C} \rightarrow \mathbb{R}$ is l.s.c. and coercive.*

Proof. It suffices to show that for all $\alpha \in \mathbb{R}$, the lower level-sets $E_\alpha = \{\gamma \in W^{1,p}([0, T]); \int_0^T \frac{|\dot{\gamma}(t)|^p}{p} dt \leq \alpha\}$ are compact for the uniform convergence (which also implies compactness for the L^1 convergence, in case the reader prefers to use \mathcal{C} as in the previous section on the jump case).

If $\alpha < 0$, then $E_\alpha = \emptyset$.

Now suppose $\alpha \geq 0$. Let $(\gamma_n)_n \subset E_\alpha$ be a sequence. The sequence $(\gamma_n)_n$ is bounded in $W^{1,p}([0, T])$ (we also use the compactness of Ω to bound the L^p norm of γ_n and not only of its derivative). By the Sobolev inequality we obtain for all $n \in \mathbb{N}$,

$$[\gamma_n]_{C^{0,1-\frac{1}{p}}([0,T])} = \sup_{s \neq t} \frac{|\gamma_n(t) - \gamma_n(s)|}{|t - s|^{1-1/p}} \leq \|\dot{\gamma}_n\|_{L^p([0,T])}.$$

Hence, the sequence $(\gamma_n)_n$ is equicontinuous and we can apply Arzelà-Ascoli's theorem. Thus, there exists a subsequence $(\gamma_{n_k})_k$ that converges uniformly towards a limit curve γ . It is also possible to extract a further subsequence so that $\dot{\gamma}_{n_k}$ weakly-* converges in L^p to a function, which can only be (identifying the limit with the distributional limit) $\dot{\gamma}$. The lower semi-continuity of the L^p -norm shows that we have

$$\int_0^T \frac{|\dot{\gamma}(t)|^p}{p} dt \leq \liminf_n \int_0^T \frac{|\dot{\gamma}_n(t)|^p}{p} dt \leq \alpha,$$

which proves that the limit γ belongs to E_α and hence the compactness of E_α . □

The last point to check is if assumption (H_4) is verified.

Proposition 1.4.12. *Let $(Q_n)_n$ be a sequence narrowly converging towards Q in $\mathcal{P}_{m_0}(\mathcal{C})$. Then*

$$e_t \# Q_n \xrightarrow{*} e_t \# Q, \forall t.$$

Proof. We remind that in this case the space \mathcal{C} has been endowed with the topology of uniform convergence. Hence, for each t , differently from what we faced in the proof of Proposition 1.4.10, the map $e_t: \mathcal{C} \rightarrow \Omega$ is continuous. Continuous maps transform narrowly converging sequences on \mathcal{C} into narrowly converging sequences on Ω , hence the result is straightforward. □

Since all the hypotheses for Theorem 1.4.7 are verified, we can conclude also in this case the existence of a fixed point $\bar{Q} \in \Gamma$ that minimizes $\mathcal{U}_{\bar{Q}}$.

Chapter 2

Equivalence between Eulerian and Lagrangian point of views and Eulerian fixed point

In this chapter, in particular in Section 2.1, we prove that the following problem formulated in a Lagrangian way, where the minimum is taken on the probability measures over the curves $Q \in \mathcal{P}(\mathcal{C})$,

$$\min_Q \left\{ \int_{\mathcal{C}} A(\gamma) dQ(\gamma) + \int_0^T I(e_t \# Q) dt + \int_{\mathcal{C}} \int_0^T V(t, \gamma(t)) dt dQ(\gamma) + \int_{\mathcal{C}} \psi(\gamma(T)) dQ(\gamma) \right\}$$

is equivalent to the following Eulerian problem

$$\min_{\rho} \mathbf{A}(\rho) + \int_0^T I(\rho(t)) dt + \int_0^T \int_{\Omega} V(t, \cdot) d\rho(t) dt + \int_{\Omega} \psi d\rho(T),$$

where the minimum is on the curves of measures, i.e. $\rho: [0, T] \rightarrow L^2(\Omega)$ such that $e_t \# Q = \rho_t$. The result is true in the following cases:

- $\mathbf{A}(\rho) = \frac{1}{2} \int_0^T \int_{\Omega} |\dot{\rho}| dx dt$ when A describes the number of jumps S .
- $\mathbf{A}(\rho) = \int_0^T \frac{1}{p} |\rho'|_{W_p}(t)^p dt$ when A is the kinetic energy K where $|\rho'|_{W_p}$ is the metric derivative defined in (1.7).

This equivalence is already known in the kinetic case, so most of the proof will be performed in the jump case where we use an important property from optimal transport.

Here $\iint |\dot{\rho}|$ is an informal notation which stands for the time derivative of ρ . More, precisely, we should use the length of the curve ρ in the space of measures endowed with the total variation norm. This identification (which is rigorous for smooth curves) comes from the fact that the total variation norm coincides with the L^1 -norm for absolutely continuous measures (and the term $\int_0^T I(\rho(t)) dt$ forces $\rho(t)$ to be absolutely continuous for a.e. t) and that the length of a curve valued into a normed vector space is the integral of the norm of its velocity.

In the Mixed Mean Field Game problem presented in Section 1.4, we showed the existence of a fixed point \bar{Q} which minimizes

$$\min_Q \left\{ \int_{\mathcal{C}} A(\gamma) dQ(\gamma) + \int_0^T I(e_t \# Q) dt + \int_{\mathcal{C}} \int_0^T V(t, \gamma(t), e_t \# \bar{Q}) dt dQ(\gamma) + \int_{\mathcal{C}} \psi(\gamma(T)) dQ(\gamma) \right\}$$

and which by Chapter 3 is in fact a Nash equilibrium.

By the equivalence result from Theorem 2.1.7, $\bar{\rho}(t) = e_t \# \bar{Q}$ minimizes

$$\min_{\rho} \mathbf{A}(\rho) + \int_0^T I(\rho(t)) dt + \int_0^T \int_{\Omega} V(t, \cdot, \bar{\rho}(t)) d\rho(t) dt + \int_{\Omega} \psi d\rho(T).$$

In Section 2.2, we show that, conversely, such a fixed point $\bar{\rho}$ exists by the Banach fixed point theorem.

2.1 Equivalence between Eulerian and Lagrangian problems

The goal of this section is to prove that, instead of considering a minimization problem among measures on curves (the *Lagrangian* viewpoint), it is possible to consider a minimization problem among curves of measures (the *Eulerian* viewpoint). In the kinetic case this is well-known and stems out of Theorem 1.2.3.

One can informally understand why the Eulerian equivalent of $\int K dQ$ involves the power p of the metric derivative in W_p . Indeed, we can see $K(\gamma)$ as a limit (or a sup) of $\frac{1}{p} \sum_k (t_{k+1} - t_k) \left| \frac{|\gamma(t_{k+1}) - \gamma(t_k)|}{t_{k+1} - t_k} \right|^p$ when the partitions $(t_k)_k$ become finer and finer. The minimal value of optimal transport problems with cost $(x, y) \mapsto |x - y|^p$ is the power p of the Wasserstein distance W_p . Hence, we obtain as an Eulerian equivalent of $\int K dQ$ a sum of the form $\frac{1}{p} \sum_k (t_{k+1} - t_k) \frac{W_p^p(\rho_{t_{k+1}}, \rho_{t_k})}{|t_{k+1} - t_k|^p}$, which is itself a discretization of $\frac{1}{p} \int_0^T |\rho'|_{W_p}(t)^p dt$.

If we want to do a similar procedure for the jump case, i.e. with S instead of K , we first note that $S(\gamma)$ can be discretized as follows

$$S(\gamma) = \sup_{t_0 < t_1 < \dots < t_N \subset [0, T]} \sum_k \mathbb{1}_{\gamma(t_{k+1}) \neq \gamma(t_k)}, \quad (2.1)$$

where, of course, one has to choose the left-continuous representative of γ (otherwise we need to take an infimum among representatives). Then, we notice two points. The first point is the absence of the dependence in $(t_{k+1} - t_k)$, corresponding to the case $p = 1$: as a consequence, we will not have a kinetic energy but a length, for a certain distance. The other is the fact that $(x, y) \mapsto \mathbb{1}_{x \neq y}$ is a distance (the so-called discrete distance) and that the associated Wasserstein distance is the total variation distance between measures, as explained in the following Lemma, whose proof can be found, for instance, in [Vil03].

Lemma 2.1.1. *For all μ and ν in $\mathcal{P}(\Omega)$,*

$$\inf_{\pi \in \Pi(\mu, \nu)} \int_{\Omega \times \Omega} \mathbb{1}_{x \neq y} d\pi(x, y) = \sup_{f: \sup f - \inf f \leq 1} \int_{\Omega} f d(\mu - \nu) = \sup_{f: |f| \leq 1/2} \int_{\Omega} f d(\mu - \nu) = \frac{1}{2} \|\mu - \nu\|_{\mathcal{M}}$$

where $\Pi(\mu, \nu)$ is the set of probability measures over $\Omega \times \Omega$ whose marginals are μ et ν and $\|\cdot\|_{\mathcal{M}}$ is the norm on the space of measures, i.e. the total variation of (signed) measures.

For these reasons, we will consider the length of a curve ρ in the space $\mathcal{P}(\Omega)$ computed according to the distance $\|\cdot\|_{\mathcal{M}}$. We will denote it by \mathbf{L} .

$$\mathbf{L}(\rho) := \sup_{0=t_0 < \dots < t_i < \dots < t_N=T} \sum_{i=0}^{N-1} \|\rho(t_{i+1}) - \rho(t_i)\|_{\mathcal{M}}.$$

Very informally we can write

$$\mathbf{L}(\rho) = \int_0^T \int_{\Omega} |\dot{\rho}(t, x)| dx dt$$

since, when all the measures ρ_t are absolutely continuous, the total variation distance coincides with the L^1 distance, and computing a length corresponds to computing the integral of the norm of the derivative.

The next proposition gives one inequality which is of interest for the equivalence of a Lagrangian and an Eulerian problem.

Proposition 2.1.2. *For all measure $Q \in \mathcal{P}(\mathcal{C})$, setting for all $t \in [0, T]$, $e_t \# Q = \rho_t$, we have*

$$\mathbf{L}(\rho) \leq 2 \int_{\mathcal{C}} S(\gamma) dQ(\gamma).$$

Proof. Let $t_0 < \dots < t_i < \dots < t_N$ be a subdivision of $[0, T]$. For all $i \in \{0, \dots, N-1\}$, if φ_i is a function such that $\|\varphi_i\|_{\infty} \leq 1$, we have

$$\begin{aligned} \sum_{i=0}^{N-1} \int_{\Omega} \varphi_i d(\rho(t_{i+1}) - \rho(t_i)) &= \sum_{i=0}^{N-1} \int_{\mathcal{C}} \varphi_i(\gamma(t_{i+1})) - \varphi_i(\gamma(t_i)) dQ(\gamma) \\ &\leq 2 \int_{\mathcal{C}} \sum_{i=0}^{N-1} \mathbb{1}_{\gamma(t_i) \neq \gamma(t_{i+1})} dQ(\gamma) \leq 2 \int_{\mathcal{C}} S(\gamma) dQ(\gamma). \end{aligned}$$

The term in the right-hand side of the inequality is independent of φ_i , so we can take the supremum over the functions bounded by 1. We obtain

$$\sum_{i=0}^{N-1} \|\rho(t_{i+1}) - \rho(t_i)\|_{\mathcal{M}} \leq 2 \int_{\mathcal{C}} S(\gamma) dQ(\gamma).$$

Next, by taking the supremum over the set of subdivisions of $[0, T]$, we obtain the desired result, by definition of length. \square

In the following proposition, we construct a sequence of measures $(Q_N)_N$ which are transport plans for the optimal transport problem with $c(x, y) = \mathbb{1}_{x \neq y}$:

Lemma 2.1.3. *Given $N + 1$ measures $\rho_0, \dots, \rho_N \in \mathcal{P}(\Omega)$ there exists a measure $\pi^N \in \mathcal{P}(\Omega^{N+1})$ which has marginals $\rho_0, \dots, \rho_N \in \mathcal{P}(\Omega)$ and such that*

$$\int_{\Omega^{N+1}} \sum_{k=0}^{N-1} \mathbb{1}_{x_k \neq x_{k+1}} d\pi^N(x_0, \dots, x_N) = \frac{1}{2} \sum_{k=0}^{N-1} \|\rho_k - \rho_{k+1}\|_{\mathcal{M}}. \quad (2.2)$$

Proof. By Lemma 2.1.1, we know that for each k we have

$$\inf_{\pi \in \Pi(\rho_k, \rho_{k+1})} \int_{\Omega \times \Omega} \mathbb{1}_{x \neq y} d\pi(x, y) = \frac{1}{2} \|\rho_k - \rho_{k+1}\|_{\mathcal{M}}.$$

Since the cost $(x, y) \mapsto \mathbb{1}_{x \neq y}$ is lower semi-continuous, we can choose for each k a measure $\pi_k \in \mathcal{P}(\Omega \times \Omega)$ attaining such an infimum. The second marginal of π_k equals the first marginal of π_{k+1} . Hence it is possible, applying several times the well-known gluing lemma about the composition of transport plans (see [San15, Lemma 5.5]), to build a measure $\pi^N \in \mathcal{P}(\Omega^{N+1})$ such that, for each $k \leq N - 1$, the projection of this measure onto the k -th and $(k + 1)$ -th coordinate (i.e. its image through the map $(x_0, \dots, x_N) \mapsto (x_k, x_{k+1})$) equals π_k . Such a measure satisfies

$$\int_{\Omega^{N+1}} \mathbb{1}_{x_k \neq x_{k+1}} d\pi^N(x_0, \dots, x_N) = \frac{1}{2} \|\rho_k - \rho_{k+1}\|_{\mathcal{M}}$$

and, summing over k , we obtain the claim. \square

Lemma 2.1.4. *Let $[-1, T] \ni t \mapsto \rho(t)$ be a left-continuous curve valued into $\mathcal{P}(\Omega)$, constant on $[-1, 0]$, and such that $\rho(0) = m_0$. Let E be a finite subset of $(-1, T) \setminus \{0\}$. Then, there exists a measure $Q \in \mathcal{P}_{m_0}(\mathcal{C})$ such that*

$$\int_{\mathcal{C}} S dQ \leq \frac{1}{2} \mathbf{L}(\rho) \quad \text{and } e_t \# Q = \rho_t \text{ for all } t \in E.$$

Proof. We order the points in E and call them t_1, t_2, \dots, t_{N-1} . We add, if needed, $t_0 = -1$ and $t_N = T$. There is an index j such that $t_j \leq 0$ and $t_{j+1} > 0$. We take $\rho_i = \rho(t_i)$ and apply the previous Lemma. Then, we consider a map $Z : \Omega^{N+1} \rightarrow \mathcal{C}$ where the points of Ω^{N+1} are denoted (x_0, x_1, \dots, x_N) . We define Z in this way: $Z(x)(t) = x_i$ for every $t \in (t_{i-1}, t_i] \cap (0, T]$ (which requires $i \geq j + 1$); $Z(x)(t) = x_j$ for every $t \in (t_j, 0]$; $Z(x)(t) = x_i$ for every $t \in (t_{i-1}, t_i]$ for $i \leq j$ and $Z(x)(-1) = x_0$. Then, we take $Q = Z \# \pi$ and we see that it satisfies the condition of the claim, since $Z(x)(t_i) = x_i$ and $S(Z(x)) = \sum_{k=0}^{N-1} \mathbb{1}_{x_k \neq x_{k+1}}$. \square

Proposition 2.1.5. *Let $[-1, T] \ni t \mapsto \rho(t)$ be a left-continuous curve valued into $\mathcal{P}(\Omega)$ with $\rho(0) = m_0$ and constant on $[-1, 0]$, with $\rho(0) = m_0$. Then there exists a measure $Q \in \mathcal{P}_{m_0}(\mathcal{C})$ such that*

$$\int_{\mathcal{C}} S dQ \leq \frac{1}{2} \mathbf{L}(\rho) \quad \text{and } e_t \# Q = \rho_t \text{ for all } t.$$

Proof. First, note that there is nothing to prove if $\mathbf{L}(\rho) = +\infty$. We then assume $\mathbf{L}(\rho) < +\infty$ and consider the measure $\bar{\mu}$ on $[-1, T]$ defined via $\bar{\mu}(I) = TV(\rho; I)$ for every open interval I , where the total variation is to be intended when endowing $\mathcal{P}(\Omega)$ with the distance induced by the $\|\cdot\|_{\mathcal{M}}$ norm. We have $\bar{\mu}([-1, T]) = \mathbf{L}(\rho) < +\infty$. Let us consider a dense countable set $E_\infty = \{t_k, k \in \mathbb{N}\}$ of points in $(-1, T) \setminus \{0\}$ which are not atoms of $\bar{\mu}$. For every N , we take t_1, t_2, \dots, t_N and we reorder them, calling t_k^N the ordered sequence that we obtain, characterized by $t_i^N < t_{i+1}^N$ and $\{t_i^N\}_{i=1, \dots, N} = \{t_i\}_{i=1, \dots, N}$. We then apply Lemma 2.1.4 to $E = E_N = \{t_i^N\}_{i=1, \dots, N}$ and we obtain the existence of a curve $Q_N \in \mathcal{P}_{m_0}(\mathcal{C})$ such that

$$\int_{\mathcal{C}} S dQ \leq \frac{1}{2} \mathbf{L}(\rho) \quad \text{and } e_t \# Q = \rho_t \text{ for all } t = 0, t_1^N, \dots, t_N^N, T.$$

We can say that $Q_N \in \Gamma_C = \{Q \in \mathcal{P}_{m_0}(\mathcal{C}) : \int_{\mathcal{C}} SdQ \leq C\}$ and such a set is compact for the narrow convergence, so that we can extract a narrowly converging subsequence $Q_{N_k} \xrightarrow{*} Q$. By construction, for each $t \in E$ and each N large enough (depending on t) we have $e_t \# Q_N = \rho(t)$. We want now to apply Lemma 1.4.10 in order to obtain $e_t \# Q_N \xrightarrow{*} e_t \# Q$, which would imply $e_t \# Q = \rho(t)$. This is not straightforward since Lemma 1.4.10 only guarantees this convergence up to subsequences (but this is not a problem, we could extract once more) and for a.e. t and we do not know if these t are concerned. Here comes the assumption $\bar{\mu}(\{t\}) = 0$. Recall the measures μ_N defined in the proof of Lemma 1.4.10: they were defined via $\mu_N(I) = \int TV(\gamma; I) dQ_N(\gamma)$ but we can say

$$\mu_N(I) = \sum_{k: t_k^N, t_{k+1}^N \in I} \int \mathbb{1}_{\gamma(t_k^N) \neq \gamma(t_{k+1}^N)} dQ_N(\gamma) = \sum_{k: t_k^N, t_{k+1}^N \in I} \|\rho(t_k^N) - \rho(t_{k+1}^N)\|_{\mathcal{M}} \leq \bar{\mu}(I).$$

This implies that any narrow limit μ of a subsequence of the μ_N should satisfy $\mu \leq \bar{\mu}$. In particular, any t which is not an atom for $\bar{\mu}$ is not an atom for μ neither, and the convergence $e_t \# Q_N \xrightarrow{*} e_t \# Q$ holds along such a subsequence.

We then obtained the existence of a measure Q which, by semi-continuity of S , satisfies

$$\int_{\mathcal{C}} S(\gamma) \leq \frac{1}{2} \mathbf{L}(\rho) \quad \text{and } e_t \# Q = \rho_t \quad \text{for all } t \in E,$$

where E is a dense set. We now want to extend this to any $t \in (-1, T]$ using left-continuity. We remind that we made the choice to use left-continuous curves and that we compute the evaluations e_t using such representatives. In particular, we have, when $s \rightarrow t^-$, the narrow convergence $e_s \# Q \xrightarrow{*} e_t \# Q$. This can be seen by testing against a test function $\phi \in C_b(\Omega)$:

$$\int_{\mathcal{C}} \phi(\gamma(s)) dQ(\gamma) \xrightarrow{s \rightarrow t^-} \int_{\mathcal{C}} \phi(\gamma(t)) dQ(\gamma),$$

this convergence being justified by the dominated convergence $\phi(\gamma(s)) \rightarrow \phi(\gamma(t))$, which itself comes from the left-continuity of γ . We also have, again for $s \rightarrow t^-$, the convergence $\rho(s) \xrightarrow{*} \rho(t)$, which is due to the assumption that ρ is left-continuous.

As a consequence, approximating an arbitrary $t \in (-1, T]$ with $s < t$, $s \in E$, we obtain $e_t \# Q = \rho(t)$ and this concludes the proof. \square

We then obtain the following equality.

Corollary 2.1.6. *For every curve $t \mapsto \rho(t)$ valued in $\mathcal{P}(\Omega)$ and left-continuous we have*

$$\frac{1}{2} \mathbf{L}(\rho) = \inf \left\{ \int_{\mathcal{C}} S(\gamma) dQ(\gamma) : Q \in \mathcal{P}(\mathcal{C}), e_t \# Q = \rho(t) \text{ for all } t \right\}.$$

Proof. The left-hand side is bounded by the right-hand side using Proposition 2.1.2 and the converse inequality is deduced from Proposition 2.1.5. \square

This allows to state the equivalence, both in the jump case and in the kinetic case, of a Lagrangian and an Eulerian problem.

Theorem 2.1.7. *Given $V : [0, T] \times \Omega \rightarrow \mathbb{R}$, an action A equal either to S or to K , and a function $I : \mathcal{P}(\Omega) \rightarrow [0 + \infty]$, the Lagrangian problem*

$$\min_Q \left\{ \mathcal{U}(Q) := \int_{\mathcal{C}} A(\gamma) dQ(\gamma) + \int_0^T I(e_t \# Q) dt + \int_{\mathcal{C}} \int_0^T V(t, \gamma(t)) dt dQ(\gamma) + \int_{\mathcal{C}} \psi(\gamma(T)) dQ(\gamma) \right\}$$

is equivalent to the following Eulerian problem

$$\min_{\rho} \left\{ \mathbf{U}(\rho) := \mathbf{A}(\rho) + \int_0^T I(\rho(t)) dt + \int_0^T \int_{\Omega} V(t, x) d\rho(t) dt + \int_{\Omega} \psi d\rho(T) \right\},$$

where the minimization is performed among left-continuous curves $t \mapsto \rho(t)$ with $\rho(0) = m_0$, and the action functional \mathbf{A} is given by $\mathbf{A} = \frac{1}{2} \mathbf{L}$ if $A = S$ or $\mathbf{A}(\rho) = \int_0^T \frac{1}{p} |\rho'|_{W_p}(t)^p dt$ if $A = K$.

The equivalence means that the minimal values are the same and that from an optimal Q we can find an optimal ρ by taking $\rho(t) = e_t \# Q$ and from an optimal ρ we can find an optimal Q by taking the one provided by Proposition 2.1.5.

Proof. Take an optimal \bar{Q} and construct $\bar{\rho}$ as described in the statement. Since all terms but the first one in the functional only depend on the measures $e_t \# Q$, then clearly we have, using Proposition 2.1.2,

$$\min \mathbf{U} \leq \mathbf{U}(\bar{\rho}) \leq \mathcal{U}(\bar{Q}) = \min \mathcal{U}. \quad (2.3)$$

Then, take an optimal $\bar{\rho}$ and construct \bar{Q} via Proposition 2.1.5. The same proposition also shows the inequality needed to obtain

$$\min \mathcal{U} \leq \mathcal{U}(\bar{Q}) \leq \mathbf{U}(\bar{\rho}) = \min \mathbf{U}. \quad (2.4)$$

Putting together (2.3) and (2.4), we obtain at the same time the equality of the minimal values and the optimality of $\bar{\rho}$ in (2.3) and \bar{Q} in (2.4). \square

2.2 Banach fixed point

In Section 1.4 we presented the fixed point problem where we look for a measure \bar{Q} which minimizes the functional $\mathcal{U}_{\bar{Q}}$ and in Chapter 3 we will explain how to prove that such a minimizer is indeed concentrated on curves optimizing $J_{\bar{Q}}$, i.e. it is a Nash equilibrium. In Section 2.1 we explained how to translate the optimization problem from the Lagrangian language to the Eulerian one.

We want now to consider the fixed-point problem in Eulerian language. First, we define an operator \mathcal{F} on the set of curves valued into $\mathcal{P}(\Omega)$ via

$$\mathcal{F}(\bar{\rho}) := \operatorname{argmin}_{\rho} \mathbf{A}(\rho) + \int_0^T I(\rho(t)) dt + \int_0^T \int_{\Omega} V(t, \bar{\rho}(t), \cdot) d\rho(t) dt + \int_{\Omega} \psi d\rho(T),$$

where the minimization is performed among left-continuous curves $t \mapsto \rho(t)$ with $\rho(0) = m_0$. Note that, the minimizer is unique as soon as I is strictly convex. Then, thanks to the results of the previous section, we look for a curve $\bar{\rho}$ such that $\bar{\rho} = \mathcal{F}(\bar{\rho})$.

We will consider the case where I is given by

$$I(\rho) = \begin{cases} \int_{\Omega} f(\rho(x)) dx & \text{if } \rho \ll \mathcal{L}^d, \\ +\infty & \text{if not,} \end{cases}$$

which is lower semi-continuous for the weak-* convergence whenever f is convex and superlinear. In this case, the function dI/dm equals $f'(\rho)$ and is thus a priori only defined a.e.

More generally, we study the dependence of the optimizer in terms of the potential appearing in the linear part of the minimization problem. Let us define a map \mathcal{O} associating with every potential $V : [0, T] \times \Omega \rightarrow \mathbb{R}$ the solution ρ of

$$\min_{\rho} \mathbf{A}(\rho) + \int_0^T \int_{\Omega} f(\rho(t, x)) dx dt + \int_0^T \int_{\Omega} V(t, x) \rho(t, x) dx dt + \int_{\Omega} \psi d\rho(T).$$

We will prove that this map is Lipschitz continuous for the L^2 norm in time-space and, for this, we first need to recall the notion of proximal operator.

Definition 2.2.1. *Given a Hilbert space H and a function $h : H \rightarrow (-\infty, \infty]$, the proximal operator of h is defined by*

$$\operatorname{prox}_h(v) = \operatorname{argmin}_{u \in H} \left\{ h(u) + \frac{1}{2} \|u - v\|^2 \right\}.$$

The prox operator has the property to be 1-Lipschitz continuous.

Proposition 2.2.2 (nonexpansivity property). *Let h be a proper closed and convex function. Then for any $v_1, v_2 \in H$ we have*

$$\| \operatorname{prox}_h(v_1) - \operatorname{prox}_h(v_2) \| \leq \|v_1 - v_2\|.$$

Proof. See for example [Bec17, Theorem 6.42] or the seminal paper [Mor65]. \square

We will prove that the operator \mathcal{O} defined by

$$\mathcal{O}(V) := \operatorname{argmin}_{\rho} \mathbf{A}(\rho) + \int_0^T \int_{\Omega} f(\rho(t, x)) dx dt + \int_0^T \int_{\Omega} V(t, x) \rho(t, x) dx dt + \int_{\Omega} \psi d\rho(T)$$

is Lipschitz continuous for the L^2 norm by showing that it can be seen as a proximal operator of a convex function. First, we analyze the convexity of \mathbf{A} . We recall the definition of \mathbf{A} in the two cases

$$\text{(jump case) } \mathbf{A}(\rho) = \mathbf{L}(\rho) \quad \text{(length for the } \mathcal{M} \text{ norm);}$$

$$\text{(kinetic case) } \mathbf{A}(\rho) = \int_0^T \frac{1}{p} |\rho'|_{W_p}(t)^p dt = \inf_{v : \partial_t \rho + \nabla \cdot (\rho v) = 0} \int_0^T \int_{\Omega} \frac{1}{p} |v|^p d\rho(t) dt \quad \text{(action in the } W_p \text{ space).}$$

Lemma 2.2.3. *Given $\rho_0, \rho_1 : [0, T] \rightarrow \mathcal{P}(\Omega)$ two curves of measures, we have*

$$\mathbf{A}(\rho_\lambda) \leq (1 - \lambda)\mathbf{A}(\rho_0) + \lambda\mathbf{A}(\rho_1),$$

where $\rho_\lambda(t) := (1 - \lambda)\rho_0(t) + \lambda\rho_1(t)$.

Proof. We have in both the jump and the kinetic case $\mathbf{A}(\rho) = \min\{\int_{\mathcal{C}} A(\gamma)dQ(\gamma) : e_t\#Q = \rho(t)\}$. Take Q_0, Q_1 optimal in this definition for ρ_0, ρ_1 , respectively. Define $Q_\lambda := (1 - \lambda)Q_0 + \lambda Q_1$. We have then $e_t\#Q_\lambda = \rho_\lambda(t)$ and hence

$$\mathbf{A}(\rho_\lambda) \leq \int_{\mathcal{C}} A(\gamma)dQ_\lambda(\gamma) = (1 - \lambda)\mathbf{A}(\rho_0) + \lambda\mathbf{A}(\rho_1).$$

□

We can now use these notions to establish the Lipschitz dependence of the optimal ρ in terms of the data V . Before stating the main estimates, we need to be clear about what is meant by *Lipschitz constant*. In the following, we will use maps defined on a subset of $L^2([0, T] \times \Omega)$ and valued into $L^2([0, T] \times \Omega)$, and denote by $\text{Lip}_{L^2([0, T] \times \Omega)}$ its Lipschitz constant w.r.t. to the L^2 norm in space-time; we will also later use maps defined on a subset of $L^2(\Omega)$ and valued into $L^2(\Omega)$, and denote by $\text{Lip}_{L^2(\Omega)}$ its Lipschitz constant w.r.t. to the L^2 norm in space only. It will also happen later that we consider curves $t \mapsto \rho(t)$ and we will consider their Lipschitz constant as a map from time to $L^2(\Omega)$, which will be denoted by Lip_t .

Proposition 2.2.4. *Suppose that $\Psi_0 : L^1(\Omega) \rightarrow \mathbb{R}$ and $\Psi_T : L^1(\Omega) \rightarrow \mathbb{R}$ are convex and lower semi-continuous on $L^1(\Omega)$. Suppose that $f : \mathbb{R} \rightarrow \mathbb{R}$ is c_0 -convex, i.e. suppose that $s \mapsto f(s) - \frac{c_0}{2}s^2$ is convex. Then, the map $\mathcal{O} : L^2([0, T] \times \Omega) \rightarrow L^2([0, T] \times \Omega)$ defined via*

$$\mathcal{O}(V) := \operatorname{argmin}_\rho \mathbf{A}(\rho) + \int_0^T \int_\Omega (V\rho + f(\rho))dxdt + \Psi_0(\rho(0)) + \Psi_T(\rho(T))$$

satisfies

$$\text{Lip}_{L^2([0, T] \times \Omega)}(\mathcal{O}) \leq \frac{1}{c_0}.$$

Note that the functional Ψ_0 can encode the constraint $\rho(0) = m_0$.

Proof. Let us rewrite the functional \mathcal{F} as a proximal operator. Since f is c_0 -convex, we note by g the function such that $f(s) = g(s) + \frac{c_0}{2}s^2$. We then denote by \mathbf{G} the function defined via

$$\mathbf{G}(\rho) := \mathbf{A}(\rho) + \int_0^T \int_\Omega g(\rho)dxdt + \Psi_0(\rho(0)) + \Psi_T(\rho(T))$$

and we note that \mathbf{G} is convex. We have the following equalities:

$$\begin{aligned} \mathcal{O}(V) &= \operatorname{argmin}_\rho \mathbf{G}(\rho) + \int_0^T \int_\Omega (V\rho + \frac{c_0}{2}\rho^2)dxdt \\ &= \operatorname{argmin}_\rho \mathbf{G}(\rho) + \int_0^T \int_\Omega \frac{c_0}{2} \left| \rho + \frac{V}{c_0} \right|^2 dxdt \\ &= \operatorname{argmin}_\rho \mathbf{G}(\rho) + \frac{c_0}{2} \left\| \rho + \frac{V}{c_0} \right\|_{L^2([0, T] \times \Omega)}^2 \\ &= \operatorname{prox}_{\mathbf{G}/c_0} \left(\frac{-V}{c_0} \right). \end{aligned}$$

Thus, by the 1-Lipschitz property of the proximal operator displayed in Proposition 2.2.2 we obtained the desired estimate. □

Thanks to these results, we can bound the Lipschitz constant of \mathcal{F} .

Proposition 2.2.5. *Suppose that for each $t \in [0, T]$ the map $\rho \mapsto F(t, \rho, \cdot)$ is L -Lipschitz continuous on $L^2(\Omega)$. Then we have*

$$\text{Lip}_{L^2([0, T] \times \Omega)}(\mathcal{F}) \leq \frac{L}{c_0}.$$

Proof. Let us take ρ_0, ρ_1 and define, for $i = 0, 1$, the functions $V_i(t, x) := F(t, \rho_i(t), x)$. We have $\mathcal{F}(\rho_i) = \mathcal{O}(V_i)$, hence

$$\|\mathcal{F}(\rho_0) - \mathcal{F}(\rho_1)\|_{L^2([0, T] \times \Omega)} = \|\mathcal{O}(V_0) - \mathcal{O}(V_1)\|_{L^2([0, T] \times \Omega)} \leq \frac{1}{c_0} \|V_0 - V_1\|_{L^2([0, T] \times \Omega)}.$$

We then compute

$$\begin{aligned} \|V_0 - V_1\|_{L^2([0, T] \times \Omega)}^2 &= \int_0^T \|V_0(t, \cdot) - V_1(t, \cdot)\|_{L^2(\Omega)}^2 dt \\ &\leq \int_0^T L^2 \|\rho_0(t, \cdot) - \rho_1(t, \cdot)\|_{L^2(\Omega)}^2 dt = L^2 \|\rho_0 - \rho_1\|_{L^2([0, T] \times \Omega)}^2 \end{aligned}$$

which allows to conclude. \square

As a consequence, we obtain the following

Theorem 2.2.6. *If the assumptions of Proposition 2.2.4 hold and if $L < c_0$, then \mathcal{F} admits a unique fixed point which is the limit of any sequence $(\mathcal{F}^n(\rho_0))_n$ with $\rho_0 \in L^2([0, T] \times \Omega)$.*

Proof. By Proposition 2.2.5 and $\frac{L}{c_0} < 1$, \mathcal{F} is a contraction. By Banach's fixed point theorem, there exists a unique $\bar{\rho}$ such that $\mathcal{F}(\bar{\rho}) = \bar{\rho}$ and such a fixed point is the limit of the sequence $(\mathcal{F}^n(\rho_0))_n$ whatever is ρ_0 . \square

Some examples fit into the framework of the above theorem.

- If $F(t, \rho, x) = V_0(t, x) + g(\eta * \rho(x))$ then we can use $L = \text{Lip}(g) \|\eta\|_{L^1}$;
- If $F(t, \rho, x) = V_0(t, x) + u_\rho$, where u_ρ is the viscosity solution of $|\nabla u| = g(\eta * \rho)$ with $u = 0$ on $\partial\Omega$ (for a given non-decreasing function g) then we can use $L = \text{Lip}(\log g) \|\eta\|_{L^2} g(\|\eta\|_{L^\infty})$;
- the last example can be made more explicit when $\Omega = [a, b]$ in 1D, and we can take $F(t, \rho, x) = V_0(t, x) + u_\rho$, where u_ρ is given by $u_\rho(x) = \min\{\int_a^x g(\rho(y)) dy; \int_x^b g(\rho(y)) dy\}$ or we can replace, in the integrals, ρ with $\eta * \rho$ (this is not strictly necessary since in dimension one $\rho \in L^2$ already guarantees $u_\rho \in C^0$, but without the convolution this example does not fit the assumptions of Section 1.4). Then we can use $L = \text{Lip}(g) \frac{(b-a)}{\pi}$, where the constant π comes from the sharp Poincaré constant on $[a, b]$. This example will be treated in Section 5.3.

Many other examples can be cooked up with convolutions or similar tools. We also want to underline an example which actually does not fit our general framework but could be treated similarly. Indeed, it is possible to also consider the case where $V(t, x)$ actually depends on the whole history $(\rho(t))_t$ and not only on $\rho(t)$. An interesting example is the following one:

$$F_{[\rho]}(t, x) = \frac{1}{t} \int_0^t \rho(s, x) ds.$$

In this example, the running cost paid by an agent at time t depends on the average of the density she saw at the same point in the past. Of course it is also possible to add an exogenous cost V_0 thus getting

$$F_{[\rho]}(t, x) = V_0(t, x) + \frac{1}{t} \int_0^t \rho(s, x) ds$$

and making the problem non-autonomous. It is not possible to treat this MFG using Proposition 2.2.5 but one can directly look at the Lipschitz dependence of F in terms of ρ in the time-space norm $L^2([0, T] \times \Omega)$. Thanks to the Hardy inequality (see [Har20; HLP88]) we have

$$\|F_0 - F_1\|_{L^2([0, T] \times \Omega)} \leq 2 \|\rho_0 - \rho_1\|_{L^2([0, T] \times \Omega)},$$

so that the map \mathcal{F} is a contraction as soon as $c_0 > 2$.

We finish this section with an observation which is specific to the jump case $A = S$. Indeed, in this setting we know from [DS23] (or Chapter 4) that the solution to any variational problem of the form

$$\min_{\rho} \mathbf{A}(\rho) + \int_0^T \int_{\Omega} f(\rho(t, x)) dx dt + \int_0^T \int_{\Omega} V(t, x) \rho(t, x) dx dt + \int_{\Omega} \psi d\rho(T)$$

is Lipschitz in time valued in $L^2(\Omega)$ whenever $t \mapsto V(t, \cdot)$ is Lipschitz in time valued in the same space and f is uniformly convex. More precisely we have

$$\text{Lip}_t(\rho) = \sup_{t \in [0, T]} \|\dot{\rho}(t, \cdot)\|_{L^2(\Omega)} \leq \frac{\text{Lip}_t(V)}{c_0}. \quad (2.5)$$

We then obtain the following result

Proposition 2.2.7. *Suppose $F(t, \rho, x) = V_0(t, x) + F_1(\rho, x)$ and suppose that the map $\rho \mapsto F_1(\rho, \cdot)$ is L -Lipschitz continuous on $L^2(\Omega)$ and that V_0 is Lipschitz continuous valued into $L^2(\Omega)$. Suppose that f is c_0 -convex and $L < c_0$. Then the unique fixed point satisfies*

$$\text{Lip}_t(\bar{\rho}) \leq \frac{\text{Lip}_t V_0}{c_0 - L}.$$

In particular, if F does not depend explicitly on time (i.e. $V_0 = 0$), then $\bar{\rho}$ is constant in time.

Proof. Using (2.5) we obtain

$$\text{Lip}_t(\bar{\rho}) \leq \frac{1}{c_0} \text{Lip}_t(V), \quad \text{where } V(t, x) = V_0(t, x) + F_1(\bar{\rho}(t), x).$$

We then have

$$\text{Lip}_t V \leq \text{Lip}_t V_0 + \text{Lip}_t(t \mapsto F_1(\bar{\rho}(t), \cdot)) \leq \text{Lip}_t V_0 + L \cdot \text{Lip}_t(\bar{\rho}).$$

Using $L < c_0$ allows to obtain the claim. □

Chapter 3

Equivalence between optimizers and Nash equilibria

In Section 1.4, we showed that there exists a fixed point, i.e. a measure $\bar{Q} \in \Gamma := \{Q \in \mathcal{P}_{m_0}(\mathcal{C}) : \int_{\mathcal{C}} A(\gamma) dQ(\gamma) \leq C\}$ which minimizes $\mathcal{U}_{\bar{Q}}$. However, this does not prove yet that \bar{Q} is a Nash equilibrium in the sense that it is supported on the set of optimal trajectories for the individual cost $J_{\bar{Q}}$. The difficulty in proving this fact comes from the variational part dI/dm . What we need to prove is that, for a fixed function $V : [0, T] \times \Omega \rightarrow \mathbb{R}$, whenever V is exogenous or depends on a fixed measure \bar{Q} such as $(t, x) \mapsto F(t, x, e_t \# \bar{Q})$, the measure Q which minimizes

$$\mathcal{U}(Q) = \int_{\mathcal{C}} A(\gamma) dQ(\gamma) + \int_0^T I(e_t \# Q) dt + \int_{\mathcal{C}} \int_0^T V(t, \gamma(t)) dt dQ(\gamma) + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ(\gamma) \quad (3.1)$$

is concentrated on curves which minimize

$$J_W(\gamma) := A(\gamma) + \int_0^T W(t, \gamma(t)) dt + \psi_T(\gamma(T)) \quad \text{for } W(t, x) = \frac{dI}{dm}[e_t \# Q](x) + V(t, x).$$

The result will be proven in the case where we take

$$I(\rho) = \begin{cases} \int_{\Omega} f(\rho(x)) dx & \text{if } \rho \ll \mathcal{L}^d, \\ +\infty & \text{if not.} \end{cases} \quad (3.2)$$

This functional is lower semi-continuous for the weak-* convergence whenever f is convex and super-linear.

In this case, the function dI/dm equals $f'(\rho)$ and is thus a priori only defined a.e. In general, this requires to choose a precise representative of it, but a first result that we will show will concern the case where *by chance* $f'(\rho)$ is a continuous function.

The regularity of the density ρ can be studied as an independent question and is better analyzed if the variational problem solved by Q is re-written as a problem involving ρ , hence in Eulerian terms. The equivalence between these two approaches were addressed in the previous chapter in the more delicate case $A = S$ (the jump case).

In this case we can see that minimizing \mathcal{U} is equivalent to solving

$$\min_{\rho} \int_0^T \int_{\Omega} (|\partial_t \rho| + V\rho + f(\rho)) dx dt + \Psi_0(\rho(0)) + \int_{\Omega} \psi_T d\rho(T) \quad (3.3)$$

for a suitable $\Psi_0 : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$. This allows to use the results in [DS23] (or in Chapter 4), which indeed provide regularity of the optimal ρ (under suitable assumptions, the solution $\bar{\rho}$ of (3.3) is Lipschitz in time valued in $L^2(\Omega)$ and its regularity in space depends on the regularity of V).

Let us begin with the next proposition which provides the optimality condition for Problem (3.1).

Proposition 3.0.1. *Let \bar{Q} be a solution of Problem (3.1) in the case where I is given by (3.2). Let $Q \in \Gamma$ be a measure such that, setting $\rho(t, \cdot) = e_t \# Q$, we have*

$$\int_0^T \int_{\Omega} f(\rho(t, x)) dx dt < \infty. \quad (3.4)$$

Then we have

$$\begin{aligned} & \int_{\mathcal{C}} A(\gamma) d\bar{Q}(\gamma) + \int_0^T \int_{\Omega} (V(t, x) + f'(\bar{\rho}(t, x))) \bar{\rho}(t, x) dx dt + \int_{\mathcal{C}} \psi(\gamma(T)) d\bar{Q}(\gamma) \\ & \leq \int_{\mathcal{C}} A(\gamma) dQ(\gamma) + \int_0^T \int_{\Omega} (V(t, x) + f'(\bar{\rho}(t, x))) \rho(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ(\gamma). \end{aligned} \quad (3.5)$$

Proof. Given $\varepsilon \in [0, 1]$, let us define $Q_\varepsilon = \bar{Q} + \varepsilon(Q - \bar{Q})$ and

$$u(\varepsilon) = \int_{\mathcal{C}} A(\gamma) dQ_\varepsilon(\gamma) + \int_0^T \int_{\Omega} V(t, x) d(e_t \# Q_\varepsilon)(x) dt + \int_0^T \int_{\Omega} f(e_t \# Q_\varepsilon)(x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ_\varepsilon(\gamma).$$

By the optimality of \bar{Q} , we have $u(\varepsilon) \geq u(0)$. We want to compute $u'(0)$. All terms in u are affine in ε except for $\varepsilon \mapsto \int_0^T \int_{\Omega} f(e_t \# Q_\varepsilon)(x) dx dt = \int_0^T \int_{\Omega} f((1 - \varepsilon)\bar{\rho} + \varepsilon\rho) dx dt$ (notice $\rho_t := e_t \# Q$ and $\bar{\rho}_t := e_t \# \bar{Q}$). Yet, the convexity of f provides

$$\frac{f((1 - \varepsilon)\bar{\rho} + \varepsilon\rho) - f(\bar{\rho})}{\varepsilon} \leq f(\rho) - f(\bar{\rho}).$$

The limit as $\varepsilon \rightarrow 0^+$ of the l.h.s. is $f'(\bar{\rho})(\rho - \bar{\rho})$. The above inequality, together with the integrability condition (3.4), show in particular that we have $f'(\bar{\rho})(\rho - \bar{\rho}) < +\infty$ a.e. and that the positive part of this function is L^1 . Then, we can obtain by Fatou's Lemma

$$\int f'(\bar{\rho})(\rho - \bar{\rho}) \geq \lim_{\varepsilon \rightarrow 0} \int \frac{f((1 - \varepsilon)\bar{\rho} + \varepsilon\rho) - f(\bar{\rho})}{\varepsilon}.$$

Therefore we have:

$$\begin{aligned} 0 \leq u'(0) & \leq \int_{\mathcal{C}} A(\gamma) d(Q - \bar{Q})(\gamma) + \int_0^T \int_{\Omega} V(t, x) d(e_t \# (Q - \bar{Q}))(x) dt \\ & \quad + \int_0^T \int_{\Omega} f'(e_t \# \bar{Q})(x) d(e_t \# (Q - \bar{Q}))(x) dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) d(Q - \bar{Q})(\gamma). \end{aligned}$$

This inequality is exactly (3.5). \square

We first re-write the inequality (3.5) in a more suitable way using the notation J_W . We observe that we have, for any $Q \in \Gamma$ with $\rho_t := e_t \# Q$,

$$\int_{\mathcal{C}} A(\gamma) dQ(\gamma) + \int_0^T \int_{\Omega} W(t, x) \rho(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ(\gamma) = \int_{\mathcal{C}} J_W(\gamma) dQ(\gamma).$$

This means that the Proposition 3.0.1 provides the inequality

$$\int_{\mathcal{C}} J_W(\gamma) d\bar{Q}(\gamma) \leq \int_{\mathcal{C}} J_W(\gamma) dQ(\gamma)$$

for $W = V + f'(\bar{\rho})$ and any Q satisfying the condition $\int_0^T \int_{\Omega} f(\rho(t, x)) dx dt < +\infty$ (where $\rho_t := e_t \# Q$). Note that because both \bar{Q} and Q are such that the measures $\bar{\rho}_t$ and ρ_t are absolutely continuous for every t , modifying W into another function which is equal to W a.e. does not change the validity of this inequality. On the other hand, what we would like is to prove that \bar{Q} is actually concentrated on curves which minimize J_W starting from their starting point. This condition, instead, could depend on the choice of the representative of W since curves are negligible.

We prove now an integral characterization of this condition.

Lemma 3.0.2. *Given a measurable function $W : [0, T] \times \Omega \rightarrow \mathbb{R}$, a measure $\bar{Q} \in \mathcal{P}_{m_0}(\mathcal{C})$ is concentrated on the curves γ such that the following conditions holds*

$$\forall \omega \in \mathcal{C} \text{ s.t. } \gamma(0) = \omega(0), J_W(\gamma) \leq J_W(\omega)$$

if and only if it satisfies

$$\int_{\mathcal{C}} J_W d\bar{Q} \leq \int_{\mathcal{C}} J_W dQ \quad \text{for every } Q \in \mathcal{P}_{m_0}(\mathcal{C}). \quad (3.6)$$

Proof. Let us define a value function φ through $\varphi(x_0) := \inf_{\omega: \omega(0)=x_0} J_W(\omega)$.

First, we suppose that \bar{Q} satisfies (3.6). Then, for every $\varepsilon > 0$ we can choose for every x_0 a curve ω_{x_0} such that $\omega_{x_0}(0) = x_0$ and $J_W(\omega_{x_0}) \leq \varphi(x_0) + \varepsilon$. Choosing the map $x_0 \mapsto \omega_{x_0}$ in a measurable way (this is possible by measurable selection arguments, for which we refer to the seminal book [CV77]) and taking the image of m_0 through such a map we obtain a measure $\tilde{Q} \in \mathcal{P}_{m_0}(\mathcal{C})$ such that

$$\int_{\mathcal{C}} J_W d\tilde{Q} \leq \int \varphi(\gamma(0)) d\tilde{Q}(\gamma) + \varepsilon = \int_{\Omega} \varphi dm_0 + \varepsilon.$$

We then have

$$\int_{\mathcal{C}} J_W d\bar{Q} \leq \int_{\mathcal{C}} J_W d\tilde{Q} \leq \int_{\Omega} \varphi + \varepsilon.$$

The number ε being arbitrarily small, we also obtain $\int_{\mathcal{C}} J_W d\bar{Q} \leq \int_{\Omega} \varphi$. Yet, we have $J_W(\gamma) \geq \varphi(\gamma(0))$, which shows that this inequality is an equality. Then, we can see that \bar{Q} is concentrated on curves satisfying $J_{\bar{Q}}(\gamma) = \varphi(\gamma(0))$, i.e. on optimal curves, which proves the claim.

For the converse implication, suppose that \bar{Q} is concentrated on optimal curves. Then we have, for \bar{Q} -a.e. γ , $J_W(\gamma) = \varphi(\gamma(0))$ and $\int_{\mathcal{C}} J_W d\bar{Q} = \int_{\mathcal{C}} \varphi(\gamma(0)) d\bar{Q} = \int_{\Omega} \varphi dm_0$. On the other hand, for any other measure Q we have $\int_{\mathcal{C}} J_W dQ \geq \int_{\mathcal{C}} \varphi(\gamma(0)) dQ = \int_{\Omega} \varphi dm_0$, which proves (3.6). \square

The goal becomes now to choose a function W which is a precise representative of the function $V + f'(\bar{\rho})$ and for which we have the validity of (3.6). The main difficulty is to prove that this inequality is valid for every Q and not only for those satisfying the integrability condition $\int_0^T \int_{\Omega} f(\rho(t, x)) dx dt < +\infty$.

We will do this in the two cases that we are considering, i.e. the jump case and the kinetic case. We start from the jump case where, under suitable conditions, it is possible to prove, thanks to regularity results contained in [DS23], that $\bar{\rho}$ is continuous. In this case there is no ambiguity, and we can take $W = V + f'(\bar{\rho})$ just choosing such a continuous representative.

Before analyzing the two cases and proving that optimizers are indeed equilibria in the sense that we explained, we want to explain why this chapter is called “equivalence between equilibria and optimizers” while we only prove that optimizers are equilibria. Indeed, the converse implication is straightforward because of the convexity of the functional \mathcal{U} . Proposition 3.0.1 provides necessary optimality conditions for the minimization of \mathcal{U} , consisting in inequality (3.5) which has to be satisfied for those Q for which (3.4) holds. Actually, because of convexity, this necessary optimality condition is also sufficient, and every equilibrium satisfies it (and even more, because it satisfies (3.5) even without (3.4)).

3.1 The jump case

Let us consider the case $A(\gamma) = S(\gamma)$ defined in Section 1.4.1. It is important to have in mind the results of [DS23], where we proved that the solution $\bar{\rho}$ of the problem (3.3) is Lipschitz in time and that according to the regularity of V and the conditions at time 0 and T , $\bar{\rho}$ can be either continuous or bounded. Because of these results, we will provide statements under the additional assumption that $\bar{\rho}$ is either continuous or bounded. In order to illustrate the strategy, we start from the case where it is continuous.

Theorem 3.1.1. *Suppose that Ω is either the d -dimensional torus or a compact Lipschitz domain in \mathbb{R}^d . Let \bar{Q} be such that (3.5) holds for all Q satisfying (3.4). Define $\bar{\rho}_t := e_t \# \bar{Q}$, suppose that the measures $\bar{\rho}_t$ are all absolutely continuous, and call $\bar{\rho}(t, x)$ their densities. Suppose that the function $\bar{\rho}: [0, T] \times \Omega \rightarrow \mathbb{R}$ is continuous (and hence bounded). Then we have*

$$\begin{aligned} \forall Q \in \Gamma, \int_{\mathcal{C}} S(\gamma) d\bar{Q}(\gamma) + \int_0^T \int_{\Omega} (V(t, x) + f'(\bar{\rho}(t, x))) \bar{\rho}(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) d\bar{Q}(\gamma) \\ \leq \int_{\mathcal{C}} S(\gamma) dQ(\gamma) + \int_0^T \int_{\Omega} (V(t, x) + f'(\bar{\rho}(t, x))) \rho(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ(\gamma). \end{aligned} \quad (3.7)$$

We observe that, in this statement, the continuity of $\bar{\rho}$ in both variables is not really needed, and continuity in x for a.e. t together with a uniform L^∞ bound would be enough.

Proof. Let us start from the case where Ω is the torus. Let Q be in Γ . Let $r > 0$. For all $y \in B_r = B(0, r) \subset \mathbb{R}^d$, we define the function

$$T_y: \mathcal{C} \rightarrow \mathcal{C}, \gamma \mapsto \left(t \mapsto \begin{cases} \gamma(t), & \text{if } t \leq t_0(\gamma), \\ \gamma(t) + y, & \text{if } t > t_0(\gamma), \end{cases} \right)$$

where $t_0(\gamma)$ is the time at which the curve γ has its first jump or discontinuity. This function is a translation of a vector y of the curves in Ω starting from the time the curve jumps (note that this very

construction requires to use the torus, otherwise $\gamma(t) + y$ could be outside the domain; the general case of other domains will be considered later). This construction leaves the curves unchanged at time 0, i.e

$$\forall \gamma \in \mathcal{C}, T_y(\gamma)(0) = \gamma(0).$$

In addition, by the construction of T_y , the function T_y does not add new jumps to the curves, so the following inequality holds:

$$S(T_y(\gamma)) \leq S(\gamma). \quad (3.8)$$

For fixed $r > 0$, we define the measure Q_r such that

$$Q_r = \int_{B_r} T_y \# Q dy,$$

so the inequality (3.8) implies

$$\int_{\mathcal{C}} S(\gamma) dQ_r(\gamma) \leq \int_{\mathcal{C}} S(\gamma) dQ(\gamma) \leq C. \quad (3.9)$$

Next, we have for all $\varphi \in C_b(\Omega)$,

$$\int_{\Omega} \varphi(x) d(e_0 \# Q_r)(x) = \int_{\mathcal{C}} \int_{B_r} \varphi(\gamma(0)) dy dQ(\gamma) = \int_{\Omega} \varphi(x) dm_0(x),$$

so $e_0 \# Q_r = m_0$, which means that the initial condition is verified.

For each t , let us define the subsets

$$\mathcal{A}(t) = \{\gamma \in \mathcal{C}; t \leq t_0(\gamma)\} \quad (3.10)$$

$$\text{and } \mathcal{B}(t) = \{\gamma \in \mathcal{C}; t > t_0(\gamma)\}. \quad (3.11)$$

Let t be in $[0, T]$. For all positive functions $\varphi \in C_b(\Omega)$, we have

$$\begin{aligned} \int_{\Omega} \varphi(x) d(e_t \# Q_r)(x) &= \int_{\mathcal{C}} \int_{B_r} \varphi(\gamma(t)) \mathbb{1}_{\mathcal{A}(t)}(\gamma) dy dQ(\gamma) + \int_{\mathcal{C}} \int_{B_r} \varphi(\gamma(t) + y) \mathbb{1}_{\mathcal{B}(t)}(\gamma) dy dQ(\gamma) \quad (3.12) \\ &= \int_{\mathcal{A}(t)} \varphi(\gamma(0)) dQ(\gamma) + \int_{\mathcal{C}} \int_{B_r} \varphi(\gamma(t) + y) \mathbb{1}_{\mathcal{B}(t)}(\gamma) dy dQ(\gamma) \\ &\leq \int_{\Omega} \varphi(x) dm_0(x) + \int_{\mathcal{C}} \int_{B_r} \varphi(\gamma(t) + y) dy dQ(\gamma) \\ &= \int_{\Omega} \varphi(x) dm_0(x) + \int_{\Omega} \int_{\mathbb{R}^d} \frac{\varphi(x+y)}{|B_r|} \mathbb{1}_{B_r}(-y) dy d(e_t \# Q)(x) \\ &\stackrel{u=x+y}{=} \int_{\Omega} \varphi(x) dm_0(x) + \int_{\Omega} \int_{\mathbb{R}^d} \frac{\varphi(u)}{|B_r|} \mathbb{1}_{B_r}(x-u) du \rho(t, x) dx \\ &= \int_{\Omega} \varphi(x) dm_0(x) + \int_{\mathbb{R}^d} \varphi(u) \int_{\Omega} \frac{\mathbb{1}_{B_r}(u-x)}{|B_r|} \rho(t, x) dx du \\ &= \int_{\Omega} \varphi(x) dm_0(x) + \int_{\mathbb{R}^d} \varphi(u) \left(\frac{\mathbb{1}_{B_r}}{|B_r|} * \rho_t \right) (u) du. \end{aligned}$$

Therefore, for almost every $x \in \Omega$, we have

$$\rho_r(t, x) := (e_t \# Q_r)(x) \leq m_0(x) + \frac{1}{|B_r|}.$$

The right-hand side is independent of t and is of the form $m_0 + C$. By assumption, m_0 is bounded, and so is ρ_r . Hence, we obtain

$$\int_0^T \int_{\Omega} f(\rho_r(t, x)) dx dt < \infty. \quad (3.13)$$

Therefore, we can apply Proposition 3.0.1 so as to write

$$\begin{aligned} \forall r > 0, \int_{\mathcal{C}} S(\gamma) d\bar{Q}(\gamma) + \int_0^T \int_{\Omega} (V(t, x) + f'(\bar{\rho}(t, x))) \bar{\rho}(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) d\bar{Q}(\gamma) \quad (3.14) \\ \leq \int_{\mathcal{C}} S(\gamma) dQ_r(\gamma) + \int_0^T \int_{\Omega} (V(t, x) + f'(\bar{\rho}(t, x))) \rho_r(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ_r(\gamma). \end{aligned}$$

We now verify that we have $\rho_r \xrightarrow{*} \rho$ by using again the sets $\mathcal{A}(t)$ and $\mathcal{B}(t)$, so that we will be able to pass to the limit in (3.14). We fix a time $t \in [0, T]$: for all test function ϕ which are Lipschitz continuous on Ω , we have

$$\begin{aligned} \int_{\Omega} \phi d(e_t \# Q_r) &= \int_{B_r} dy \int_{\mathcal{C}} \phi(\gamma(t)) d(T_y \# Q)(\gamma) \\ &= \int_{\mathcal{A}(t)} \phi(\gamma(t)) dQ(\gamma) + \int_{\mathcal{B}(t)} \int_{B_r} \phi(\gamma(t) + y) dy dQ(\gamma). \end{aligned}$$

Since ϕ is Lipschitz continuous and Q is a probability measure, we have

$$\left| \int_{\mathcal{B}(t)} \int_{B_r} \phi(\gamma(t) + y) dy dQ(\gamma) - \int_{\mathcal{B}(t)} \phi(\gamma(t)) dQ(\gamma) \right| \leq \text{Lip}(\phi) r \rightarrow 0.$$

This shows that we have $\int_{\Omega} \phi d(e_t \# Q_r) \rightarrow \int_{\Omega} \phi d(e_t \# Q)$, i.e. $\rho_r(t) \xrightarrow{*} \rho(t)$ for every t . Then, using the continuity of V , ψ_T and $\bar{\rho}$, we finally obtain (using dominated convergence in time)

$$\begin{aligned} \lim_{r \rightarrow 0} \int_0^T \int_{\Omega} (V(t, x) + f'(\bar{\rho}(t, x))) \rho_r(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ_r(\gamma) \\ = \int_0^T \int_{\Omega} (V(t, x) + f'(\bar{\rho}(t, x))) \rho(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ(\gamma). \end{aligned} \quad (3.15)$$

Combining the inequalities (3.9) and (3.14) and the limit (3.15), we have the desired inequality:

$$\begin{aligned} \int_{\mathcal{C}} S(\gamma) d\bar{Q}(\gamma) + \int_0^T \int_{\Omega} (V(t, x) + f'(\bar{\rho}(t, x))) \bar{\rho}(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) d\bar{Q}(\gamma) \\ \leq \limsup_{r \rightarrow 0} \int_{\mathcal{C}} S(\gamma) dQ_r(\gamma) + \int_0^T \int_{\Omega} (V(t, x) + f'(\bar{\rho}(t, x))) \rho_r(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ_r(\gamma) \\ \leq \int_{\mathcal{C}} S(\gamma) dQ(\gamma) + \int_0^T \int_{\Omega} (V(t, x) + f'(\bar{\rho}(t, x))) \rho(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ(\gamma), \end{aligned}$$

which ends the proof in the case of the torus.

Let us consider now a Lipschitz domain Ω . For $r_0 > 0$ small enough, there exists a map $\Pi : \Omega_{r_0} \rightarrow \Omega$ (where $\Omega_r := \{x \in \mathbb{R}^d : d(x, \Omega) \leq r\}$) with the following properties:

- $\Pi(x) = x$ for all $x \in \Omega$;
- there exist a constant C_1 such that $|\Pi(x) - x| \leq C_1 r$ for all $x \in \Omega_r$;
- there exist a constant C_2 such that for any measure on Ω_{r_0} with density bounded by a constant M , its image through Π is a measure on Ω with density bounded by $C_2 M$.

Note that we do not need Π to be continuous and that it can be defined using charts where Ω is bounded by a graph: if locally we have $\Omega = \{x = (x', x_d) \in \mathbb{R}^d : x_d > h(x')\}$ we can define $\Pi(x', x_d) = (x', 2h(x') - x_d)$ for points $x \notin \Omega$. The whole domain can be covered by a finite number of these charts, and for points in the intersection of different charts we can choose arbitrarily which definition of Π to choose. The constant C_2 would depend on the maximal number of intersecting charts, in this case.

We replace now the map T_y used in the case of the torus with a map \tilde{T}_y defined via $\tilde{T}_y(\gamma)(t) = \Pi(T_y(\gamma)(t))$. We thus obtain a family of measures \tilde{Q}_r . Note that it is still true that we have $S(\tilde{T}_y(\gamma)) \leq S(\gamma)$. We then go on with the same procedure as before, defining $\tilde{\rho}_r(t) := e_t \# \tilde{Q}_r = \Pi_{\#}(\rho_r(t))$. The densities of the measures ρ_r were bounded by $\|m_0\|_{L^\infty} + \frac{1}{|B_r|}$, thus the densities of $\tilde{\rho}_r$ will be bounded by another constant, but the argument does not change. Moreover, we will again have $\tilde{\rho}_r(t) \xrightarrow{*} \rho(t)$. This is due to the fact that $\rho_r(t)$ is supported in Ω_r for every t , and hence for every Lipschitz test function ϕ we have $|\int \phi d\tilde{\rho}_r - \int \phi d\rho_r| \leq C \text{Lip} \phi r \rightarrow 0$. This allows to conclude the argument in the very same way as in the case of the torus. \square

We now consider a variant, where we only suppose $\bar{\rho}$ to be bounded instead of continuous. First of all, we consider the map Π defined on the set Ω_{r_0} at the end of the proof of the previous theorem and we define, for $r < r_0$ the following quantities

$$\forall x \in \Omega, [f'(\bar{\rho})]_r(t, x) := \int_{B_r} f'(\bar{\rho}(t, \Pi(x + y))) dy, \quad \widehat{f'(\bar{\rho})}(t, x) := \limsup_{r \rightarrow 0} [f'(\bar{\rho})]_r(t, x). \quad (3.16)$$

Of course the map Π is useless in the case of the torus and could be omitted for any internal point x in the definition of $\widehat{f'(\bar{\rho})}(t, x)$ and in the definition of $[f'(\bar{\rho})]_r(t, x)$ as soon as $r < d(x, \partial\Omega)$. Also note that we have $\widehat{f'(\bar{\rho})} = f'(\bar{\rho})$ if $\bar{\rho}$ is a continuous function and that in any case the equality $\widehat{f'(\bar{\rho})} = f'(\bar{\rho})$ holds a.e. (more precisely, at any interior point x which is a Lebesgue point of $f'(\bar{\rho})$; this only excludes a negligible set of points, and the boundary, which is also negligible).

Theorem 3.1.2. *Suppose that Ω is either the d -dimensional torus or a compact Lipschitz domain in \mathbb{R}^d . Let \bar{Q} be such that (3.5) holds for all Q satisfying (3.4). Suppose that the densities of all the measures $e_t\#\bar{Q}$, denoted by $\bar{\rho}(t, \cdot)$, are bounded by a common constant. Then for all $Q \in \Gamma$, we have*

$$\int_{\mathcal{C}} S(\gamma)d\bar{Q}(\gamma) + \int_0^T \int_{\Omega} \left(V(t, x) + \widehat{f'(\bar{\rho})}(t, x) \right) \bar{\rho}(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T))d\bar{Q}(\gamma) \quad (3.17)$$

$$\leq \int_{\mathcal{C}} S(\gamma)dQ(\gamma) + \int_0^T \int_{\Omega} \left(V(t, x) + \widehat{f'(\bar{\rho})}(t, x) \right) \rho(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T))dQ(\gamma). \quad (3.18)$$

Proof. The proof is similar to Theorem 3.1.1. The main difference is that the function $\widehat{f'(\bar{\rho})}$ is no longer continuous in space, but only bounded. We directly treat the case where we use the map Π , as it is now quite standard. We will omit, by the way, the symbol $\bar{\cdot}$.

We start from the same construction: for all $r > 0$, we define as in Theorem 3.1.1 the measures

$$Q_r = \int_{B_r} T_y\#Q dy.$$

As a consequence, $e_0\#Q_r = m_0$ and $\int_{\mathcal{C}} S(\gamma)dQ_r(\gamma) \leq \int_{\mathcal{C}} S(\gamma)dQ(\gamma) \leq C$. We have seen that the densities ρ_r are bounded by a constant (by $\|m_0\| + \frac{1}{|B_r|}$ in the case of the torus – note that in this case as well our assumption implies that m_0 is bounded – or by a multiple of this number in case we need to use Π). This yields

$$\int_0^T \int_{\Omega} f(\rho_r(t, x)) dx dt < \infty.$$

By applying Proposition 3.0.1, we obtain

$$\begin{aligned} \forall r > 0, \int_{\mathcal{C}} S(\gamma)d\bar{Q}(\gamma) + \int_0^T \int_{\Omega} (V(t, x) + f'(\bar{\rho}(t, x))) \bar{\rho}(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T))d\bar{Q}(\gamma) \\ \leq \int_{\mathcal{C}} S(\gamma)dQ_r(\gamma) + \int_0^T \int_{\Omega} (V(t, x) + f'(\bar{\rho}(t, x))) \rho_r(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T))dQ_r(\gamma), \end{aligned}$$

and we need to pass to the limit in the right-hand side. We use again the sets $\mathcal{A}(t)$ and $\mathcal{B}(t)$ defined in (3.10) and (3.11). The only term which has to be treated differently can be dealt with via the definition of $\widehat{f'(\bar{\rho})}$. Indeed, we have

$$\begin{aligned} \int_0^T \int_{\Omega} f'(\bar{\rho}(t, x)) \rho_r(t, x) dx dt &= \int_0^T \int_{\mathcal{A}(t)} f'(\bar{\rho}(t, \gamma(t))) dQ(\gamma) dt + \int_0^T \int_{\mathcal{B}(t)} \int_{B_r} f'(\bar{\rho}(t, \Pi(\gamma(t) + y))) dy dQ(\gamma) dt \\ &= \int_0^T \int_{\mathcal{A}(t)} f'(\bar{\rho}(t, \gamma(t))) dQ(\gamma) dt + \int_0^T \int_{\mathcal{B}(t)} [f'(\bar{\rho})]_r(t, \gamma(t)) dQ(\gamma) dt. \end{aligned}$$

The first term on the right-hand side does not depend on r . Moreover we note that we have

$$\int_0^T \int_{\mathcal{A}(t)} f'(\bar{\rho}(t, \gamma(t))) dQ(\gamma) dt = \int_0^T \int_{\mathcal{A}(t)} \widehat{f'(\bar{\rho})}(t, \gamma(t)) dQ(\gamma) dt.$$

This is due to the a.e. equality $\widehat{f'(\bar{\rho})} = f'(\bar{\rho})$ and to the fact that $e_t\#(Q\mathbb{1}_{\mathcal{A}(t)})$ is absolutely continuous (it is indeed a measure bounded from above by m_0). For the second term we can use a reverse Fatou's Lemma together with $\limsup_{r \rightarrow 0} [f'(\bar{\rho})]_r = \widehat{f'(\bar{\rho})}$. For this, we need to upper bound all these functions, but the assumption guarantees that they are all bounded by $f'(M)$, where $M := \sup \bar{\rho}$. We then obtain

$$\limsup_{r \rightarrow 0} \int_0^T \int_{\mathcal{B}(t)} [f'(\bar{\rho})]_r(t, \gamma(t)) dQ(\gamma) dt \leq \int_0^T \int_{\mathcal{B}(t)} \widehat{f'(\bar{\rho})}(t, \gamma(t)) dQ(\gamma) dt$$

and hence

$$\limsup_{r \rightarrow 0} \int_0^T \int_{\Omega} f'(\bar{\rho}(t, x)) \rho_r(t, x) dx dt \leq \int_0^T \int_{\Omega} \widehat{f'(\bar{\rho})}(t, x) \rho(t, x) dx dt.$$

The other terms (including the integral of the potential V , which is now separated from $f'(\bar{\rho})$) are treated exactly as in Theorem 3.1.1. By combining the different terms we obtain

$$\begin{aligned} & \int_{\mathcal{C}} S(\gamma) d\bar{Q}(\gamma) + \int_0^T \int_{\Omega} (V(t, x) + f'(\bar{\rho}(t, x))) \bar{\rho}(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) d\bar{Q}(\gamma) \\ & \leq \limsup_{r \rightarrow 0} \int_{\mathcal{C}} S(\gamma) dQ_r(\gamma) + \int_0^T \int_{\Omega} (V(t, x) + f'(\bar{\rho}(t, x))) \rho_r(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ_r(\gamma) \\ & \leq \int_{\mathcal{C}} S(\gamma) dQ(\gamma) + \int_0^T \int_{\Omega} (V(t, x) + \widehat{f'(\bar{\rho})}(t, x)) \rho(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ(\gamma). \end{aligned}$$

We then conclude by replacing $f'(\bar{\rho})$ with $\widehat{f'(\bar{\rho})}$ in the left-hand side, since these two functions agree a.e. and for a.e. t the measure $\bar{\rho}_t = e_t \# \bar{Q}$ is absolutely continuous. \square

We can then conclude about the equilibrium properties of the measure \bar{Q} that we obtained as a fixed point in Section 1.4.

Corollary 3.1.3. *Suppose that the measures $e_t \# \bar{Q} = \bar{\rho}(t, \cdot)$ are all bounded by a same constant. Then, the measure \bar{Q} is concentrated on the curves $\bar{\gamma}$ which are minimizers, for fixed starting point, of the action function J_W with $W = V + \widehat{f'(\bar{\rho})}$, i.e. on curves $\bar{\gamma}$ such that*

$$\begin{aligned} \forall \omega \in \mathcal{C} \text{ with } \omega(0) = \bar{\gamma}(0) \quad & S(\bar{\gamma}) + \int_0^T \left[V(t, \bar{\gamma}(t)) + \widehat{f'(\bar{\rho})}(t, \bar{\gamma}(t)) \right] dt + \psi_T(\bar{\gamma}(T)) \\ & \leq S(\omega) + \int_0^T \left[V(t, \omega(t)) + \widehat{f'(\bar{\rho})}(t, \omega(t)) \right] dt + \psi_T(\omega(T)). \end{aligned}$$

The proof is a combination of the previous results. The case where $\bar{\rho}$ is continuous is included in this same statement, as in this case we simply have $\widehat{f'(\bar{\rho})} = f'(\bar{\rho})$.

3.2 The kinetic case

In this part, we consider the case $A(\gamma) = K(\gamma)$ (see (1.30)). The goal is to perform a similar computation as in the case $A = S$, i.e. showing by a convolution argument that the inequality (3.7) actually holds for all $Q \in \Gamma$.

The regularity theory for the kinetic case is a classical issue in MFG theory (see for eg. [Muñ22] and [Por23]), and one of the strongest results is the one contained in [LS18], which proves L^∞ bounds on $\bar{\rho}$. Continuity for the density of $\bar{\rho}$ is a much harder task, so we will only rely on its boundedness. We refer to [LS18] for the proofs of this boundedness, and to [San20] for an overview of the applications of such a result to the equivalence between optimizers and equilibria. Indeed, a general strategy first introduced by [AF09] for fluid mechanics applications allowed to prove the optimality of \bar{Q} -almost every curve within a restricted class of curves on which the maximal function of $f'(\bar{\rho})$ is integrable in time. This has later been used in [CMS16] in the framework of MFG, following essentially the same ideas. As explained in [San20], the upper bounds on $\bar{\rho}$ allow to state a much simpler result, showing optimality in the class of all $W^{1,p}$ curves. The proof of [AF09] and [CMS16] is based on a multiple approximation procedure strongly relying on the separability of the space $W^{1,p}$ and on the continuity of A for the strong $W^{1,p}$ convergence. Such a procedure was not possible for the jump case, which is naturally set in the space BV which is non-separable. This motivated the different approach that we presented in the previous section. A variant of such approach allows to obtain the following result in the kinetic case. Such a result recovers already known results, and our approach requires a strong assumption on the growth of the function f in terms of the exponent p in the kinetic energy and on the space dimension d . Yet, it is interesting to observe that such a numeric assumption is not new in the setting of deterministic (first-order) MFG (see, for instance, [CG15]). Moreover, such a proof is considerably simpler than the one in [AF09; CMS16].

Compared to the theorems in the jump case, we need to require extra regularity on the domain Ω . We need indeed an extra property for the map Π : we need Π to be $(1 + Cr)$ -Lipschitz continuous on Ω_r . The reason for this is that we need now to preserve Sobolev bounds on the composition of a curve with Π . It is possible to guarantee this by requiring Ω to be C^2 and defining $\Pi(x) = 2P_\Omega(x) - x$, where P_Ω is the projection onto Ω , well-defined for smooth domains in a neighborhood of Ω . This map is a sort of reflection across the boundary. The Jacobian matrix of this map Π is symmetric and its eigenvalues are very close to -1 in the direction of $x - P_\Omega(x)$ and to 1 in the orthogonal directions. The precise bounds depend on r and on the maximal curvature of $\partial\Omega$.

Lemma 3.2.1. *Let $\Pi(x) = x + 2(P_\Omega(x) - x) = 2P_\Omega(x) - x$ be the reflexion such that $P_\Omega(x) = \operatorname{argmin}_{y \in \Omega} d(y, \Omega)$. Then we have*

$$\|D\Pi\| \leq 1 + \frac{r}{R+r} \text{ and } \det D\Pi \geq \frac{1}{2},$$

where $D\Pi$ is the Jacobian matrix of Π .

Proof. The function Π associates with a point $x \in \Omega^c$ its reflexion with respect to the boundary $\partial\Omega$. We can see that

$$\forall x \in \Omega_r, |\Pi(x) - x| = 2|P_\Omega(x) - x| \leq 2r.$$

Next, we can estimate $\|D\Pi\|$ by estimating $\|I_d - DP_\Omega\|$. Indeed, if we define $g(x) = \frac{1}{2}d^2(x, \partial\Omega)$, then we have $x - P_\Omega(x) = \nabla g(x)$. By differentiating again, we obtain $I_d - DP_\Omega = \operatorname{Hess}_g$ the Hessian matrix of g . Let $x_0 \in \Omega^c$ and $p_0 = P_\Omega(x_0)$. There exists $R > 0$ and two points $y_0 \in \Omega^c$ and $z_0 \in \Omega$ such that $B(y_0, R) \subset \Omega^c$ and $B(z_0, R) \subset \Omega$ and $p_0 \in \overline{B(y_0, R)} \cap \overline{B(z_0, R)}$. Let us note that x_0 belongs to the segment $[p_0, y_0]$ and the points x_0, p_0, y_0 and z_0 are aligned. Consequently, we have for all $x \in \Omega^c$,

$$\frac{1}{2}d^2(x, \partial B(y_0, R)) \leq g(x) \leq \frac{1}{2}d^2(x, \partial B(z_0, R)), \quad (3.19)$$

and in particular, at $x = x_0$,

$$\frac{1}{2}(R - |x_0 - y_0|)^2 = \frac{1}{2}d^2(x_0, \partial B(y_0, R)) \text{ and } \frac{1}{2}d^2(x_0, \partial B(z_0, R)) = \frac{1}{2}(|x_0 - z_0| - R)^2.$$

The Hessian of $\frac{1}{2}(R - |x - y_0|)^2$ is

$$(H^l)_{i,j}(x) = \begin{cases} 1 - R \left(\frac{1}{|x-y_0|} - \frac{(x_i - y_{0,i})^2}{|x-y_0|^3} \right) & \text{if } i = j, \\ R \frac{(x_i - y_{0,i})(x_j - y_{0,j})}{|x-y_0|^3} & \text{if } i \neq j, \end{cases}$$

and the Hessian of $\frac{1}{2}(|x - z_0| - R)^2$ is

$$(H^r)_{i,j}(x) = \begin{cases} 1 - R \left(\frac{1}{|x-z_0|} - \frac{(x_i - z_{0,i})^2}{|x-z_0|^3} \right) & \text{if } i = j, \\ R \frac{(x_i - z_{0,i})(x_j - z_{0,j})}{|x-z_0|^3} & \text{if } i \neq j. \end{cases}$$

If we denote $r := |x_0 - p_0|$, then we have

$$|x_0 - y_0| = R - r \text{ and } |x_0 - z_0| = R + r.$$

Therefore, we have

$$H^l(x_0) = \left(1 - \frac{R}{R-r} \right) I_d + \frac{R}{R-r} \frac{(x_0 - y_0) \otimes (x_0 - y_0)}{|x_0 - y_0|^2}$$

and $H^r(x_0) = \left(1 - \frac{R}{R+r} \right) I_d + \frac{R}{R+r} \frac{(x_0 - z_0) \otimes (x_0 - z_0)}{|x_0 - z_0|^2}.$

Next, we use the property that $(v \otimes v)v = v \cdot (v \cdot v) = v$ with $\|v\| = 1$ which implies that 1 is an eigenvalue of $H^l(x_0)$ and $H^r(x_0)$ for the eigenvectors $\frac{x_0 - y_0}{|x_0 - y_0|}$ and $\frac{x_0 - z_0}{|x_0 - z_0|}$ respectively. By using the same property, we also obtain that $1 - \frac{R}{R-r}$ and $1 - \frac{R}{R+r}$ are eigenvalues of $H^l(x_0)$ and $H^r(x_0)$ respectively for any eigenvector in the spaces $\operatorname{Vect}(x_0 - y_0)^\perp$ and $\operatorname{Vect}(x_0 - z_0)^\perp$ respectively.

Since the equality in (3.19) is reached at $x = p_0$, we obtain

$$\begin{pmatrix} 1 & 0 \\ 0 & -\frac{r}{R-r} I_{d-1} \end{pmatrix} \leq \operatorname{Hess}_g(x_0) \leq \begin{pmatrix} 1 & 0 \\ 0 & \frac{r}{R+r} I_{d-1} \end{pmatrix}, \quad (3.20)$$

where the matrices are written in a basis from $\left\{ \frac{x_0 - y_0}{|x_0 - y_0|} \right\} \cup \operatorname{Vect}(x_0 - y_0)^\perp$ in the left-hand side and from $\left\{ \frac{x_0 - z_0}{|x_0 - z_0|} \right\} \cup \operatorname{Vect}(x_0 - z_0)^\perp$ in the right-hand side.

In particular, the inequality (3.20) implies that

$$\|D\Pi\| = \|2DP_\Omega - I_d\| = \|-2\operatorname{Hess}_g + I_d\| \leq \frac{r}{R+r} + 1,$$

so that the kinetic energy remains bounded.

The inequality (3.20) also gives

$$\det D\Pi = \det(-2\text{Hess}_g + I_d) = -1 \cdot \left(1 + \frac{2r}{R-r}\right)^{d-1},$$

so $\det D\Pi \geq \frac{1}{2}$ for sufficiently small r which implies that Π sends bounded densities by M to bounded densities by CM . \square

Theorem 3.2.2. *Suppose that Ω is either the torus or a C^2 compact domain in \mathbb{R}^d . Let \bar{Q} be such that (3.5) holds for all Q satisfying (3.4). Suppose that the densities of all the measures $e_t\#\bar{Q}$, denoted by $\bar{\rho}(t, \cdot)$, are bounded by a common constant. Let $f: \mathbb{R} \rightarrow \mathbb{R}$ be a C^1 convex function such that $f(u) \leq C(u^q + 1)$. Let p, q and d verify $q < 1 + \frac{p'}{d}$ where p' is the conjugate exponent of p . Let $\widehat{f'(\bar{\rho})}$ be defined as in (3.16). Then for all $Q \in \Gamma$, we have*

$$\begin{aligned} & \int_{\mathcal{C}} K(\gamma) d\bar{Q}(\gamma) + \int_0^T \int_{\Omega} \left(V(t, x) + \widehat{f'(\bar{\rho})}(t, x) \right) \bar{\rho}(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) d\bar{Q}(\gamma) \\ & \leq \int_{\mathcal{C}} K(\gamma) dQ(\gamma) + \int_0^T \int_{\Omega} \left(V(t, x) + \widehat{f'(\bar{\rho})}(t, x) \right) \rho(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ(\gamma). \end{aligned} \quad (3.21)$$

Proof. Let $Q \in \Gamma$ and $\rho(t, \cdot) := e_t\#Q$. Let $r > 0$ and $\eta(t) = t^\alpha$. For all $y \in B(0, 1) \subset \mathbb{R}^d$, we define the function

$$\begin{aligned} T_y : W^{1,p}([0, T]) & \rightarrow W^{1,p}([0, T]), \\ \gamma & \mapsto (t \mapsto \Pi(\gamma(t) + r\eta(t)y)). \end{aligned}$$

Then, we have $\|T_y(\gamma)'\|_{L^p([0, T])} \leq (1 + Cr)(\|\dot{\gamma}\|_{L^p([0, T])} + \|r\dot{\eta}y\|_{L^p([0, T])})$ (where the multiplying factor $1 + Cr$ comes from the composition with Π , see Lemma 3.2.1), so a necessary condition for $T_y(\gamma)'$ to be in $L^p([0, T])$ is that $\dot{\eta}$ is in $L^p([0, T])$ and this is true if $p(\alpha - 1) > -1$, i.e. $\alpha > 1 - 1/p$.

Next, we define similarly to Theorem 3.1.1 the sequence of measures $(Q_r)_r$:

$$Q_r = \int_{B(0, 1)} T_y\#Q dy,$$

and we denote by $\rho_r(t, \cdot) = e_t\#Q_r$.

We now perform similar computations as those of (3.12). We have

$$e_t\#Q_r = \Pi\# \left((e_t\#Q) * \frac{\mathbb{1}_{B_{r\eta(t)}}}{|B_{r\eta(t)}|} \right).$$

Computing the volume of the ball in \mathbb{R}^d gives $|B_{r\eta(t)}| = c(r\eta(t))^d$ for some $c > 0$. Then we have

$$\|e_t\#Q_r\|_{L^\infty(\Omega)} \leq C_2 \|e_t\#Q\|_{L^1(\Omega)} \left\| \frac{\mathbb{1}_{B_{r\eta(t)}}}{|B_{r\eta(t)}|} \right\|_{L^\infty(\Omega)} \leq C(r\eta(t))^{-d}.$$

Since $\eta(t) \rightarrow 0$ as $t \rightarrow 0$, the L^∞ norm will not be uniformly bounded, but we are actually interested in the L^q norm. We recall that for every probability density u we have

$$\|u\|_{L^q(\Omega)} = \left(\int u^q dx \right)^{1/q} \leq \left(\|u\|_{L^\infty}^{q-1} \int u dx \right)^{1/q} \leq \|u\|_{L^\infty}^{\frac{q-1}{q}}.$$

Thus,

$$\|e_t\#Q_r\|_{L^q(\Omega)} \leq C(r\eta(t))^{\frac{d(1-q)}{q}}.$$

We want to impose the integrability condition $\int_0^T \int_{\Omega} f(\rho_r) dx dt < +\infty$, i.e. (thanks to the assumption on f) $\int_0^T \|\rho_r(t)\|_{L^q}^q dt < +\infty$. For this we need to require another condition on α , i.e. $\alpha < \frac{1}{d(q-1)}$.

Such an α exists, because of the assumption on p and q . Indeed, we assumed $q < 1 + \frac{p'}{d}$, which is equivalent to $\frac{1}{p'} := 1 - \frac{1}{p} < \frac{1}{d(q-1)}$.

The condition (3.4) is now satisfied and we obtain by Proposition 3.0.1 the inequality

$$\begin{aligned} & \int_{\mathcal{C}} K(\gamma) d\bar{Q}(\gamma) + \int_0^T \int_{\Omega} (V(t, x) + f'(\bar{\rho}(t, x))) \bar{\rho}(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) d\bar{Q}(\gamma) \\ & \leq \int_{\mathcal{C}} K(\gamma) dQ_r(\gamma) + \int_0^T \int_{\Omega} (V(t, x) + [f'(\bar{\rho})]_r(t, x)) \rho_r(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ_r(\gamma). \end{aligned} \quad (3.22)$$

We want to take the limit as $r \rightarrow 0$ in the right-hand side. The last term can be treated just by noting that we have $Q_r \xrightarrow{*} Q$. The term in the middle can be treated by taking its limsup exactly as in the proof of Theorem 3.1.2. We are only left to consider the first term, which, in the proof of Theorem 3.1.2, was dealt with by noting that we had $S(T_y(\gamma)) \leq S(\gamma)$.

Here this inequality is no longer true, but almost. Indeed, we noted that we have the inequality $\|T_y(\gamma)'\|_{L^p([0,T])} \leq (1 + Cr)\|\dot{\gamma}\|_{L^p([0,T])} + \|r\dot{\eta}y\|_{L^p([0,T])}$. This means $K(T_y(\gamma)) \leq (1 + Cr)^{1/p}((K(\gamma))^{1/p} + Cr)^p$. We then write

$$\int_{\mathcal{C}} K(\gamma) dQ_r(\gamma) = \int_{B(0,1)} \int_{\mathcal{C}} K(T_y(\gamma)) dQ(\gamma) \leq (1 + Cr)^{1/p} \int_{B(0,1)} \int_{\mathcal{C}} (K(\gamma))^{1/p} + Cr)^p dQ(\gamma),$$

and we can take the limit by dominated convergence once we note that we have $(K(\gamma))^{1/p} + Cr)^p \leq C(K(\gamma) + 1)$ and K is supposed to be Q -integrable. We then obtain

$$\begin{aligned} & \int_{\mathcal{C}} K(\gamma) d\bar{Q}(\gamma) + \int_0^T \int_{\Omega} (V(t, x) + f'(\bar{\rho}(t, x))) \bar{\rho}(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) d\bar{Q}(\gamma) \\ & \leq \int_{\mathcal{C}} K(\gamma) dQ(\gamma) + \int_0^T \int_{\Omega} (V(t, x) + \widehat{f'(\bar{\rho})}(t, x)) \rho(t, x) dx dt + \int_{\mathcal{C}} \psi_T(\gamma(T)) dQ_r(\gamma) \end{aligned}$$

and we conclude by replacing once more $f'(\bar{\rho})$ with $\widehat{f'(\bar{\rho})}$ in the left-hand side, since these functions agree a.e. \square

Remark 3.2.3. We note that in the above proof the condition on p, q and d is required to find a function $\eta \in W^{1,p}([0, T])$ with $\eta(0) = 0$ and such that $\eta^{(1-q)d}$ is integrable. The difficulty is due to the fact that this integrability condition requires η not to tend to 0 too fast and the Sobolev regularity of η imposes to tend to 0 fast enough. . . If we could remove the condition $\eta(0) = 0$ we could just use $\eta(t) = 1$ which satisfies both conditions at the same time. The reason for imposing $\eta(0) = 0$ lies in the need to preserve the initial condition of the curves, but actually we only need $e_0 \# Q_r = m_0$. Hence, an example where the condition on p, q and d could be dropped is the case where Ω is the torus and m_0 is the uniform measure on it. In this case with $\eta(t) = 1$ the measure $e_0 \# Q_r$ would be the convolution of m_0 with the uniform measure on the ball, and thus it would also be equal to m_0 . It would be interesting to investigate whether smoothness properties on m_0 , which do not allow to say that the convolution equals m_0 but that it is close enough to it, would be enough to “correct” the error in the initial condition and obtain the same result under less stringent assumptions on p, q and d .

Again, we can then conclude about the equilibrium properties of the measure \bar{Q} that we obtained as a fixed point in Section 1.4.

Corollary 3.2.4. Suppose that the measures $e_t \# \bar{Q} = \bar{\rho}(t, \cdot)$ are all bounded by a same constant. Then, the measure \bar{Q} is concentrated on the curves $\bar{\gamma}$ which are minimizers, for fixed starting point, of the action function J_W with $W = V + \widehat{f'(\bar{\rho})}$, i.e. such that such

$$\begin{aligned} \forall \omega \in \mathcal{C} \text{ with } \omega(0) = \bar{\gamma}(0) \quad & K(\bar{\gamma}) + \int_0^T \left[V(t, \bar{\gamma}(t)) + \widehat{f'(\bar{\rho})}(t, \bar{\gamma}(t)) \right] dt + \psi_T(\bar{\gamma}(T)) \\ & \leq K(\omega) + \int_0^T \left[V(t, \omega(t)) + \widehat{f'(\bar{\rho})}(t, \omega(t)) \right] dt + \psi_T(\omega(T)). \end{aligned}$$

Chapter 4

Regularity for Variational Mean Field Games with jumps

In this chapter, essentially taken from [DS23], we are interested in the regularity of the density $\rho(t) = e_t \# Q$ which is the Eulerian representation of a Nash equilibrium $Q \in \mathcal{P}_{m_0}(\mathcal{C})$ for which we proved its existence in Section 1.4. We recall that Q minimizes

$$\min_{Q \in \mathcal{P}_{m_0}(\mathcal{C})} \mathcal{U}(Q) = \int_{\mathcal{C}} S(\gamma) dQ(\gamma) + \int_0^T I(e_t \# Q) dt + \int_{\mathcal{C}} \int_0^T V(t, \gamma(t)) dt dQ(\gamma) + \int_{\mathcal{C}} \Psi(\gamma(T)) dQ(\gamma) \quad (4.1)$$

and in Chapter 3, we showed that Q is concentrated on curves which minimize

$$J_W(\gamma) = S(\gamma) + \int_0^T W(t, \gamma(t)) dt + \Psi(\gamma(T)) \quad \text{for } W(t, x) = \frac{dI}{dm}[e_t \# Q](x) + V(t, x)$$

which includes the number of jumps S defined in (1.28) for piecewise constant trajectories γ . The function $V: [0, T] \times \Omega \rightarrow \mathbb{R}$ is supposed to be continuous and I is of the form

$$I(\rho) = \begin{cases} \int_{\Omega} f(\rho(x)) dx & \text{if } \rho \ll \mathcal{L}^d, \\ +\infty & \text{if not.} \end{cases}$$

In Chapter 2, we showed that the problem (4.1) is equivalent to

$$\min_{\substack{\rho \in \mathcal{E}, \rho \geq 0 \\ \forall t \in [0, T], \int_{\Omega} \rho(t, x) dx = 1}} \mathcal{V}(\rho) := \int_0^T \int_{\Omega} \left(\frac{1}{2} |\dot{\rho}(t, x)| + V(t, x) \rho(t, x) + f(\rho(t, x)) \right) dx dt + \psi_0(\rho(0)) + \psi_T(\rho(T)),$$

where $\mathcal{E} := \text{BV}([0, T]; L^1(\Omega)) \cap L^2([0, T] \times \Omega)$.

Because of the BV behavior of the curves, the Dirichlet condition $\rho(0) = \rho_0$, as well as the final penalization on ρ_T , have to be suitably interpreted. Indeed, it is always possible to jump exactly at time $t = 0$ or $t = T$, so that the Dirichlet condition at $t = 0$ can be replaced by a penalization $\frac{1}{2} \|\rho(0^+) - \rho_0\|_{L^1}$ and for the final penalization, we can replace $\int_{\Omega} \Psi d\rho_T$ with $\inf_{\mu \in \mathcal{P}(\Omega)} \|\mu - \rho_T\| + \int \tilde{\Psi} d\mu$. This last quantity can be computed and equals $\int \tilde{\Psi} d\rho_T$, where $\tilde{\Psi} = \min\{\Psi, \inf \Psi + 1\}$. This is perfectly coherent with the individual optimization problem: if agents are allowed to jump at a cost 1, the final cost Ψ is automatically replaced by $\tilde{\Psi}$ as there is no point in paying $\Psi(x)$ whenever $\Psi(x) > \Psi(x') + 1$ for some point x' . Up to subtracting a constant to the final penalization, we can thus suppose that we have $|\Psi| \leq 1/2$.

More generally, we will study in this chapter the variational problem

$$\min_{\rho} \int_0^T \int_{\Omega} (\lambda |\dot{\rho}| + f(\rho_t) + V_t \rho_t) dt + \int_{\Omega} \Psi d\rho_T$$

with a final cost Ψ satisfying $|\Psi| \leq \lambda$ and an initial condition $\rho(0) = \rho_0$ which can also be replaced by a penalization $\lambda \|\rho(0^+) - \rho_0\|_{L^1}$. A variant will be the infinite-horizon case with a discount factor $r > 0$, i.e. the variational problem

$$\min_{\rho} \int_0^{\infty} \int_{\Omega} e^{-rt} (\lambda |\dot{\rho}| + f(\rho_t) + V_t \rho_t) dt$$

under the same initial condition.

For simplicity, we will only consider the case where the function f is uniformly convex (think at $f(\rho) := \frac{1}{2}\rho^2$). We still establish regularity results in both time and space for the optimal solution (which is unique because of strict convexity). This result, besides its mathematical interest, has also at least two applications in the MFG theory which motivates the problem. First, it proves that, despite individual trajectories being discontinuous, the global behavior of the density of agents $\rho(t, x)$ is smooth, coherently with the experience about the evolution of residential areas. Second, it provides the necessary mathematical properties on the individual running cost $g(\rho) + V$ so as to rigorously prove that minimizers of the variational problem are indeed equilibria of the game (see Chapter 3). This requires, as we will briefly explain at the end of Section 4.4, the continuity (in space) of the running cost, or at least its boundedness; we refer to [San20]) for more details.

Since proving regularity in a problem set on BV curves could be challenging, in order to develop the relevant techniques (which will be based on a suitable use of the maximum principle) we will first start from a simpler, yet not-so-standard, case, where curves, instead of being valued in the functional space L^1 , will be simply valued in the euclidean space \mathbb{R}^d . This will be object of Section 4.1, where we will prove Lipschitz behavior in the open interval $(0, T)$. Some explicit examples will also be analyzed, in particular for $d = 1$, in order to have some cases which could be used as a test for the numerical methods of Chapter 5.

The analysis of the infinite-dimensional case, valued in L^1 , will be the object of Section 4.2, and will be performed by means of many approximations. Some of them are common with the Euclidean case, but there is an extra approximation which comes from discretization: the infinite-dimensional problem is indeed approximated by a sequence of finite-dimensional ones, where the domain Ω is divided into small cells and only piecewise constant densities are considered. Note that in this Eulerian discretization (different from a Lagrangian one where one should follow the jumping trajectories of the particles) the problem becomes very similar to the one studied in Section 4.1 with the only exception that the norm used on the finite-dimensional space is not the Euclidean one (i.e. ℓ^2) but the ℓ^1 norm. This makes things slightly more involved, but, surprisingly, the final Lipschitz regularity result will be expressed anyway in the L^2 norm.

Section 4.3 contains the modifications of the strategy proof which are needed to handle the infinite-horizon case, both in the finite and infinite dimensional case. Then, Section 4.4 addresses the problem of space regularity. We will prove that the solution $\rho(t, x)$ shares the same modulus of continuity in x (uniformly in t) of the Dirichlet data and of the time-dependent potential $V(t, x)$, and in this proof, differently from what is done in Sections 4.1, 4.2 and 4.3, the Dirichlet data will be attacked by approximation but without replacing them with penalizations (the L^1 cost will be approximated by other superlinear costs on the velocity, as it was already done in the other sections as well). Indeed, for the previous results, it was crucial to use the transversality conditions coming from a suitable approximation of these penalizations, while here, on the contrary, the transversality conditions are harder to consider. This is why the first regularity result is presented in the case where the problem is endowed with Dirichlet conditions at both $t = 0$ and $t = T$, which is not the natural framework we will be interested in. In order to consider some instances of the problem which are more interesting for applications, we will consider the infinite time horizon with Dirichlet data at $t = 0$ (in this case we do not have penalization at the end), where the whole analysis can be performed, as well as the finite-horizon case when the final penalization ψ is piecewise constant: in this last case we can prove continuity of ρ on each piece where ψ is constant, which is enough, for instance, to obtain $\rho \in L^\infty$.

4.1 Lipschitz regularity in the Euclidean setting

The result of this section was originally proved in [DS23] and presented in the book [San23].

As a starting point for our analysis, we consider here the following easier problem

$$\min \left\{ TV(\gamma; [0, 1]) + \int_0^1 F(t, \gamma(t)) dt + \psi_0(\gamma(0)) + \psi_1(\gamma(1)) : \gamma \in BV([0, 1]; \mathbb{R}^d) \right\}. \quad (4.2)$$

Here $TV(\gamma; [0, 1])$ denotes the total variation of γ on $[0, 1]$, i.e.

$$TV(\gamma; [0, 1]) := \sup \left\{ \sum_{k=0}^{N-1} |\gamma(t_k) - \gamma(t_{k+1})| : 0 = t_0 < t_1 < \dots < t_N = 1 \right\},$$

a value which also coincides with the total mass of the vector measure γ' .

The functions $\psi_0, \psi_1 : \mathbb{R}^d \rightarrow [0, +\infty]$ are just supposed to be l.s.c. and bounded from below, and

among possible choices we mention those which impose Dirichlet boundary conditions, i.e.

$$t = 0, 1 \quad \psi_t(x) = \begin{cases} 0 & \text{if } x = x_t, \\ +\infty & \text{if not.} \end{cases}$$

We stress the fact that functions in BV spaces are not continuous and can have jumps ; even if we consider that BV functions of one variable are defined pointwisely, it is possible to change very easily their value at a point. In particular, Dirichlet boundary conditions have a very particular meaning: a curve which takes the value x_0 at $t = 0$ but immediately jumps at another point at $t = 0^+$ is considered to satisfy the condition $\gamma(0) = x_0$. In particular, it is possible to freely choose a curve γ on $(0, 1)$ and then add a jump to x_0 or x_1 at the boundary in order to satisfy the corresponding Dirichlet boundary condition, of course adding a price $|\gamma(0^+) - x_0|$ or $|\gamma(1^-) - x_1|$ to the total variation. In this way, we could decide to identify the values of γ at $t = 0$ or $t = 1$ with their right or left limits at these points, respectively, and replace Dirichlet boundary conditions with a boundary penalization. This could also be done for more general penalizations ψ_0, ψ_1 , for which it is useful to define the relaxed functions

$$\tilde{\psi}_t(x) := \inf_y |y - x| + \psi_t(y).$$

It is important to observe that the functions $\tilde{\psi}_i$ are automatically 1-Lipschitz continuous, as an inf of Lip_1 functions of the variable x , indexed with the parameter y .

In this way the problem (4.2) becomes

$$\min \left\{ TV(\gamma; [0, 1]) + \int_0^1 F(t, \gamma(t)) dt + \tilde{\psi}_0(\gamma(0^+)) + \tilde{\psi}_1(\gamma(1^-)) : \gamma \in BV([0, 1]; \mathbb{R}^d) \right\}$$

or, equivalently, we can replace $\tilde{\psi}_0(\gamma(0^+)) + \tilde{\psi}_1(\gamma(1^-))$ with $\tilde{\psi}_0(\gamma(0)) + \tilde{\psi}_1(\gamma(1))$ and impose continuity of γ at $t = 0, 1$.

Lemma 4.1.1. *Let $L : \mathbb{R}^d \rightarrow \mathbb{R}$ be a smooth and uniformly convex function which is supposed to be radial: $L(v) := \ell(|v|)$ for a convex and increasing function $\ell : \mathbb{R}_+ \rightarrow \mathbb{R}$. Let $F : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a C^2 time-dependent potential satisfying $D_{xx}^2 F(t, x) \geq c_0 \mathbf{I}$ for a certain constant $c_0 > 0$ and $|\partial_t \nabla_x F(t, x)| \leq C_0$, and $\psi_0, \psi_1 : \mathbb{R}^d \rightarrow \mathbb{R}$ two Lipschitz continuous functions. Consider a solution γ of*

$$\min \left\{ \int_0^1 (L(\gamma'(t)) + F(t, \gamma(t))) dt + \psi_0(\gamma(0)) + \psi_1(\gamma(1)) : \gamma \in H^1([0, 1]) \right\}.$$

Then γ is Lipschitz continuous and satisfies $|\gamma'| \leq C$ where C is defined by

$$C := \max \left\{ \frac{C_0}{c_0}, (\ell')^{-1}(\text{Lip} \psi_0), (\ell')^{-1}(\text{Lip} \psi_1) \right\}.$$

Proof. Let us start from the Euler-Lagrange system of the above optimization problem. We have

$$\begin{cases} (\nabla L(\gamma'))'(t) = \nabla_x F(t, \gamma(t)) \\ \nabla L(\gamma'(0)) = \nabla \psi_0(\gamma(0)) \\ \nabla L(\gamma'(1)) = -\nabla \psi_1(\gamma(1)). \end{cases}$$

First we observe that $\gamma \in C^0$ and $F \in C^1$ imply that the right-hand side in the first equation is a continuous function, so that we have $\nabla L(\gamma') \in C^1$. Inverting the injective function ∇L we obtain $\gamma \in C^2$ and, since $F \in C^2$, we obtain $\gamma \in C^3$.

Then, the transversality conditions show $|\nabla L(\gamma'(t))| \leq \text{Lip} \psi_t$ for $t = 0, 1$. Using $|\nabla L(v)| = \ell'(|v|)$ we see $|\gamma'(t)| \leq C$ for $t = 0, 1$.

Let us now consider the maximal value of $|\gamma'(t)|$. This maximum exists on $[0, 1]$ since $\gamma \in C^1$ and if it is attained on the boundary $t = 0, 1$ the desired Lipschitz bound $|\gamma'| \leq C$ is satisfied. We can now suppose that it is attained in $(0, 1)$. Since ℓ' is increasing and non-negative, the maximal points of $|\gamma'|$ and of $|\nabla L(\gamma')|^2$ are the same. We can then write the optimality condition differentiating once and twice in t : we do have

$$\nabla L(\gamma'(t)) \cdot (\nabla L(\gamma'))'(t) = 0; \quad \nabla L(\gamma'(t)) \cdot (\nabla L(\gamma'))''(t) + |(\nabla L(\gamma'))'(t)|^2 \leq 0.$$

In the last condition we can ignore the positive term $|(\nabla L(\gamma'))'(t)|^2$ and observe that, since $\nabla L(\gamma'(t))$ and $\gamma'(t)$ are vectors with the same orientation (we do have $\nabla L(\gamma'(t)) = \frac{\ell'(|\gamma'(t)|)}{|\gamma'(t)|} \gamma'(t)$), we have $\gamma'(t) \cdot (\nabla L(\gamma'))''(t) \leq 0$.

We now differentiate in time the Euler-Lagrange equation and take the scalar product times γ' , and obtain

$$0 \geq \gamma'(t) \cdot (\nabla L(\gamma'))''(t) = (\nabla_x F(t, \gamma(t)))' \cdot \gamma'(t) = \partial_t \nabla_x F(t, \gamma(t)) \cdot \gamma'(t) + \gamma'(t) \cdot D_{xx}^2 F(t, \gamma(t)) \gamma'(t).$$

We deduce

$$c_0 |\gamma'(t)|^2 \leq |\gamma'(t)| |\partial_t \nabla_x F(t, \gamma(t))|,$$

which implies $|\gamma'(t)| \leq \frac{C_0}{c_0} \leq C$ and concludes the proof. \square

We now use the above result on an approximation of the original problem in BV.

Proposition 4.1.2. *Consider*

$$\min \left\{ TV(\gamma; [0, 1]) + \int_0^1 F(t, \gamma(t)) dt + \psi_0(\gamma(0^+)) + \psi_1(\gamma(1^-)) : \gamma \in BV([0, 1]; \mathbb{R}^d) \right\} \quad (4.3)$$

where $F : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}$ is a C^2 time-dependent potential satisfying $D_{xx}^2 F(t, x) \geq c_0 I$ for a certain constant $c_0 > 0$ and $|\partial_t \nabla_x F(t, x)| \leq C_0$, and $\psi_0, \psi_1 \in \text{Lip}_1(\mathbb{R}^d)$ are two penalization functions.

Then a minimizer γ for the above problem exists, is unique, and is actually Lipschitz continuous with $|\gamma'| \leq \frac{C_0}{c_0}$.

Note that we directly state the problem using Lip_1 penalizations instead of first fixing ψ_0 and ψ_1 and then passing to $\tilde{\psi}_0$ and $\tilde{\psi}_1$, but we have already explained why we can restrict to this case. Yet, an important remark is needed:

Remark 4.1.3. *The solutions with penalizations ψ_t and $\tilde{\psi}_t$ ($t = 0, 1$) coincide in $(0, 1)$, but the solution with the original (non- Lip_1) penalizations could jump at $t = 0$ or $t = 1$, and this jump is intrinsically considered in the definition of ψ_t .*

Proof. Given $\varepsilon > 0$, we define $\ell_\varepsilon : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ via $\ell_\varepsilon(s) := \sqrt{\varepsilon^2 + s^2} + \varepsilon h(s)$, where $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a smooth, convex, and increasing function, with $\liminf_{s \rightarrow \infty} h''(s) > 0$. We then define $L_\varepsilon : \mathbb{R}^d \rightarrow \mathbb{R}$ via $L_\varepsilon(v) = \ell_\varepsilon(|v|)$, so that L_ε is smooth, uniformly convex, and radial¹.

We also choose some numbers $\alpha_\varepsilon < 1$ in such a way that $\lim_{\varepsilon \rightarrow 0} \alpha_\varepsilon = 1$ and $\lim_{\varepsilon \rightarrow 0} \frac{\varepsilon^2}{1 - \alpha_\varepsilon^2} = 0$ (for instance $\alpha_\varepsilon = \sqrt{1 - \varepsilon}$).

We consider γ_ε the solution of the variational problem

$$\min \left\{ \int_0^1 (L_\varepsilon(\gamma'(t)) + F(t, \gamma(t))) dt + \alpha_\varepsilon (\psi_0(\gamma(0)) + \psi_1(\gamma(1))) : \gamma \in H^1([0, 1]) \right\}.$$

Since L_ε is convex, a solution exists by the direct method of the calculus of variations, and it is unique because of the strict convexity of the function F .

We want to apply Lemma 4.1.1 to this approximated optimization problem. We first compute $\ell'_\varepsilon(s) = \frac{s}{\sqrt{\varepsilon^2 + s^2}} + \varepsilon h'(s) \geq \frac{s}{\sqrt{\varepsilon^2 + s^2}}$ and observe that we have $(\ell'_\varepsilon)^{-1}(r) \leq \frac{r\varepsilon}{\sqrt{1 - r^2}}$. Since $\text{Lip}(\alpha_\varepsilon \psi_i) = \alpha_\varepsilon \text{Lip}(\psi_i) \leq \alpha_\varepsilon$ we obtain from Lemma 4.1.1

$$|\gamma'_\varepsilon| \leq \max \left\{ \frac{C_0}{c_0}, \frac{\alpha_\varepsilon \varepsilon}{\sqrt{1 - \alpha_\varepsilon^2}} \right\},$$

and we observe that our choice of α_ε implies that the second term in the max above tends to 0 as $\varepsilon \rightarrow 0$. This means that the Lipschitz constant of γ_ε is at most $\frac{C_0}{c_0}$ if ε is small enough.

By comparing γ_ε with the constant curve $\gamma = 0$ we obtain

$$\int_0^1 F(t, \gamma_\varepsilon(t)) dt + \alpha_\varepsilon (\psi_0(\gamma_\varepsilon(0)) + \psi_1(\gamma_\varepsilon(1))) \leq \int_0^1 F(t, 0) dt + \alpha_\varepsilon (\psi_0(0) + \psi_1(0)) \leq C.$$

This estimate includes an L^2 estimate on γ_ε and, because of the uniform Lipschitz condition on γ_ε , it also implies that the curves γ_ε are equibounded. We can then apply Arzelà-Ascoli's theorem to obtain a limit curve $\gamma_\varepsilon \rightarrow \gamma_0$. This curve γ_0 is of course $\frac{C_0}{c_0}$ -Lipschitz continuous, and we can prove that it solves Problem (4.3).

¹Note that the easiest choice for h is $h(s) = \frac{1}{2}s^2$, but other choices are possible and reasonable, and the only role of h is to guarantee a lower bound on the Hessian of L_ε (and in particular, to provide a quadratic behavior to ℓ_ε so that the problem is well-posed in H^1). Later, one will see the interest for other choices of h .

Indeed, the optimality of γ_ε , together with the inequality $L_\varepsilon(v) \geq |v|$ (which allows to bound the total variation from above with the integral of L_ε), shows that we have

$$\begin{aligned} TV(\gamma_\varepsilon; [0, 1]) + \int_0^1 F(t, \gamma_\varepsilon(t)) dt + \alpha_\varepsilon(\psi_0(\gamma_\varepsilon(0)) + \psi_1(\gamma_\varepsilon(1))) \\ \leq \int_0^1 L_\varepsilon(\gamma'_\varepsilon(t)) dt + \int_0^1 F(t, \gamma(t)) dt + \alpha_\varepsilon(\psi_0(\gamma(0)) + \psi_1(\gamma(1))) \end{aligned}$$

for every $\gamma \in H^1$. If we let ε tend to 0 and use the lower semi-continuity of TV for the uniform convergence, we obtain

$$\begin{aligned} TV(\gamma_0; [0, 1]) + \int_0^1 F(t, \gamma_0(t)) dt + \psi_0(\gamma_0(0)) + \psi_1(\gamma_0(1)) \\ \leq \int_0^1 |\gamma'(t)| dt + \int_0^1 F(t, \gamma(t)) dt + \psi_0(\gamma(0)) + \psi_1(\gamma(1)), \end{aligned}$$

where we used the dominated convergence $L_\varepsilon(\gamma') \rightarrow |\gamma'|$ as $\varepsilon \rightarrow 0$, and $\alpha_\varepsilon \rightarrow 1$.

This shows the optimality of γ_0 compared to any H^1 curve. It is now enough to approximate any BV curve with H^1 curves. We take $\gamma \in BV([0, 1])$, we define it as equal to $\gamma(0^+)$ on $[-1, 0]$ and to $\gamma(1^-)$ on $[1, 2]$ and we convolve it with a smooth compactly supported kernel η_δ tending to the identity so as to smooth it, thus obtaining a sequence of curves $\gamma * \eta_\delta$ such that $TV(\gamma * \eta_\delta; [-1, 2]) = \int_{-1}^2 |(\gamma * \eta_\delta)'(t)| dt \leq TV(\gamma, (0, 1))$; moreover, $\gamma * \eta_\delta$ is uniformly bounded and converges to γ at all continuity points of γ , which means that the convergence holds a.e. and at the boundary point. This proves

$$\begin{aligned} \limsup_{\delta \rightarrow 0} \int_0^1 |(\gamma * \eta_\delta)'(t)| dt + \int_0^1 F(t, \gamma * \eta_\delta(t)) dt + \psi_0(\gamma * \eta_\delta(0)) + \psi_1(\gamma * \eta_\delta(1)) \\ \leq TV(\gamma, (0, 1)) + \int_0^1 F(t, \gamma(t)) dt + \psi_0(\gamma(0)) + \psi_1(\gamma(1)) \end{aligned}$$

and concludes the proof of the optimality of γ_0 .

What we proved implies the existence of an optimal curve for (4.3), and its uniqueness comes, again, from the strict convexity of the term with F . \square

In the case where the space of curves is $BV([0, 1]; \mathbb{R}^d)$ with $d = 1$, we can obtain a very interesting behavior.

Proposition 4.1.4. *When $d = 1$, i.e. the target space of the curves in Problem (4.3) is one-dimensional, the minimizer γ satisfies $|\gamma'(t)| |\nabla_x F(t, \gamma(t))| = 0$ a.e., i.e. at each instant of time either γ does not move or it is already located at the optimal point for $F(t, \cdot)$.*

Proof. We consider the same approximation as in Proposition 4.1.2, using the function $h(s) = (s - M)_+^2$ for a very large M . The uniform Lipschitz bound proven in Lemma 4.1.1 and Proposition 4.1.2 makes it irrelevant the choice of $\ell_\varepsilon(s)$ for large values of s , so that we can write the Euler-Lagrange equation for the minimizer γ_ε in the form

$$\frac{\varepsilon^2}{(\varepsilon^2 + |\gamma'_\varepsilon|^2)^{3/2}} \gamma''_\varepsilon = (L'_\varepsilon(\gamma'_\varepsilon))' = \nabla_x F(t, \gamma_\varepsilon),$$

where we explicitly computed² the second derivative of L_ε ignoring the term in h .

We write this as $\varepsilon^2 \gamma''_\varepsilon = (\varepsilon^2 + |\gamma'_\varepsilon|^2)^{3/2} \nabla_x F(t, \gamma_\varepsilon)$. First, note that this implies a uniform bound $\varepsilon^2 |\gamma''_\varepsilon| \leq C$. Then, we differentiate it in time, thus obtaining

$$\varepsilon^2 \gamma'''_\varepsilon = 3(\varepsilon^2 + |\gamma'_\varepsilon|^2)^{1/2} \gamma'_\varepsilon \cdot \gamma''_\varepsilon \nabla_x F(t, \gamma_\varepsilon) + (\varepsilon^2 + |\gamma'_\varepsilon|^2)^{3/2} (\nabla_x F(t, \gamma_\varepsilon))'.$$

Re-using the Euler-Lagrange equation we have

$$\varepsilon^2 \gamma'''_\varepsilon = 3\varepsilon^2 (\varepsilon^2 + |\gamma'_\varepsilon|^2)^{-1} (\gamma'_\varepsilon \cdot \gamma''_\varepsilon) \gamma''_\varepsilon + (\varepsilon^2 + |\gamma'_\varepsilon|^2)^{3/2} (\nabla_x F(t, \gamma_\varepsilon))'.$$

²Note that this computation is based on a 1D cancellation effect, since in higher dimension we have

$$D^2 L_\varepsilon(v) = \frac{(\varepsilon^2 + |v|^2)I - v \otimes v}{(\varepsilon^2 + |v|^2)^{3/2}}$$

and the matrices $|v|^2 I$ and $v \otimes v$ do not cancel out.

We observe that the last term $(\varepsilon^2 + |\gamma'_\varepsilon|^2)^{3/2} (\nabla_x F(t, \gamma_\varepsilon))'$ is bounded thanks to the Lipschitz bound on γ_ε and the regularity of F . We multiply by γ'_ε and obtain

$$|\varepsilon^2 \gamma_\varepsilon''' \cdot \gamma'_\varepsilon - 3\varepsilon^2 (\varepsilon^2 + |\gamma'_\varepsilon|^2)^{-1} (\gamma'_\varepsilon \cdot \gamma_\varepsilon'')^2| \leq C. \quad (4.4)$$

We then compute

$$\varepsilon^2 \int_0^1 |\gamma_\varepsilon''|^2 dt = -\varepsilon^2 \int_0^1 \gamma_\varepsilon''' \cdot \gamma'_\varepsilon dt + [\varepsilon^2 \gamma'_\varepsilon \cdot \gamma_\varepsilon'']_0^1 \leq C.$$

The last inequality is justified by (4.4), by the sign of the term $\varepsilon^2 (\varepsilon^2 + |\gamma'_\varepsilon|^2)^{-1} (\gamma'_\varepsilon \cdot \gamma_\varepsilon'')$, and by the bound on the boundary term, which is the product of two bounded quantities: γ'_ε and $\varepsilon^2 \gamma_\varepsilon''$.

Coming back to the equality $\varepsilon^2 \gamma_\varepsilon'' = (\varepsilon^2 + |\gamma'_\varepsilon|^2)^{3/2} \nabla_x F(t, \gamma_\varepsilon)$ we take the L^2 norms of both sides, thus obtaining

$$\int_0^1 |\gamma'_\varepsilon|^6 |\nabla_x F(t, \gamma_\varepsilon)|^2 dt \leq \int_0^1 (\varepsilon^2 + |\gamma'_\varepsilon|^2)^3 |\nabla_x F(t, \gamma_\varepsilon)|^2 dt = \int_0^1 \varepsilon^4 |\gamma_\varepsilon''|^2 dt \leq C\varepsilon^2.$$

We deduce $\int_0^1 |\gamma'_\varepsilon|^6 |\nabla_x F(t, \gamma_\varepsilon)|^2 dt \rightarrow 0$ and, at the limit, by lower semi-continuity, we have the claim. \square

We will now analyze in details a simple example, both for future use and for better understanding the properties of the minimizers.

We consider the periodic problem

$$\min \left\{ J(\gamma) := \lambda TV(\gamma; \mathbb{S}^1) + \int_{\mathbb{S}^1} \frac{1}{2} |\gamma(t) - \omega(t)|^2 dt : \gamma \in BV(\mathbb{S}^1; \mathbb{R}) \right\}, \quad (4.5)$$

where $\lambda > 0$ and $\omega : \mathbb{S}^1 \rightarrow \mathbb{R}$ is a fixed curve. We suppose that ω is a Lipschitz continuous function such that there exists a finite decomposition of \mathbb{S}^1 into essentially disjoint intervals $I_k = [a_k, b_k]$ ($k = 1, \dots, 4N$ for $N \geq 1$) such that $a_{k+1} = b_k$ and $b_{4N} = a_1$ and satisfying for each $n = 0, \dots, N-1$, the following conditions

- ω is non-decreasing on I_{4n+1} ;
- $\omega(a_{4n+2}) = \omega(b_{4n+2}) := c_{4n+2}$ and $\omega \geq c_{4n+2}$ on I_{4n+2} ;
- ω is non-increasing on I_{4n+3} ;
- $\omega(a_{4n+4}) = \omega(b_{4n+4}) := c_{4n+4}$ and $\omega \leq c_{4n+4}$ on I_{4n+4} ;
- $\int_{I_{4n+2}} \omega - c_{4n+2} = \int_{I_{4n+4}} c_{4n+4} - \omega = 2\lambda$.

We then define a curve γ and a function z via

$$\begin{array}{llll} \text{on } I_{4n+1} & \gamma = \omega & \text{and} & z = \lambda, \\ \text{on } I_{4n+2} & \gamma = c_{4n+2} & \text{and} & z(t) = \lambda - \int_{a_{4n+2}}^t (\omega - c_{4n+2}), \\ \text{on } I_{4n+3} & \gamma = \omega & \text{and} & z = -\lambda, \\ \text{on } I_{4n+4} & \gamma = c_{4n+4} & \text{and} & z(t) = -\lambda - \int_{a_{4n+2}}^t (\omega - c_{4n+4}). \end{array}$$

We can check that γ and z are Lipschitz continuous functions, that we have $z' = \omega - \gamma$ as well as $|z| \leq \lambda$, $z = \pm\lambda$ when $\gamma' \neq 0$ and z and γ' have the same sign. Hence, $z(t) \in \partial(\lambda|\cdot|)(\gamma'(t))$. This means that we have, for any curve $\tilde{\gamma}$, the inequality

$$\lambda TV(\tilde{\gamma}; \mathbb{S}^1) \geq \lambda TV(\gamma; \mathbb{S}^1) + \int_{\mathbb{S}^1} (\tilde{\gamma}' - \gamma') \cdot z. \quad (4.6)$$

We claim that γ is a solution of Problem (4.5). Indeed, for any other competitor $\tilde{\gamma}$, we have

$$J(\tilde{\gamma}) = \lambda TV(\tilde{\gamma}; \mathbb{S}^1) + \int_{\mathbb{S}^1} \frac{1}{2} |\tilde{\gamma} - \omega|^2 \geq J(\gamma) + \int_{\mathbb{S}^1} (\tilde{\gamma}' - \gamma') \cdot z + (\tilde{\gamma} - \gamma) \cdot (\gamma - \omega) = J(\gamma),$$

where the last equality is obtained by integrating by parts and using $z' = \omega - \gamma$. The previous inequality comes from the use of (4.6) and from expanding the square.

This explicit example confirms the behavior predicted in Proposition 4.1.4: the optimizer in the scalar case either coincides with the minimal point of $F(t, \cdot)$ (here such a point is equal to $\omega(t)$) or it does not move. By using the numerical method described in Chapter 5, the solution to the problem (4.5) is displayed in Figure 4.1 when $\omega : \mathbb{S}^1 \rightarrow \mathbb{R}$ is a curve such that $\omega(t) = \sin(2\pi t) + 3 \sin(3 \cdot 2\pi t)$. The periodic function ω verifies the hypotheses listed above and we can see that the solution γ either follows ω or it is constant on each interval I_k . This numerical simulation confirms the solution and at the same time it validates the numerical method which will be presented more precisely in Chapter 5.

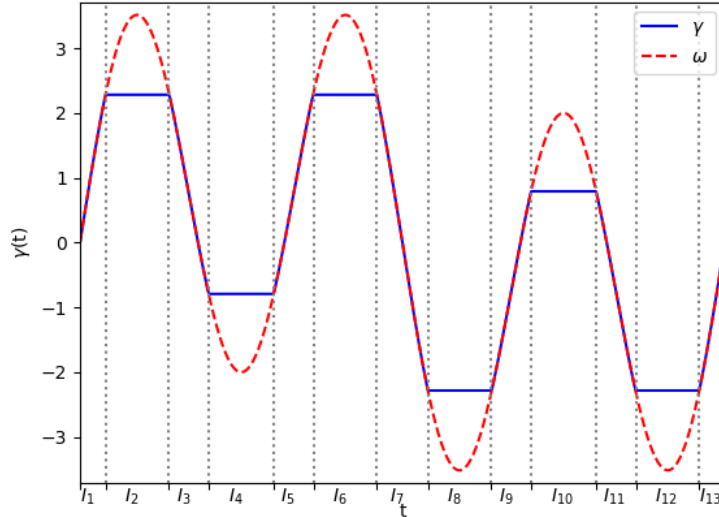


Figure 4.1: The simulation of the solution γ to the 1D periodic case (4.5) with $\omega(t) = \sin(2\pi t) + 3 \sin(3 \cdot 2\pi t)$. The parameters are displayed in Table 5.1 except from $S = 1$ and $\lambda = 0.04$. The blue solid line corresponds to the solution γ , while the red dashed line is the profile of ω . Each interval I_k is delimited by the vertical dotted gray lines.

4.2 Lipschitz regularity in the L^1 setting

In this section, we prove the regularity in time of the density ρ which solves the following problem:

$$\min_{\substack{\rho \in \mathcal{E}, \rho \geq 0 \\ \forall t \in [0, T], \int_{\Omega} \rho(t, x) dx = 1}} \int_0^T \int_{\Omega} (|\dot{\rho}(t, x)| + V(t, x)\rho(t, x) + f(\rho(t, x))) dx dt + \psi_0(\rho(0)) + \psi_T(\rho(T)) := F(\rho), \quad (4.7)$$

where $\mathcal{E} := \text{BV}([0, T]; L^1(\Omega)) \cap L^2([0, T] \times \Omega)$ (for one-variable BV functions valued in a Banach space such as L^1 we refer, for instance, to [Mor87; Mor89]). The function f will be supposed to be uniformly convex (i.e. $f'' \geq c_0 > 0$), and the time-dependent potential V will be supposed to belong to $\text{Lip}([0, T]; L^2(\Omega))$. The domain Ω is a finite measure set in \mathbb{R}^d , and we assume for simplicity that it has unit volume.

A few words on the penalizations ψ_0 and ψ_T :

We aim to apply the results when ψ_T is of the following form

$$\begin{aligned} \psi_T: L^1(\Omega) &\rightarrow \mathbb{R} \\ \rho &\mapsto \int_{\Omega} \varphi_T(x) \rho(x) dx, \end{aligned}$$

where $\varphi_T: \Omega \rightarrow \mathbb{R}$ is L^∞ and $\|\varphi_T\|_\infty \leq 1$. In this case, ψ_T is 1 -Lipschitz for the norm $\|\cdot\|_{L^1(\Omega)}$, and it is also weakly continuous in $L^1(\Omega)$ (and hence in $L^2(\Omega)$).

When it comes to ψ_0 , we aim to consider the following case:

$$\begin{aligned} \psi_0: L^1(\Omega) &\rightarrow \mathbb{R} \\ \rho &\mapsto \|\rho - m_0\|_{L^1(\Omega)}, \end{aligned}$$

where $m_0 \in L^1(\Omega)$. In this case as well, ψ_0 is 1 -Lipschitz for the norm $\|\cdot\|_{L^1(\Omega)}$. It is also weakly lower semi-continuous in $L^1(\Omega)$ and $L^2(\Omega)$.

More generally, our results apply when the penalizations ψ_0 and ψ_1 are of the following form:

$$\psi_t(\rho) := \int_{\Omega} a_t(x, \rho(x)) dx,$$

for two functions a_t which are 1 -Lipschitz continuous and convex in the second variable. In this way the functionals ψ_t are both continuous for the strong L^1 convergence (actually, Lip_1) and lower semi-continuous for the weak L^1 convergence. This general framework includes the two examples above.

Finally, if ψ_0 and ψ_T are defined as previously, we will define $\psi_{0,\alpha} := \alpha\psi_0$ and $\psi_{T,\alpha} := \alpha\psi_T$.

Approximations:

As in the previous section, the absolute value $|\dot{\rho}|$ will be approximated by a smoother function L_ε which will be specified later. The mass constraint $\int \rho_t = 1$ will be imposed via a penalization method, adding $\frac{(\int \rho - 1)^2}{2\delta}$. The penalizations on the boundary on $[0, T]$, ψ_0 and ψ_T , will also be approximated by multiplying them by $\alpha < 1$. Finally, the positivity constraint will be handled by approximating f with a sequence $f_n : \mathbb{R} \rightarrow \mathbb{R}$ obtained via $f_n(\rho) := f(\rho) + n(\rho_-)^2$ so that negative values are penalized but the uniform convexity of f is preserved since we have $f_n'' \geq c_0$ for the same $c_0 > 0$.

Taking together these approximations, we obtain the following problem:

$$\min_{\rho \in \mathcal{E}} \int_0^T \left(\int_\Omega L_\varepsilon(\dot{\rho}(t, x)) + f_n(\rho(t, x)) + V(t, x)\rho(t, x) dx + \frac{(\int \rho(t, y) dy - 1)^2}{2\delta} \right) dt + \psi_{0,\alpha}(\rho(0)) + \psi_{T,\alpha}(\rho(T)) := F_n(\rho).$$

Note that the approximated functional is written as F_n , meaning that we have fixed a suitable sequence of values $(\varepsilon, \delta, \alpha)$ such that $\varepsilon_n \rightarrow 0$, $\delta_n \rightarrow 0$ and $\alpha_n \rightarrow 1$.

In order not to have difficulties with the infinite-dimensional space $L^1(\Omega)$, we will also use a finite-dimensional discretization. This consists in imposing that the functions $\rho(t)$ belong to a finite-dimensional subspace. More precisely, we divide the space Ω into n small areas of volume $\frac{1}{n}$ called A_i^n (we need their diameter to tend to 0) and we take $\rho_n(t) : \Omega \rightarrow \mathbb{R}$ such that it is constant on each area. This means that $\rho_n(t)$ takes at most n different values and its mass is constant equal to $\rho_i(t)$ on each region A_i^n (and its density equals $n\rho_i(t)$). The problem can be considered as a restriction of the previous one to the subset of \mathcal{E} composed of densities which are piecewise constant functions (constant on each A_i^n) for every t , or it can be rewritten as follows:

$$\min_{\rho \in BV([0, T]; \mathbb{R}^n)} \int_0^T \left(\left(\sum_{i=1}^n \frac{1}{n} (L_\varepsilon(n\dot{\rho}_i(t)) + f_n(n\rho_i(t))) + \rho_i(t) \int_{A_i^n} V(t, x) dx \right) + \frac{(\sum_j \rho_j(t) - 1)^2}{2\delta} \right) dt + \psi_{0,\alpha}(\rho(0)) + \psi_{T,\alpha}(\rho(T)),$$

where the functionals $\psi_{0,\alpha}$ and $\psi_{T,\alpha}$ can also be written in terms of values $\rho_i(0)$ and $\rho_i(T)$: they are of the form

$$\psi_{0,\alpha}(\rho(0)) := \alpha \sum_i a_{i,0}^n(\rho_i(0)) \quad \text{and} \quad \psi_{T,\alpha}(\rho(T)) := \alpha \sum_i a_{i,T}^n(\rho_i(T))$$

for some Lip_1 functions $a_{i,t}^n$, which are precisely given by

$$a_{i,t}^n(u) := \frac{1}{n} \int_{A_i^n} a_t(x, nu) dx.$$

In the set \mathcal{E} , we say that ρ_n converges to ρ in \mathcal{E} in the sense of (4.8), if

$$\exists C \text{ s.t. } \|\rho_n(t)\|_{L^2(\Omega)} \leq C \text{ for every } n \text{ and every } t \text{ and } \rho_n(t) \rightharpoonup \rho(t) \text{ uniformly in } t \quad (4.8)$$

where the uniform L^2 bound allows to metrize the weak L^2 convergence and the uniform convergence is defined accordingly.

We observe that the convergence in the sense of (4.8) implies the weak convergence $\rho_n \rightharpoonup \rho$ weakly in $L^2([0, T] \times \Omega)$.

Lemma 4.2.1. *Let $(\rho_n)_n$ be a sequence converging to $\rho \in \mathcal{E}$ in the sense of (4.8) such that $F_n(\rho_n)$ is bounded. Then, we have $\rho \geq 0$, for a.e. $t \in [0, T]$ we have $\int_\Omega \rho(t, x) dx = 1$, and moreover*

$$F(\rho) \leq \liminf_n F_n(\rho_n).$$

Proof. First, we have

$$\int_0^T \int_\Omega |\dot{\rho}_n(t, x)| dx dt \leq \int_0^T \int_\Omega L_\varepsilon(\dot{\rho}_n(t, x)) dx dt \leq F_n(\rho_n) \leq C,$$

so that $\|\dot{\rho}_n\|_{L^1([0, T] \times \Omega)}$ is bounded. By embedding $L^1([0, T] \times \Omega)$ into $\mathcal{M}([0, T] \times \Omega)$, the space of Radon measures, there exists a subsequence of $(\dot{\rho}_n)_n$ which converges weakly in $\mathcal{M}([0, T] \times \Omega)$ towards a measure which can only be $\dot{\rho}$. The semi-continuity of the mass for the weak convergence provides

$$\int_0^T \int_\Omega |\dot{\rho}(t, x)| dx dt \leq \liminf_{n \rightarrow \infty} \|\dot{\rho}_n\|_{L^1([0, T] \times \Omega)} \leq \int_0^T \int_\Omega L_\varepsilon(\dot{\rho}_n(t, x)) dx dt$$

(where, actually, the first integral is to be intended as a mass in the sense of measures, or as the total variation of the curve $t \mapsto \rho(t)$ in $L^1(\Omega)$). Then, using again $\rho_n \xrightarrow{L^2} \rho$, we obtain

$$\int_0^T \int_{\Omega} V(t, x) \rho(t, x) dx dt = \lim_{n \rightarrow \infty} \int_0^T \int_{\Omega} V(t, x) \rho_n(t, x) dx dt.$$

Using $f \leq f_n$ and the convexity of f , which implies the weak lower semi-continuity of $\rho \mapsto \iint f(\rho(t, x)) dx dt$, we have

$$\int_0^T \int_{\Omega} f(\rho(t, x)) dx dt \leq \liminf_{n \rightarrow \infty} \int_0^T \int_{\Omega} f_n(\rho_n(t, x)) dx dt.$$

Moreover, by the definition of f_n , we also obtain

$$\int_0^T \int_{\Omega} (\rho_-)^2(t, x) dx dt \leq \liminf_n \int_0^T \int_{\Omega} ((\rho_n)_-)^2(t, x) dx dt = 0$$

since $\iint ((\rho_n)_-)^2(t, x) dx dt \leq \frac{C}{n}$. This shows $\rho \geq 0$.

Since for all $t \in [0, T]$, $\rho_n(t) \rightharpoonup \rho(t)$ weakly in $L^2(\Omega)$, we have in particular $\rho_n(0) \rightharpoonup \rho(0)$ and $\rho_n(T) \rightharpoonup \rho(T)$ weakly in $L^2(\Omega)$ and consequently weakly in $L^1(\Omega)$, so by the lower semi-continuity of ψ_0 and ψ_T , we have

$$\psi_0(\rho(0)) \leq \liminf_n \psi_0(\rho_n(0)) \quad \text{and} \quad \psi_T(\rho(T)) \leq \liminf_n \psi_T(\rho_n(T)).$$

Since $(\|\rho_n(0)\|_{L^2(\Omega)})_n$ and $(\|\rho_n(T)\|_{L^2(\Omega)})_n$ are bounded we also have L^1 bounds and hence

$$\lim_n \alpha_n \psi_0(\rho_n(0)) = \lim_n \psi_0(\rho_n(0)) \quad \text{and} \quad \lim_n \alpha_n \psi_T(\rho_n(T)) = \lim_n \psi_T(\rho_n(T)).$$

Finally, we use the positivity of the term

$$\int_0^T \frac{(\int \rho_n(t, y) dy - 1)^2}{2\delta} dt$$

to obtain

$$F(\rho) \leq \liminf_n F_n(\rho_n),$$

and its boundedness to obtain $\int \rho(t) = \lim \int \rho_n(t) = 1$ for a.e. t . \square

The following lemmas will be useful for the existence and regularity of the solution to (4.7) and will be referred to when needed.

Lemma 4.2.2. *Suppose that m_0 is such that $\int f(m_0) < +\infty$ (in particular $m_0 \in L^2(\Omega)$). For all $\rho \in \mathcal{E}$ there exists a sequence $(\rho_n)_n$ in $D = H^1([0, T]; L^2(\Omega))$ which converges to ρ strongly in $L^2_{t,x}$ and which satisfies*

$$\rho_n(0) = m_0$$

$$\begin{aligned} \|\dot{\rho}_n\|_{L^1([0, T] \times \Omega)} &\leq \|\dot{\rho}\|_{L^1([0, T] \times \Omega)} + \|\rho(0^+) - m_0\|_{L^1(\Omega)} \\ &= \|\dot{\rho}\|_{L^1([0, T] \times \Omega)} + \psi_0(\rho(0^+)) \end{aligned} \quad (4.9)$$

$$\limsup_n \int_0^T \int_{\Omega} f(\rho_n(t, x)) dx dt \leq \int_0^T \int_{\Omega} f(\rho(t, x)) dx dt \quad (4.10)$$

$$\lim_{n \rightarrow \infty} \int_0^T \int_{\Omega} \rho_n(t, x) V(t, x) dx dt = \int_0^T \int_{\Omega} \rho(t, x) V(t, x) dx dt \quad (4.11)$$

$$\lim_n \psi_T(\rho_n(T)) = \psi_T(\rho(T^-)) \quad (4.12)$$

In addition, if for all $t \in [0, T]$, $\int_{\Omega} \rho(t, x) dx = 1$, then for all $n \in \mathbb{N}$ and all $t \in [0, T]$, $\int_{\Omega} \rho_n(t, x) dx = 1$.

Proof. Let $\rho \in \mathcal{E}$. We denote by η_n a sequence of mollifiers in the t variable (i.e. η_n is a sequence of smooth probability measures on \mathbb{R} such that $\eta_n \rightharpoonup \delta_0$), and we suppose $\text{spt} \eta_n \subset [0, \frac{1}{n}]$.

The function ρ is only defined on the time interval $[0, T]$, but we can extend ρ to a function $\tilde{\rho} : [-1, T] \times \Omega \rightarrow \mathbb{R}$ setting $\tilde{\rho}(t) = m_0$ for all $t < 0$. The total variation in L^1 of this extension is equal to the sum of that of ρ and the L^1 distance between m_0 and $\rho(0^+)$.

We then define ρ_n as the convolution of $\tilde{\rho}$ with the mollifiers η_n in time:

$$\rho_n(t, x) := \int \eta_n(t - s) \tilde{\rho}(s, x) ds.$$

With the assumption on the support of η_n this convolution is well-defined on $[0, T]$ even if $\tilde{\rho}$ has not been extended on $]T, +\infty[$. Moreover, we have $\rho_n(0) = m_0$.

Note that we have

$$|\dot{\rho}_n(t, x)| = \left| \int \eta'_n(t-s) \tilde{\rho}(s, x) ds \right| \leq C(n) \int_{t-1/n}^t \tilde{\rho}(s, x) ds$$

so that we have

$$\|\dot{\rho}_n(t)\|_{L^2(\Omega)} \leq C(n) \int_{t-1/n}^t \|\tilde{\rho}(s)\|_{L^2(\Omega)} ds \leq C(n)$$

(we use here the assumption $m_0 \in L^2(\Omega)$). This proves $\rho_n \in H^1([0, T]; L^2(\Omega))$.

The convexity of the total variation easily implies (4.9). Again by convexity, for any convex function g we have

$$\int_0^T \int_{\Omega} g(\rho_n(t, x)) dx dt \leq \int_{-1/n}^T \int_{\Omega} g(\tilde{\rho}(t, x)) dx dt \rightarrow \int_0^T \int_{\Omega} g(\rho(t, x)) dx dt,$$

where the last limit is valid whenever $\int_{\Omega} g(m_0) < +\infty$. Applying this to $g(\rho) = \rho^2$ proves that ρ_n strongly convergence in $L^2_{t,x}$ to ρ (it provides an L^2 bound, hence a weak limit up to subsequences; this weak limit can be identified as ρ by testing against continuous functions; the limit is actually strong because the L^2 norm converges to that of the limit). This implies, in particular, (4.11). As for (4.10), it is enough to use $g = f$.

To prove (4.12), we use the property of bounded variation functions: ρ admits a left limit at T in L^1 , i.e for every $\varepsilon > 0$ there exists $\delta > 0$ such that $\|\rho(t) - \rho(T^-)\|_{L^1} \leq \varepsilon$ for every $t \in [T - \delta, T]$. By convexity this implies, as soon as $\frac{1}{n} < \delta$, $\|\rho_n(T) - \rho(T^-)\| \leq \varepsilon$ and shows $\rho_n(T) \rightarrow \rho(T^-)$ strongly in L^1 . This implies (4.12).

If we suppose in addition that for all $t \in [0, T]$, $\int_{\Omega} \rho(t, x) dx = 1$ the same will be true for ρ_n by convexity, which concludes the proof of the statement. \square

Lemma 4.2.3. *Let $(\rho_n)_n$ be a sequence in \mathcal{E} . Suppose that there exists $C_1 > 0$ and $C_2 > 0$ such that*

$$\forall n \in \mathbb{N}^*, \quad \sup_{t \in [0, T]} \|\dot{\rho}_n(t)\|_{L^2(\Omega)} \leq C_1, \quad (4.13)$$

and

$$\int_0^T \|\rho_n(t)\|_{L^2(\Omega)}^2 dt \leq C_2. \quad (4.14)$$

Then, there exists $C_3 > 0$ such that

$$\forall n \in \mathbb{N}^*, \quad \sup_{t \in [0, T]} \|\rho_n(t)\|_{L^2(\Omega)} \leq C_3.$$

Proof. Hypothesis (4.13) implies,

$$\forall t, s \in [0, T], \quad \|\rho_n(t) - \rho_n(s)\|_{L^2(\Omega)} \leq C_1 |t - s|. \quad (4.15)$$

By the continuity of $\|\rho_n(\cdot)\|_{L^2(\Omega)}$, there exists a time $t_{0,n} \in [0, T]$ at which the mean-value is reached:

$$\|\rho_n(t_{0,n})\|_{L^2(\Omega)}^2 = \frac{\int_0^T \|\rho_n(t)\|_{L^2(\Omega)}^2 dt}{T},$$

so by Hypothesis (4.14),

$$\|\rho_n(t_{0,n})\|_{L^2(\Omega)}^2 \leq \frac{C_2}{T}. \quad (4.16)$$

By (4.15) and (4.16), we have for all $t \in [0, T]$,

$$\|\rho_n(t)\|_{L^2(\Omega)} \leq \|\rho_n(t) - \rho_n(t_{0,n})\|_{L^2(\Omega)} + \|\rho_n(t_{0,n})\|_{L^2(\Omega)} \leq C_1 T + \frac{C_2}{T} := C_3.$$

\square

Theorem 4.2.4. *Suppose that $\psi_0: L^1(\Omega) \rightarrow \mathbb{R}$ and $\psi_T: L^1(\Omega) \rightarrow \mathbb{R}$ are 1-Lipschitz and weakly lower semi-continuous on $L^1(\Omega)$ and that $V: [0, T] \times \Omega \rightarrow \mathbb{R}$ belongs to $\text{Lip}([0, T]; L^2(\Omega))$. Suppose also that $f: \mathbb{R} \rightarrow \mathbb{R}$ is c_0 -convex, i.e $f'' \geq c_0$ on \mathbb{R} .*

Then, there exists a unique minimizer ρ to the problem

$$\min_{\substack{\rho \in \mathcal{E}, \rho \geq 0 \\ \forall t \in [0, T], \int_{\Omega} \rho(t, x) dx = 1}} F(\rho). \quad (4.17)$$

This solution belongs to $\text{Lip}([0, T]; L^2(\Omega))$ and it satisfies

$$\sup_{t \in [0, T]} \int_{\Omega} |\dot{\rho}(t, x)|^2 dx \leq C, \quad (4.18)$$

where $C = \frac{C_0^2}{c_0^2}$ with $C_0^2 = \sup_{t \in [0, T]} \|V'(t, \cdot)\|_{L^2(\Omega)}^2$.

Proof. To obtain the regularity (4.18), we will approximate the problem (4.17) as already described at the beginning of this section; i.e. solving

$$\begin{aligned} \min_{\rho \in \mathcal{E}_n} F_n(\rho) := & \int_0^T \left(\left(\sum_{i=1}^n \frac{1}{n} (L_{\varepsilon}(n\rho_i(t)) + f_n(n\rho_i(t))) + n\rho_i(t) \int_{A_i^n} V(t, x) dx \right) + \frac{(\sum_j \rho_j(t) - 1)^2}{2\delta} \right) dt \\ & + \psi_{0, \alpha}(\rho(0)) + \psi_{T, \alpha}(\rho(T)), \end{aligned} \quad (4.19)$$

where we denote by \mathcal{E}_n the set of piecewise constant (in space) functions $\rho \in \mathcal{E}$ such that

$$\forall t \in [0, T], \forall x \in \Omega, \rho(t, x) = \sum_{i=1}^n n\rho_i(t) \mathbb{1}_{A_i^n}(x),$$

where $\rho_i: [0, T] \rightarrow \mathbb{R}$ is a real-valued function. With the definition, the mass of ρ on each A_i^n equals $\rho_i(t)$.

In particular, if we consider the curves $\rho_i(t)$, for each i the curve ρ_i solves

$$\begin{aligned} \min_{\rho_i \in H^1([0, T]; \mathbb{R})} \int_0^T & \left(L_{\varepsilon}(n\rho_i(t)) + f_n(n\rho_i(t)) + n\rho_i(t) V_i^n(t) + \frac{(\sum_j \rho_j(t) - 1)^2}{2\delta} \right) dt \\ & + \alpha a_{i,0}^n(\rho_i(0)) + \alpha a_{i,T}(\rho_i(T)) \end{aligned} \quad (4.20)$$

where $V_i^n(t) := \frac{1}{|A_i^n|} \int_{A_i^n} V(t, x) dx = n \int_{A_i^n} V(t, x) dx$. The function $V^n(t)$ defined to be equal to $V_i^n(t)$ in A_i^n satisfies

$$\|V^n(t)\|_{L^2(\Omega)} \leq \|V(t)\|_{L^2(\Omega)}, \quad \|V^n\|_{L^2([0, T] \times \Omega)} \leq \|V\|_{L^2([0, T] \times \Omega)}, \quad \|\partial_t V^n(t)\|_{L^2(\Omega)} \leq \|\partial_t V(t)\|_{L^2(\Omega)}.$$

All these inequalities are a consequence of Jensen's inequality. Indeed, we have for all $n \in \mathbb{N}$ and $t \in [0, T]$,

$$\begin{aligned} \int_{A_i^n} V_i^n(t, x)^2 dx &= \int_{A_i^n} n^2 \left(\int_{A_i^n} V(t, y) dy \right)^2 dx = n \left(\int_{A_i^n} V(t, y) dy \right)^2 = \frac{1}{n} \left(n \int_{A_i^n} V(t, y) dy \right)^2 \\ &\leq \frac{1}{n} \left(n \int_{A_i^n} V(t, y)^2 dy \right) = \int_{A_i^n} V(t, y)^2 dy. \end{aligned}$$

The proof is divided in several steps.

STEP 1 The minimizers of (4.19), which is a finite-dimensional variational problem in H^1 , exist by the direct method. A presentation of this method is described for example in [San23, Section 1.2.1] or in [Dac07, Section 3.4.1].

STEP 2 Let ρ_n be a minimizer of F_n for all n . In this step, we bound $\|\dot{\rho}_n(t)\|_{L^2(\Omega)}$ independently of n so that we will be able to pass to the limit $n \rightarrow \infty$. We remind that ρ_n is piecewise constant in space: $\rho_n(t, x) := \sum_{i=1}^n n\rho_{n,i}(t) \mathbb{1}_{A_i^n}(x)$. In the following, we fix n and write $\rho_i(t)$ instead of $\rho_{n,i}(t)$ to enlighten the notation.

The Euler-Lagrange equation of (4.20) is

$$(L'_{\varepsilon}(n\rho_i(t)))' = V_i^n(t) + f'_n(n\rho_i(t)) + \frac{(\sum_j \rho_j(t) - 1)}{n\delta}. \quad (4.21)$$

Differentiating the equation (4.21) yields

$$(L'_\varepsilon(n\dot{\rho}_i(t)))'' = (V_i^n)'(t) + n\dot{\rho}_i(t)f''_n(n\rho_i(t)) + \frac{\sum_j \dot{\rho}_j(t)}{n\delta}.$$

Multiplying by $\dot{\rho}_i(t)$ and summing over i gives

$$\sum_{i=1}^n \dot{\rho}_i(t)(L'_\varepsilon(n\dot{\rho}_i(t)))'' = \sum_{i=1}^n (V_i^n)'(t)\dot{\rho}_i(t) + n(\dot{\rho}_i(t))^2 f''_n(n\rho_i(t)) + \frac{\left(\sum_j \dot{\rho}_j(t)\right)^2}{n\delta}.$$

Since the term $\left(\sum_j \dot{\rho}_j(t)\right)^2$ is positive, using $f''_n \geq c_0$ as well, we obtain the inequality

$$\sum_{i=1}^n \dot{\rho}_i(t) (L'_\varepsilon(n\dot{\rho}_i(t)))'' \geq \sum_{i=1}^n (V_i^n)'(t)\dot{\rho}_i(t) + c_0 n(\dot{\rho}_i(t))^2. \quad (4.22)$$

Now, we need to estimate the left-hand side of (4.22). By expanding the second derivative, we have

$$\begin{aligned} \sum_{i=1}^n \dot{\rho}_i(t) (L'_\varepsilon(n\dot{\rho}_i(t)))'' &= \sum_{i=1}^n \dot{\rho}_i(t) (n\ddot{\rho}_i(t)L''_\varepsilon(n\dot{\rho}_i(t)))' \\ &= \sum_{i=1}^n \dot{\rho}_i(t) \left((n\ddot{\rho}_i(t))^2 L'''_\varepsilon(n\dot{\rho}_i(t)) + n\ddot{\rho}_i(t)L''_\varepsilon(n\dot{\rho}_i(t)) \right). \end{aligned} \quad (4.23)$$

Let us define the function $h: \mathbb{R} \rightarrow \mathbb{R}$ such that $h(s) = sL'_\varepsilon(s) - L_\varepsilon(s)$. Then, h verifies

$$h'(s) = sL''_\varepsilon(s) \quad \text{and} \quad h''(s) = L''_\varepsilon(s) + sL'''_\varepsilon(s).$$

We now consider

$$\max_{t \in [0, T]} \sum_{i=1}^n h(n\dot{\rho}_i(t)). \quad (4.24)$$

Two cases will be distinguished:

- the maximum is reached on $(0, T)$,
- the maximum is reached on $\{0, T\}$.

(i) Suppose there exists $t_0 \in]0, T[$ such that $\sum_{i=1}^n h(n\dot{\rho}_i(t_0)) = \max_{t \in [0, T]} \sum_{i=1}^n h(n\dot{\rho}_i(t))$.

In particular, we have

$$\begin{aligned} \sum_{i=1}^n n\ddot{\rho}_i(t_0)h'(n\dot{\rho}_i(t_0)) &= \sum_{i=1}^n n^2\ddot{\rho}_i(t_0)\dot{\rho}_i(t_0)L''_\varepsilon(n\dot{\rho}_i(t_0)) = 0 \\ \text{and} \quad \sum_{i=1}^n n\ddot{\rho}_i(t_0)h'(n\dot{\rho}_i(t_0)) &+ (n\ddot{\rho}_i(t_0))^2 h''(n\dot{\rho}_i(t_0)) \leq 0, \\ \text{i.e.} \quad \sum_{i=1}^n n^2\ddot{\rho}_i(t_0)\dot{\rho}_i(t_0)L''_\varepsilon(n\dot{\rho}_i(t_0)) &+ (n\ddot{\rho}_i(t_0))^2 (L''_\varepsilon(n\dot{\rho}_i(t_0)) + n\dot{\rho}_i(t_0)L'''_\varepsilon(n\dot{\rho}_i(t_0))) \leq 0. \end{aligned}$$

Inserting (4.22) and (4.23) in t_0 , the last inequality becomes

$$\sum_{i=1}^n \left((V_i^n)'(t_0)\dot{\rho}_i(t_0) + c_0 n\dot{\rho}_i(t_0)^2 \right) \leq - \sum_{i=1}^n (n\ddot{\rho}_i(t_0))^2 L''_\varepsilon(n\dot{\rho}_i(t_0)) \leq 0. \quad (4.25)$$

Let us precise the expression of h :

$$\begin{aligned} h(s) &= sL'_\varepsilon(s) - L_\varepsilon(s) \\ &= \frac{s^2}{\sqrt{s^2 + \varepsilon^2}} + 2\varepsilon s^2 - \sqrt{s^2 + \varepsilon^2} - \varepsilon s^2 = \frac{s^2 - s^2 - \varepsilon^2}{\sqrt{s^2 + \varepsilon^2}} + \varepsilon s^2 \\ &= -\frac{\varepsilon^2}{\sqrt{s^2 + \varepsilon^2}} + \varepsilon s^2 < \varepsilon s^2. \end{aligned}$$

Since t_0 is a maximizer of $\sum_{i=1}^n h(n\dot{\rho}_i)$, for all $t \in [0, T]$, we obtain

$$\begin{aligned} \sum_{i=1}^n h(n\dot{\rho}_i(t)) &= \sum_{i=1}^n -\frac{\varepsilon^2}{\sqrt{(n\dot{\rho}_i(t))^2 + \varepsilon^2}} + \varepsilon(n\dot{\rho}_i(t))^2 \\ &\leq \sum_{i=1}^n h(n\dot{\rho}_i(t_0)) < \varepsilon \sum_{i=1}^n (n\dot{\rho}_i(t_0))^2. \end{aligned} \quad (4.26)$$

In particular, we have

$$\sum_{i=1}^n -\frac{\varepsilon}{\sqrt{(n\dot{\rho}_i(t))^2 + \varepsilon^2}} + (n\dot{\rho}_i(t))^2 < \sum_{i=1}^n (n\dot{\rho}_i(t_0))^2.$$

Since $\frac{\varepsilon}{\sqrt{(n\dot{\rho}_i(t))^2 + \varepsilon^2}} \leq 1$, for all $t \in [0, T]$,

$$\sum_{i=1}^n n\dot{\rho}_i(t)^2 < 1 + \sum_{i=1}^n n\dot{\rho}_i(t_0)^2. \quad (4.27)$$

Besides, thanks to Cauchy-Schwarz inequality, the inequality (4.25) becomes

$$\begin{aligned} c_0 \sum_{i=1}^n n\dot{\rho}_i(t_0)^2 &\leq -\sum_{i=1}^n (V_i^n)'(t_0)\dot{\rho}_i(t_0) \\ &\leq \sqrt{\sum_{i=1}^n \frac{1}{n} (V_i^n)'(t_0)^2} \sqrt{\sum_{i=1}^n n\dot{\rho}_i(t_0)^2} \\ &\leq \sqrt{\sup_{t \in [0, T]} \sum_{i=1}^n \int_{A_i^n} (V^n)'(t)^2 dx} \sqrt{\sum_{i=1}^n n\dot{\rho}_i(t_0)^2} \\ &\leq \sup_{t \in [0, T]} \|(V^n)'(t, \cdot)\|_{L^2(\Omega)} \sqrt{\sum_{i=1}^n n\dot{\rho}_i(t_0)^2} \\ &\leq \sup_{t \in [0, T]} \|V'(t, \cdot)\|_{L^2(\Omega)} \sqrt{\sum_{i=1}^n n\dot{\rho}_i(t_0)^2}. \end{aligned} \quad (4.28)$$

Finally, (4.28) gives

$$c_0^2 \sum_{i=1}^n n\dot{\rho}_i(t_0)^2 \leq \sup_{t \in [0, T]} \|V'(t, \cdot)\|_{L^2(\Omega)}^2. \quad (4.29)$$

By gathering the inequalities (4.27) and (4.29), there exists a constant $C_0^2 := \sup_{t \in [0, T]} \|V'(t, \cdot)\|_{L^2(\Omega)}^2 \geq 0$ independent from n such that for all $t \in [0, T]$,

$$\sum_{i=1}^n n\dot{\rho}_i(t)^2 \leq 1 + \sum_{i=1}^n n\dot{\rho}_i(t_0)^2 \leq 1 + \frac{C_0^2}{c_0^2}. \quad (4.30)$$

Yet, for all $t \in [0, T]$, the L^2 -norm of $\dot{\rho}_n(t)$ is

$$\|\dot{\rho}_n(t)\|_{L^2(\Omega)}^2 = \sum_{i=1}^n \int_{A_i^n} (n\dot{\rho}_i(t))^2 = \sum_{i=1}^n n\dot{\rho}_i(t)^2,$$

so, for all $n \in \mathbb{N}$, a minimizer of (4.20) verifies

$$\sup_{t \in [0, T]} \|\dot{\rho}_n(t)\|_{L^2(\Omega)}^2 \leq 1 + \frac{C_0^2}{c_0^2}. \quad (4.31)$$

(ii) Suppose that the maximum of (4.24) is reached at 0, i.e $\sum_{i=1}^n h(n\dot{\rho}_i(0)) = \max_{t \in [0, T]} \sum_{i=1}^n h(n\dot{\rho}_i(t))$.

The transversality condition yields

$$L'_\varepsilon(n\dot{\rho}_i(0)) = (\alpha a_{i,0}^n)'(\rho_i(0)),$$

and we know that $\alpha a_{i,0}^n$ is Lip_α . Using $L'_\varepsilon(s) = \frac{s}{\sqrt{s^2 + \varepsilon^2}} + 2\varepsilon s$, we get

$$\left| \frac{n\dot{\rho}_i(0)}{\sqrt{(n\dot{\rho}_i(0))^2 + \varepsilon^2}} + 2\varepsilon n\dot{\rho}_i(0) \right| \leq \alpha.$$

This inequality is equivalent to

$$\left(\frac{1}{\sqrt{(n\dot{\rho}_i(0))^2 + \varepsilon^2}} + 2\varepsilon \right) |n\dot{\rho}_i(0)| \leq \alpha$$

which yields

$$\begin{aligned} \frac{1}{\sqrt{(n\dot{\rho}_i(0))^2 + \varepsilon^2}} |n\dot{\rho}_i(0)| &\leq \alpha, \\ \text{i.e. } |n\dot{\rho}_i(0)| &\leq \alpha \sqrt{(n\dot{\rho}_i(0))^2 + \varepsilon^2}, \\ \text{i.e. } |n\dot{\rho}_i(0)|^2 &\leq \alpha^2 [(n\dot{\rho}_i(0))^2 + \varepsilon^2], \\ \text{i.e. } (1 - \alpha^2) |n\dot{\rho}_i(0)|^2 &\leq \alpha^2 \varepsilon^2. \end{aligned}$$

This leads to

$$|n\dot{\rho}_i(0)| \leq \frac{\alpha\varepsilon}{\sqrt{1 - \alpha^2}}.$$

(Note that we could have directly referred to the computation in Proposition 4.1.2.)

Thanks to this inequality, we can conclude similarly to (4.30) that

$$\begin{aligned} \sum_{i=1}^n n\dot{\rho}_i(t)^2 &< 1 + \sum_{i=1}^n n\dot{\rho}_i(0)^2, \\ &\leq 1 + \frac{n\alpha^2\varepsilon^2}{n(1 - \alpha^2)} = 1 + \frac{\alpha^2\varepsilon^2}{1 - \alpha^2}, \end{aligned}$$

so, if $\alpha \rightarrow 1$ and $\varepsilon \rightarrow 0$ in a way that $\frac{\alpha^2\varepsilon^2}{1 - \alpha^2}$ remains below 1, we have

$$\sup_{t \in [0, T]} \|\dot{\rho}_n(t)\|_{L^2(\Omega)}^2 = \sup_{t \in [0, T]} \sum_{i=1}^n n\dot{\rho}_i(t)^2 \leq 2. \quad (4.32)$$

Remark 4.2.5. *The upper bounds in (4.31) and (4.32) can be improved to quantities which are arbitrarily close to C_0^2/c_0^2 and 0, respectively.*

For the first one, it suffices to define a slightly different approximation of $|\cdot|$. If one take $L_{\varepsilon, A}(s) = \sqrt{s^2 + \varepsilon^2} + A\varepsilon s^2$ instead of L_ε , the calculus remains the same, except for (4.26), which becomes

$$\begin{aligned} \sum_{i=1}^n h(n\dot{\rho}_i(t)) &= \sum_{i=1}^n -\frac{\varepsilon^2}{\sqrt{(n\dot{\rho}_i(t))^2 + \varepsilon^2}} + A\varepsilon(n\dot{\rho}_i(t))^2 \\ &\leq \sum_{i=1}^n h(n\dot{\rho}_i(t_0)) < A\varepsilon \sum_{i=1}^n (n\dot{\rho}_i(t_0))^2, \end{aligned}$$

and yields the inequality

$$\begin{aligned} A \sum_{i=1}^n n\dot{\rho}_i(t)^2 &< 1 + A \sum_{i=1}^n n\dot{\rho}_i(t_0)^2, \\ \text{i.e. } \sum_{i=1}^n n\dot{\rho}_i(t)^2 &< \frac{1}{A} + \sum_{i=1}^n n\dot{\rho}_i(t_0)^2. \end{aligned}$$

Since the constant A can be chosen as large as we want, the bound can be C_0^2/c_0^2 .

For the second estimate, it is enough to choose $\alpha \rightarrow 1$ such that $\frac{\alpha^2\varepsilon^2}{1 - \alpha^2} \rightarrow 0$ which allows to replace the second term in (4.32) by almost 0 (and the first one can be taken small as well, as we have just explained).

The same computations lead to the same conclusion if the maximum were reached on T .

To conclude, by (4.31), (4.32) and Remark 4.2.5, we have shown that

$$\forall n, \sup_{t \in [0, T]} \|\dot{\rho}_n(t)\|_{L^2(\Omega)}^2 \leq \max \left\{ \frac{C_0^2}{c_0^2}, 0 \right\} = \frac{C_0^2}{c_0^2} := C.$$

STEP 3 In this step, we prove that there exists a subsequence of $(\rho_n)_n$ which converges to a function $\rho \in C^0([0, T], L^2(\Omega))$.

The sequence $(\rho_n(t))_n$ is equi-Lipschitz in $L^2(\Omega)$ but the norm $L^2_{t,x}$ is also bounded, which provides a uniform bound $\|\rho_n(t)\|_{L^2(\Omega)} \leq C$ (see Lemma 4.2.3). Bounded sets in $L^2(\Omega)$ are not compact for the strong convergence, but they are compact, by Banach-Alaoglu's theorem, for the weak convergence. The equicontinuity that we have is in the strong sense, so it also holds in the weak sense, and we can then apply the Arzelà-Ascoli's theorem to extract a subsequence of $(\rho_n)_n$ which converges to ρ in the sense of (4.8).

Additionally, we obtain by the lower semi-continuity property of the L^2 -norm that the limit curve ρ is also Lipschitz in time for the strong L^2 norm:

$$\forall (t, s) \in [0, T]^2, \|\rho(t) - \rho(s)\|_{L^2(\Omega)} \leq C|t - s|.$$

STEP 4 The goal of this step is to prove that the limit ρ found previously is actually a minimizer of (4.17). Comparing to a suitable competitor (for instance a discretization of m_0 , constant in time), we can see that there exists $C > 0$ such that $F_n(\rho_n) \leq C$ for all n .

First, let us notice that $\int_{\Omega} \rho(t, x) dx = 1$, for all $t \in [0, T]$. Indeed, we have

$$\int_0^T \left(\int_{\Omega} \rho(t, x) dx - 1 \right)^2 dt \leq \liminf_n \int_0^T \left(\int_{\Omega} \rho_n(t, x) dx - 1 \right)^2 dt = 0.$$

Second, by Lemma 4.2.1, we have that $F(\rho) \leq \liminf_n F_n(\rho_n)$.

Third, let $m \in D = H^1([0, T]; L^2(\Omega))$ be such that $\int_{\Omega} m(t, y) dy = 1$. In what follows, we prove that there exists a sequence $(m_n)_n$ such that $\limsup_n F_n(m_n) \leq F(m)$.

We define the sequence $(m_n)_n$ by choosing piecewise constant functions such that

$$\forall t \in [0, T], \forall x \in A_i^n, m_n(t, x) = n \int_{A_i^n} m(t, y) dy \quad (4.33)$$

which verifies $\int_{\Omega} m_n(t, y) dy = \int_{\Omega} m(t, y) dy = 1$. This lets the term penalizing the mass of m_n disappear in $F_n(m_n)$ and, using $m \geq 0$ and $m_n \geq 0$, the penalization of the negative part disappears as well:

$$F_n(m_n) = \int_0^T \int_{\Omega} L_{\varepsilon}(\dot{m}_n(t, x)) + f(m_n(t, x)) + m_n(t, x) V^n(t, x) dx dt + \psi_{0,\alpha}(m_n(0)) + \psi_{T,\alpha}(m_n(T)).$$

We observe that, by Jensen's inequality, exactly as it happens for V^n , the sequence m_n is bounded in L^2 (the norm is bounded by that of m) and clearly weakly converges to m (it is enough to test against continuous test functions). So, it strongly converges to m in L^2 . This is true both in $L^2_{t,x}$ and in L^2_x for every t . In particular, the boundary terms and the linear term $\int m_n V^n$ converge to the corresponding terms with m . As for the term $\int f(m_n)$, it can be bounded (again thanks to Jensen's inequality) by $\int f(m)$. We are left to bound the term involving the time-derivative.

Using again the Jensen inequality we have $\int L_{\varepsilon}(\dot{m}_n) \leq \int L_{\varepsilon}(\dot{m})$ and using $m \in D$ (i.e. $\dot{m} \in L^2_{t,x}$) the quantity in the right-hand side tends to $\|\dot{m}\|_{L^1([0,T] \times \Omega)}$.

Summing up, we have proved that for all $m \in D$ such that $\int_{\Omega} m(t, x) dx = 1$, there exists a sequence $(m_n)_n$ which converges to m strongly in $L^2([0, T] \times \Omega)$ and which verifies

$$\limsup_n F_n(m_n) \leq F(m). \quad (4.34)$$

Now, we can prove that ρ minimizes F . Let $m \in D$ be a competitor such that $\int_{\Omega} m(t, x) dx = 1$ and $(m_n)_n$ the sequence defined in (4.33). Since for all $n \in \mathbb{N}$, ρ_n minimizes F_n , we have

$$\forall n \in \mathbb{N}, F_n(\rho_n) \leq F_n(m_n).$$

By Lemma 4.2.1 and the property (4.34), we obtain

$$\forall m \in D, F(\rho) \leq F(m). \quad (4.35)$$

The inequality is true for all $m \in D$, there remains to show that it is true for all $m \in \mathcal{E}$.

Let $m \in \mathcal{E}$ be such that $\int_{\Omega} m(t, x) dx = 1$. By Lemma 4.2.2, there exists a sequence $(m_n)_n \subset D$ which converges to m strongly in $L^2_{t,x}$ and such that $\int_{\mathbb{R}^N} m_n(t, x) dx = 1$ for all $n \in \mathbb{N}$. This sequence may be different from (4.33). By Inequality (4.35), we have

$$\forall n \in \mathbb{N}, F(\rho) \leq F(m_n).$$

By using the properties (4.9), (4.10) and (4.11) and the fact that m is supported in $[0, T] \times \Omega$, we obtain

$$\forall m \in \mathcal{E}, F(\rho) \leq F(m),$$

which shows that the limit ρ is a minimizer of F .

To conclude, the function ρ is a solution to the problem (4.17) and verifies the property (4.18).

Moreover, the solution ρ is unique by the strict convexity of F , given by f . \square

4.3 Infinite horizon problems

Similar results can be proven in the infinite horizon case, with an exponential discount factor:

$$\min_{\substack{\rho \in \mathcal{E}_\infty; \rho \geq 0 \\ \forall t \in \mathbb{R}_+, \int_\Omega \rho(t, x) dx = 1}} \int_0^\infty e^{-rt} \int_\Omega (|\dot{\rho}(t, x)| + V(t, x)\rho(t, x) + f(\rho(t, x))) dx dt + \psi_0(\rho(0)) := F_\infty(\rho)$$

where $\mathcal{E}_\infty := \text{BV}_{loc}(\mathbb{R}_+; L^1(\Omega)) \cap L^2_{loc}(\mathbb{R}_+ \times \Omega)$. This kind of infinite horizon problem is very classical in economical models and we will not discuss any more its economical motivations. From the mathematical point of view, it is interesting as we get rid of some difficulties (which will appear in the next section) related to the transversality condition at $t = T$ but the computations which provide the time regularity can be re-done up to minor modifications.

We have the following result:

Theorem 4.3.1. *Suppose that $\psi_0 : L^1(\Omega) \rightarrow \mathbb{R}$ is 1-Lipschitz and weakly lower semi-continuous on $L^1(\Omega)$ and that $V : \mathbb{R}_+ \times \Omega \rightarrow \mathbb{R}$ belongs to $\text{Lip}(\mathbb{R}_+; L^2(\Omega))$. Suppose also that $f : \mathbb{R} \rightarrow \mathbb{R}$ is c_0 -convex, i.e $f'' \geq c_0$. Then, there exists a unique solution ρ to the problem*

$$\min_{\substack{\rho \in \mathcal{E}_\infty; \rho \geq 0 \\ \forall t \in \mathbb{R}_+, \int_\Omega \rho(t, x) dx = 1}} F_\infty(\rho),$$

and it satisfies

$$\sup_{t \in \mathbb{R}_+} \int_\Omega |\dot{\rho}(t, x)|^2 dx \leq C_\infty, \quad (4.36)$$

where $C_\infty = \frac{C_0^2}{c_0^2} > 0$ only depends on V with $C_0^2 = \sup_{t \in \mathbb{R}_+} \|V'(t, \cdot)\|_{L^2(\Omega)}^2$ and on c_0 .

Proof. The proof will be very similar to the case discussed in Section 4.2, and we will only highlight the differences. First, there is an extra approximation due to the infinite horizon: we fix T (and, later, we will consider $T \rightarrow \infty$) and we define the following approximated functional

$$F_T(\rho) := \int_0^T e^{-rt} \int_\Omega (|\dot{\rho}(t, x)| + V(t, x)\rho(t, x) + f(\rho(t, x))) dx dt + \psi_0(\rho(0))$$

and its approximation

$$F_T^n(\rho) := \int_0^T e^{-rt} \left(\sum_{i=1}^n \frac{1}{n} \left(L_\varepsilon(n\rho_i(t)) + f_n(n\rho_i(t)) + \frac{(\sum_j \rho_j(t) - 1)^2}{2\delta} \right) + n\rho_i(t) \int_{A_i^n} V(t, x) dx \right) dt + \psi_{0, \alpha}(\rho(0)),$$

where L_ε and the partition $\Omega = \cup_{i=1}^n A_i^n$ are the same as in the proof of Theorem 4.2.4.

We solve the problem

$$\min_{\rho \in \mathcal{E}_n} F_T^n(\rho). \quad (4.37)$$

We adapt the proof of Theorem 4.2.4 and we mention here the main differences.

STEP 1 Using the fact that $e^{-rT} \leq e^{-rt}$, we can conclude similarly to Step 1 in Theorem 4.2.4 that there exists a solution to problem (4.37).

STEP 2 Let ρ_n (denoted ρ in this step for simplicity) be a minimizer of F_T^n . We exploit the Euler-Lagrange equation for each component $\rho_i(t)$ in order to obtain bounds. The equation is slightly different now, due to the coefficient e^{-rt} . We have

$$\begin{aligned} (e^{-rt} L'_\varepsilon(\dot{\rho}_i))' &= e^{-rt} \left(V_i^n(t, x_i) + f'_n(n\rho_i) + \frac{\sum_j \rho_j - 1}{\delta} \right) \\ \text{i.e. } -rL'_\varepsilon(\dot{\rho}_i) + (L'_\varepsilon(\dot{\rho}_i))' &= V_i^n(t, x_i) + f'_n(n\rho_i) + \frac{\sum_j \rho_j - 1}{\delta n}. \end{aligned} \quad (4.38)$$

Similarly to the proof of Theorem 4.2.4, we differentiate Eq. (4.38), multiply it by $\dot{\rho}_i$ and sum over i :

$$\sum_i (-r\ddot{\rho}_i\dot{\rho}_i L'_\varepsilon(\dot{\rho}_i) + \dot{\rho}_i(L'_\varepsilon(\dot{\rho}_i))'') = \sum_i (\dot{\rho}_i(V_i^n)'(t, x_i) + n\dot{\rho}_i^2 f''_n(n\rho_i)) + \frac{(\sum_i \dot{\rho}_i)^2}{\delta n}.$$

Compared to Section 4.2, there is an extra term equal to $\sum_i (-r\ddot{\rho}_i\dot{\rho}_i L'_\varepsilon(\dot{\rho}_i))$. We now take again h the function such that

$$h(s) = sL'_\varepsilon(s) - L_\varepsilon(s)$$

and look at $\max_{t \in [0, T]} \sum_{i=1}^n h(n\dot{\rho}_i(t))$.

The extra term in the Euler-Lagrange equation appears in considering the maximum in $(0, T)$, but in this case we also have

$$\sum_{i=1}^n h'(n\dot{\rho}_i(t))\ddot{\rho}_i(t) = 0.$$

Using $h'(s) = sL''_\varepsilon(s)$ we see that this lets the extra term vanish, and the computations are then unchanged.

As for the case where the maximum is reached at $t = 0$, the transversality condition is exactly the same as in Section 4.2 because of $e^{-rt} = 1$. When the maximum is reached at $t = T$, the transversality condition now gives $e^{-rT} L'_\varepsilon(n\dot{\rho}_i(T)) = 0$ for every i . This allows to obtain $\dot{\rho}_i(T) = 0$ and the maximum cannot be reached at $t = T$.

This allows to bound $\sum_i n\dot{\rho}_i(t)^2$ by a constant $C_\infty := C_0^2/c_0^2$ which only depends on c_0 and the quantity $\sup_{t \in [0, T]} \|V'(t, \cdot)\|_{L^2(\Omega)}^2 \leq \sup_{t \in \mathbb{R}_+} \|V'(t, \cdot)\|_{L^2(\Omega)}^2 := C_0^2$.

STEP 3 Now, we pass to the limit in n for fixed T . This follows the very same procedure based on the Arzelà-Ascoli's theorem, as in Section 4.2. In this way, we obtain a family $(\rho_T)_T$ of equilipschitz functions in t valued in $L^2(\Omega)$, satisfying the very same uniform bound on the time derivative, and each ρ_T minimizes F_T by proceeding the same way as in Step 4 of Theorem 4.2.4.

STEP 4 We pass to the limit $T \rightarrow \infty$. The previous step provides a family $(\rho_T)_{T \in \mathbb{N}}$ of equilipschitz minimizers of F_T . By choosing a density m such that $\int f(m) < +\infty$ and comparing ρ_T to the constant curve equal to m , using the integrability of the exponential discount coefficient e^{-rt} , we obtain a uniform bound $F_T(\rho_T) \leq C$ (independent of T). Let us fix $T_0 < \infty$. Since ρ_T is a minimizer of F_T , we have $F_{T_0}(\rho_T) \leq F_T(\rho_T) \leq C$. Thus we have a bound (depending on T_0 , because of the coefficient e^{-rt})

$$\sup_{t \in [0, T_0]} \|\rho_T(t)\|_{L^2(\Omega)} \leq C_{T_0}.$$

This allows to apply Arzelà-Ascoli's theorem for the convergence in the sense of (4.8), and extracting a diagonal subsequence we obtain a subsequence of $(\rho_T)_T$ which converges towards $\rho \in C^0(\mathbb{R}_+, L^2(\Omega))$ in the sense of (4.8), on each interval $[0, T_0]$.

STEP 5 In this step, we show that ρ defined previously is a minimizer of F . Let us denote by $(\rho_{T_k})_k$, the subsequence which converges to ρ . Let $m \in \mathcal{E}_\infty$ be a competitor and T a fixed value. For all $T_k \geq T$, we have

$$F_T(\rho_{T_k}) \leq F_T(m) \leq F_\infty(m).$$

By the lower semi-continuity of F_T we obtain $F_T(\rho) \leq F_\infty(m)$ and, taking the limit $T \rightarrow \infty$, we see $F_\infty(\rho) \leq F_\infty(m)$.

This shows that ρ is a minimizer of F . Additionally, it verifies

$$\forall t, s \in \mathbb{R}_+, \|\rho(t) - \rho(s)\|_{L^2(\Omega)} \leq C_\infty |t - s|,$$

hence (4.36). □

4.4 Regularity in space

While the regularity shown in the previous sections is in time, we can show regularity in space under additional conditions on V and on the boundary conditions. We will start from the problem with Dirichlet boundary conditions:

$$\min_{\substack{\rho \in \mathcal{E} \\ \rho(0, \cdot) = m_0, \rho(T, \cdot) = m_T \\ \forall t \in [0, T], \int_\Omega \rho(t, x) dx = 1}} \int_0^T \int_\Omega (|\dot{\rho}(t, x)| + V(t, x)\rho(t, x) + f(\rho(t, x))) dx dt := F(\rho), \quad (4.39)$$

where we remind that $\mathcal{E} := \text{BV}([0, T]; L^1(\Omega)) \cap L^2([0, T] \times \Omega)$. Similarly to Section 4.2 and 4.3, we approximate the problem (4.39) by

$$\min_{\substack{\rho \in \mathcal{E}_n \\ \rho(0, \cdot) = m_0^n, \\ \rho(T, \cdot) = m_T^n}} \int_0^T \left(\left(\sum_{i=1}^n \frac{1}{n} L_\varepsilon(n\rho_i(t)) + n\rho_i(t) \int_{A_i^n} V(t, x) dx + \frac{1}{n} f(n\rho_i(t)) \right) + \frac{(\sum_j \rho_j(t) - 1)^2}{2\delta} \right) dt := F_n(\rho), \quad (4.40)$$

with the usual choices for $L_\varepsilon(s) = \sqrt{s^2 + \varepsilon^2} + \varepsilon s^2$ and \mathcal{E}_n , the set of piecewise constant functions $\rho \in \mathcal{E}$ such that

$$\forall t \in [0, T], \forall x \in \Omega, \rho(t, x) = \sum_{i=1}^n n\rho_i(t) \mathbb{1}_{A_i^n}(x),$$

where $\rho_i: [0, T] \rightarrow \mathbb{R}$ is a real-valued function. Note that here we prefer not to replace the Dirichlet boundary conditions with L^1 penalizations, so that we need to discretize the initial and final data as well. We then define m_t^n to be piecewise constant approximations of m_t (for $t = 0, T$), taking in particular $m_{i,t}^n := \frac{1}{|A_i^n|} \int_{A_i^n} m_t(x) dx$. We also set, as usual, $V_i^n(t) := \frac{1}{|A_i^n|} \int_{A_i^n} V(t, x) dx$.

We will consider the problem solved by $\rho_i(t)$ for all $i \in \{1, \dots, n\}$:

$$\min_{\substack{n\rho_i(0) = m_{i,0}^n, \\ n\rho_i(T) = m_{i,T}^n}} \int_0^T \left(L_\varepsilon(n\rho_i(t)) + f(n\rho_i(t)) + n\rho_i(t) V_i^n + \frac{(\sum_j \rho_j(t) - 1)^2}{2\delta} \right) dt. \quad (4.41)$$

The Euler-Lagrange system of (4.41) is

$$\begin{cases} (L'_\varepsilon(n\rho_i(t)))' &= V_i^n(t, x_i) + f'(n\rho_i(t)) + c(t), \\ n\rho_i(0) &= m_{i,0}^n, \\ n\rho_i(T) &= m_{i,T}^n, \end{cases} \quad (4.42)$$

where $c(t) := \frac{\int \rho(t, y) dy - 1}{\delta}$.

Lemma 4.4.1. *If ρ is a minimizer of (4.40), then for every i, j we have the following inequality:*

$$n \sup_{t \in [0, T]} \rho_i(t) - \rho_j(t) \leq \max \left(\sup_{t \in [0, T]} V_j^n(t) - V_i^n(t), m_{i,0}^n - m_{j,0}^n, m_{i,T}^n - m_{j,T}^n \right).$$

Proof. We consider

$$\max_{t \in [0, T]} \rho_i(t) - \rho_j(t).$$

We distinguish three cases:

- the maximum is reached on $(0, T)$,
- the maximum is reached at 0
- the maximum is reached at T .

(i) If the maximum is attained at $t_0 \in (0, T)$, we have $\dot{\rho}_i(t_0) = \dot{\rho}_j(t_0)$ as well as $\ddot{\rho}_i(t_0) \leq \ddot{\rho}_j(t_0)$. This implies $\ddot{\rho}_i(t_0) L''_\varepsilon(n\rho_i(t_0)) \leq \ddot{\rho}_j(t_0) L''_\varepsilon(n\rho_j(t_0))$. By using (4.42) and subtracting the equation for i and that for j , we have

$$\begin{aligned} n\rho_i(t_0) - n\rho_j(t_0) &\leq \frac{1}{c_0} (f'_n(n\rho_i(t_0)) - f'_n(n\rho_j(t_0))) \\ &= \ddot{\rho}_i(t_0) L''_\varepsilon(n\rho_i(t_0)) - \ddot{\rho}_j(t_0) L''_\varepsilon(n\rho_j(t_0)) + V_j^n(t_0) - V_i^n(t_0) \\ &\leq V_j^n(t_0) - V_i^n(t_0) \leq \sup_{t \in [0, T]} V_j^n(t) - V_i^n(t). \end{aligned}$$

Consequently, $n(\rho_i(t_0) - \rho_j(t_0))$ is bounded by $\sup_{t \in [0, T]} V_j^n(t) - V_i^n(t)$.

(ii) If the maximum is reached at $t = 0$ we have

$$n\rho_i(t) - n\rho_j(t) \leq m_{i,0}^n - m_{j,0}^n$$

using the Dirichlet condition.

(iii) If the maximum is reached at $t = T$ we have

$$n\rho_i(t) - n\rho_j(t) \leq m_{i,T}^n - m_{j,T}^n$$

using the other Dirichlet condition.

The claim follows by putting together the three cases. \square

Theorem 4.4.2. *Let ρ be a minimizer of (4.39) and ω a function such that*

$$\frac{1}{c_0}(V(t, x) - V(t, x')), m_s(x) - m_s(x') \leq \omega(x - x')$$

for all t, x, x' with $s = 0, T$. Suppose either that ω is a constant or that $\lim_{z \rightarrow 0} \omega(z) = 0$ (i.e., ω is a modulus of continuity).

Then we have for a.e. t, x, x'

$$\rho(t, x) - \rho(t, x') \leq \omega(x - x').$$

Proof. We first consider the curves ρ_n minimizing (4.40). We write the estimate from Lemma 4.4.1 in terms of the densities $\rho_n(t, x) = n\rho_i(t)$. We obtain

$$\rho_n(t, x) - \rho_n(t, x') \leq \omega(x - x') + \varepsilon_n.$$

Indeed, if the function ω is a constant M , then the oscillations of $\frac{V}{c_0}, m_0$ and m_T are bounded by M and so $n\rho_i - n\rho_j \leq M$ as well. Otherwise, if ω is a modulus of continuity, then the functions $\frac{V}{c_0}, m_0$ and m_T are continuous and their averages on small pieces A_i^n can be replaced by the values at the center of these pieces up to a small error ε_n (which also takes into account that x and x' can differ from the centers of the corresponding pieces).

The family $(\rho_n)_n$ is bounded in $L^2([0, T] \times \Omega)$. Differently from the case of Section 4.2, we do not have Lipschitz estimates in time (note that this is not the same approximation as in Section 4.2, since we impose the Dirichlet boundary conditions instead of penalizing them), but luckily we will not need them. Indeed, the equicontinuity in time was essential to obtain uniform and pointwise bounds and deal with the boundary terms. Here we just use weak convergence in $L^2([-1, T+1] \times \Omega)$ after extending ρ_n to m_0 on $[-1, 0]$ and to m_T on $[T, T+1]$. It is easy to see that ρ_n admits a weakly converging subsequence and that the limit solves (4.39). Note that the extension before $t = 0$ and after $t = T$ is needed to include the possible jump at those instants of time in the total variation and hence play the role of the Dirichlet boundary condition.

The conclusion comes from the following Lemma 4.4.3. \square

Lemma 4.4.3. *Let ρ_n be a sequence weakly converging to ρ in $L^2([0, T] \times \Omega)$ and suppose that there exists a function ω such that*

$$\forall n \in \mathbb{N}^*, \forall t \in [0, T], \forall x, x' \in \Omega, \rho_n(t, x) - \rho_n(t, x') \leq \omega(x - x') + \varepsilon_n$$

for a sequence $\varepsilon_n \rightarrow 0$.

Then we have

$$\rho(t, x) - \rho(t, x') \leq \omega(x - x')$$

whenever (t, x) and (t, x') are Lebesgue points of ρ .

Proof. Let us take (t_0, x_0) and (t_0, y_0) in $[0, T] \times \Omega$, two Lebesgue points of ρ . Let $r > 0$ be such that $]t_0 - r, t_0 + r[\subset [0, T]$, $B(x_0, r) \subset \Omega$ and $B(y_0, r) \subset \Omega$, we have in particular

$$\int_{t_0-r}^{t_0+r} \int_{B(x_0, r)} \rho_n(t, x) dx dt \xrightarrow{n \rightarrow \infty} \int_{t_0-r}^{t_0+r} \int_{B(x_0, r)} \rho(t, x) dx dt, \quad (4.43)$$

$$\text{and } \int_{t_0-r}^{t_0+r} \int_{B(y_0, r)} \rho_n(t, x) dx dt \xrightarrow{n \rightarrow \infty} \int_{t_0-r}^{t_0+r} \int_{B(y_0, r)} \rho(t, x) dx dt. \quad (4.44)$$

By assumption we have the inequality

$$\begin{aligned} \int_{t_0-r}^{t_0+r} \left(\int_{B(x_0, r)} \rho_n(t, x) dx - \int_{B(y_0, r)} \rho_n(t, x') dx' \right) dt &= \int_{t_0-r}^{t_0+r} \int_{B(x_0, r)} (\rho_n(t, x) - \rho_n(t, x - x_0 + y_0)) dx dt \\ &\leq \omega(x_0 - y_0) + \varepsilon_n. \end{aligned}$$

By taking the limit $n \rightarrow \infty$ and applying (4.43) and (4.44) to the left-hand side of the inequality, we get

$$\int_{t_0-r}^{t_0+r} \left(\int_{B(x_0,r)} \rho(t,x) dx - \int_{B(y_0,r)} \rho(t,x') dx' \right) dt \leq \omega(x_0 - y_0).$$

Since (t_0, x_0) and (t_0, y_0) are Lebesgue points, we can pass to the limit $r \rightarrow 0$ and obtain

$$\rho(t_0, x_0) - \rho(t_0, y_0) \leq \omega(x_0 - y_0).$$

□

The case where the Dirichlet conditions are replaced by penalizations are harder to deal with. The only case that is easy to consider requires that the transversality condition is the same for ρ_i and ρ_j . We can obtain the following results for which we just sketch the modifications to the previous proofs.

Theorem 4.4.4. *Let ρ be a minimizer of*

$$\min_{\substack{\rho \in \mathcal{E} \\ \rho(0,\cdot) = m_0, \\ \forall t \in [0,T], \int_{\Omega} \rho(t,x) dx = 1}} \int_0^T \int_{\Omega} (|\dot{\rho}(t,x)| + V(t,x)\rho(t,x) + f(\rho(t,x))) dx dt + \int \Psi_T \rho(T,x).$$

Let $A \subset \Omega$ be a set where Ψ_T is constant. Suppose that ω is a function (either constant or a modulus of continuity) such that $\frac{1}{c_0}(V(t,x) - V(t,x')), m_0(x) - m_0(x') \leq \omega(x - x')$ for all t, x, x' . Then we have for a.e. $x, x' \in A$

$$\rho(t,x) - \rho(t,x') \leq \omega(x - x').$$

Proof. The approximation will be the same as before, but the Dirichlet boundary condition at $t = T$ is replaced by a transversality condition. Choosing a decomposition into pieces A_i^t such that x, x' belong to two pieces contained in A , this transversality condition will be the same for the two curves $\rho_i(t)$ and $\rho_j(t)$ that we need to consider to estimate $\rho(t,x) - \rho(t,x')$. In particular, we will have $\dot{\rho}_i(T) = \dot{\rho}_j(T)$. Hence, when considering $\max_i \rho_i(t) - \rho_j(t)$, the maximum could be attained on $t_0 = T$ but in this case the first-order optimality condition will be satisfied, and this allows to obtain the second-order one, which is the main tool to estimate $\rho_i(t_0) - \rho_j(t_0)$. The rest of the analysis goes as in the rest of the section. □

Theorem 4.4.5. *Let ρ be a minimizer of*

$$\min_{\substack{\rho \in \mathcal{E}_{\infty} \\ \rho(0,\cdot) = m_0, \\ \forall t \in \mathbb{R}_+, \int_{\Omega} \rho(t,x) dx = 1}} \int_0^{\infty} e^{-rt} \int_{\Omega} (|\dot{\rho}(t,x)| + V(t,x)\rho(t,x) + f(\rho(t,x))) dx dt := F(\rho).$$

Suppose that ω is a function (either constant or a modulus of continuity) such that for all t, x, x'

$$\frac{1}{c_0}(V(t,x) - V(t,x')), m_0(x) - m_0(x') \leq \omega(x - x').$$

Then we have for a.e. x, x'

$$\rho(t,x) - \rho(t,x') \leq \omega(x - x').$$

Proof. We first replace F with F_T as in Section 4.3. Then, the claim is the same as in Theorem 4.4.4 with $\Psi_T = 0$ (hence we can take $A = \Omega$) with the only difference that have an extra coefficient e^{-rt} . This lets an extra term $-rL'_{\varepsilon}(n\dot{\rho}_i)$ appear in the Euler-Lagrangian equation, but this term cancels when taking the difference between i and j because of the first-order optimality condition for t_0 . The rest of the analysis goes as in Theorem 4.4.4 and in the rest of the section. This provides a uniform estimate on ρ_T , independent of T , and we can then take the limit $T \rightarrow \infty$. □

We conclude summarizing the result that we can obtain in terms of spatial regularity:

- For the problem with two Dirichlet boundary conditions, if V, m_0 , and m_T are continuous, then the solution ρ shares the same modulus of continuity of $\frac{V}{c_0}, m_0$, and m_T . Combining this with the L^2 Lipschitz regularity in time obtained in Section 4.2 this gives a uniform continuity result in (t, x) .
- In the same problem, if V, m_0 , and m_T are only bounded, then ρ is bounded, since its oscillation is bounded by that of $\frac{V}{c_0}, m_0$, and m_T and its L^2 norm is also bounded.

- For the problem with a Dirichlet boundary condition at $t = 0$ and a penalization Ψ_T at $t = T$, if Ψ_T is piecewise constant and V and m_0 are continuous, then ρ is piecewise continuous, and hence bounded. The solution ρ is also bounded if we only assume V and m_0 to be bounded, and Ψ_T to be piecewise constant.
- In the infinite-horizon problem with Dirichlet boundary condition at $t = 0$, the solution shares the same modulus of continuity of $\frac{V}{c_0}$ and m_0 and is uniformly continuous in (t, x) . It is uniformly bounded if V and m_0 are bounded.
- In the periodic case (which we briefly presented in Section 4.1 but did not develop here) the solution ρ shares the same continuity or boundedness of $\frac{V}{c_0}$.

All the above continuity results are crucial for applications to the theory of Mean Field Games, since the variational problem we studied, with a congestion penalization in the form of $f(\rho)$, corresponds to a MFG where every agent minimizes a cost over trajectories γ involving the integral of a running cost of the form $\int_0^T (V + f'(\rho))(t, \gamma(t)) dt$. This cost is not even well-defined if $\rho(t, x)$ is not a continuous function of x ! When ρ is not continuous but it is bounded, a clever construction due to Ambrosio and Figalli ([AF09]) and re-used in the frameworks of MFG in [CMS16] allows to choose a particular representative of this running cost for which it is possible to prove the desired equilibrium results. This explains our interest for the spatial regularity and/or the boundedness of ρ .

Chapter 5

Numerical simulations

In this chapter numerical simulations are carried out on the following problem

$$\min_{\substack{\rho \in \mathcal{E}, \rho \geq 0 \\ \forall t \in [0, T], \int_{\Omega} \rho(t, x) dx = 1}} \mathcal{V}(\rho) := \int_0^T \int_{\Omega} \left(\lambda |\dot{\rho}(t, x)| + V(t, x) \rho(t, x) + \frac{\rho(t, x)^2}{2} \right) dx dt + \psi_0(\rho(0)) + \psi_T(\rho(T)) \quad (5.1)$$

and on some of its variants (Dirichlet conditions, periodic case...). The set \mathcal{E} is $\text{BV}([0, T]; L^1(\Omega)) \cap L^2([0, T] \times \Omega)$. We study a particular case of \mathcal{V} where $f(\rho) = \rho^2/2$. The parameter $\lambda > 0$ allows us to add more or less importance to the L^1 -norm.

In 2009, Beck and Teboulle [BT09] introduced an algorithm which approximates the solution to a minimization problem of the form

$$\min_x f(x) + g(x),$$

where f is convex and sufficiently smooth and g is convex and can be less regular. The algorithm is called the *Proximal Gradient Method* and is also known as the *Iterative Shrinkage-Thresholding Algorithm* (ISTA). At each step, the method consists in composing the proximal operator (introduced by Moreau in 1965 [Mor65]) of g with the step of the gradient method in f . Beck and Teboulle showed that the sequence of the values $(f(x_n) + g(x_n))_n$ converges towards the optimal value at a speed of order $O(1/n)$. By using an acceleration due to Nesterov [Nes83], the speed of the convergence can be improved to the order $O(1/n^2)$. The latter method is called the *Fast Iterative Shrinkage-Thresholding Algorithm* (FISTA).

To compute numerically the solution of our problem, we will use the *Fast Dual Proximal Gradient Method*, also introduced by Beck and Teboulle in 2014 [BT14], which consists in approximating the solution to a problem of the form

$$\min_x f(x) + g(\mathcal{A}x),$$

where f is σ -strongly convex for some $\sigma > 0$, g is convex and \mathcal{A} is a linear operator. The discretization of the term $\dot{\rho}(t, x)$ involves a sparse matrix and the constraints $\rho \geq 0$ and $\int \rho(t) = 1$ have to be taken into account. This is why the formulation with the linear operator is very useful. In addition, Beck and Teboulle proved that the sequence $(x_n)_n$ converges towards the minimizer at a speed of $O(1/n)$, which is slower than FISTA but occurs directly on the approximating sequence $(x_n)_n$ and not on the values of the function.

The first section is an introduction to non-smooth optimization algorithms which will lead to the algorithm that we use.

For these numerical examples, we take the domain $\Omega := [0, S]$ to be one-dimensional with $S > 0$. In Section 5.2, we study the following cases:

- periodic solutions 1D in time (only the time variable, Section 5.2.1, Figure 5.1).
- periodic solutions 1D in time and 1D in space (Section 5.2.2, Figures 5.2, 5.3).
- non-periodic solutions 1D in time and 1D in space, with or without Dirichlet conditions or penalizations at the time boundary (Section 5.2.3, Figures 5.4, 5.5, and 5.6).

In Section 5.3, we study the case where the solution $\bar{\rho}$ is the fixed point of $\bar{\rho} \mapsto \arg\min_{\rho} \mathcal{V}_{\bar{\rho}}(\rho)$ described in Section 2.2. We consider three different examples:

(i) $V(\rho)(t, x) = \left(\int_{x-\delta}^{x+\delta} \rho(t, y) dy \right)^2 + V_0(t, x)$ (Figure 5.7).

(ii) $V(\rho)(t, x) = \min \left(\int_0^x \arctan(\rho(t, y)) dy, \int_x^S \arctan(\rho(t, y)) dy \right) + V_0(t, x)$ (Figure 5.8).

(iii) $V(\rho)(t, x) = \int_0^t \rho(s, x) ds + V_0(t, x)$ (Figures 5.9).

Each of these examples is *non-variational*, i.e. we cannot find a functional for which they are the first variation. The example (i) corresponds to a problem where each player considers the running cost depending on the density in his neighborhood. The example (ii) describes an escape problem where agents would like to exit the domain either from one side or from the other. In the last example (iii), the players keep in memory the past experience.

5.1 Introduction to non-smooth optimization algorithms

To carry out the numerical simulations, we used the Fast Dual Proximal Gradient method whose presentation can be found in [Bec17, Chapter 12]. This section is a small introduction to this method which is inspired by [Bec17, Chapter 10]. The following methods are usually called *first-order methods*, since the gradient of the function is involved.

5.1.1 The Proximal Gradient method

To explain the idea of the Proximal Gradient method, one can start with the Gradient Descent method. To construct a sequence that approximates the solution to the problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

where f is differentiable and convex, the iteration at each step is

$$x^{k+1} = x^k - t_k \nabla f(x^k).$$

The real number t_k is given according to the type of algorithm we choose. For example, it can either be constant for every k , or it can be computed in optimal way so that t_k minimizes $t \mapsto f(x^k - t \nabla f(x^k))$. The latter is called the *Optimal Step Gradient method* and it is known that this method converges when f is \mathcal{C}^1 and α -convex (see for example [Cia07, Theorem 8.4-3]).

Why is the direction $-\nabla f(x^k)$ the best we can choose?

If $v \in \mathbb{R}^n$ such that $\|v\| = 1$ is a direction which makes f diminish (i.e. $f(x^k + v) \leq f(x^k)$), we have that

$$\langle \nabla f(x^k) | v \rangle = \lim_{t \rightarrow 0^+} \frac{f(x^k + tv) - f(x^k)}{t} \leq 0. \quad (5.2)$$

Then, by the Cauchy-Schwarz inequality we have

$$-\|\nabla f(x^k)\| \|v\| \leq \langle \nabla f(x^k) | v \rangle$$

whose equality is reached when v and $\nabla f(x^k)$ are colinear. It is better to choose the opposite direction of $\nabla f(x^k)$, i.e. $v = -\nabla f(x^k) / \|\nabla f(x^k)\|$, so that the scalar product (5.2) is as negative as possible. A more obvious way to see is when the scalar product is written as

$$\langle u | w \rangle = \|u\| \cos(\widehat{u, w}), \forall u, w \in \mathbb{R}^n \text{ and } \|w\| = 1.$$

Indeed, the angle that minimizes $\langle u | w \rangle$ is π and it corresponds to w being in the opposite direction of u .

Remark 5.1.1. *A little curiosity is that to reach the minimum in a finite number of steps, the direction $-\nabla f(x^k)$ may not always be optimal. For example, if $f(x, y) = \frac{1}{2}(\alpha x^2 + \beta y^2)$ with $0 < \alpha < \beta$, then the minimum point is $(0, 0)$. We seek for an iteration $k + 1$ such that $(x^{k+1}, y^{k+1}) = (0, 0)$. In particular, there would exist t verifying*

$$\begin{cases} x^k - \alpha t x^k & = 0 \\ y^k - \beta t y^k & = 0. \end{cases}$$

Since α and β are different, this implies that one of the coordinates (x^k, y^k) is null. In addition, if we apply the *Optimal Step Gradient Descent method*, we would have $t^k = \frac{\alpha^2(x^k)^2 + \beta^2(y^k)^2}{\alpha^3(x^k)^2 + \beta^3(y^k)^2}$, so the next step would be

$$x^{k+1} = x^k - t^k \alpha x^k = \frac{\beta^2(\beta - \alpha)x^k(y^k)^2}{\alpha^3(x^k)^2 + \beta^3(y^k)^2} \text{ and } y^{k+1} = y^k - t^k \beta y^k = \frac{\alpha^2(\alpha - \beta)y^k(x^k)^2}{\alpha^3(x^k)^2 + \beta^3(y^k)^2}.$$

Therefore, if $x^0 \neq 0$ and $y^0 \neq 0$, then $x^k \neq 0$ and $y^k \neq 0$ for all k . Consequently, the *Optimal Step Gradient Descent method* cannot reach the minimum.

This motivates the use of the *Conjugate Gradient method* which, at step k , consists in going in the direction that is a linear combination of $\nabla f(x^k)$, $\nabla f(x^{k-1})$, \dots , $\nabla f(x^0)$, which should be a family of two by two orthogonal vectors. Since we are in a space of finite dimension, the algorithm should end at the minimum point. This example is presented in [Cia07, Section 8.5].

If we restrict the problem to a nonempty closed and convex set $C \subset \mathbb{R}^n$ such that

$$\min_{x \in C} f(x), \quad (5.3)$$

then at each step of the algorithm one just needs to project the iteration onto the set C :

$$x^{k+1} = P_C(x^k - t_k \nabla f(x^k)),$$

where P_C is the projection on the set C . In this way, the iteration x^{k+1} can be written as

$$x^{k+1} = \operatorname{argmin}_{x \in C} \|x - (x^k - t^k \nabla f(x^k))\|^2.$$

Since t^k is positive, x^{k+1} is also a minimizer of $\frac{1}{2t^k} \|x - (x^k - t^k \nabla f(x^k))\|^2$ in C . By developing the latter, we obtain

$$\frac{1}{2t^k} \|x - (x^k - t^k \nabla f(x^k))\|^2 = \frac{1}{2t^k} \|x - x^k\|^2 + \langle x - x^k | \nabla f(x^k) \rangle + \frac{1}{2} t^k \|\nabla f(x^k)\|^2.$$

The last term in the sum does not depend on x , so it can be cancelled in the minimization problem and one can add $f(x^k)$ such that the iteration x^{k+1} becomes

$$x^{k+1} = \operatorname{argmin}_{x \in C} \left\{ f(x^k) + \langle x - x^k | \nabla f(x^k) \rangle + \frac{1}{2t^k} \|x - x^k\|^2 \right\}.$$

In this form, x^{k+1} is said to be a minimizer of the linearisation of f in the previous iteration plus a proximity term that prevents the next iteration to be too far from the previous one.

The problem (5.3) can be rewritten as

$$\min_{x \in \mathbb{R}^n} f(x) + g(x)$$

where $g(x) = \delta_C(x)$ is the *indicator function* such that

$$\delta_C(x) = \begin{cases} 0, & \text{if } x \in C, \\ +\infty & \text{otherwise} \end{cases} \quad (5.4)$$

and the iteration term becomes

$$x^{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} \left\{ f(x^k) + \langle x - x^k | \nabla f(x^k) \rangle + g(x) + \frac{1}{2t^k} \|x - x^k\|^2 \right\}.$$

By adding and cancelling terms that does not depend on x and multiplying by t^k , we have

$$x^{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} \left\{ t^k g(x) + \frac{1}{2} \|x - (x^k - t^k \nabla f(x^k))\|^2 \right\}.$$

By the definition of the *proximal operator* which is

$$\operatorname{prox}_f(x) := \operatorname{argmin}_{u \in \mathbb{R}^n} \left\{ f(u) + \frac{1}{2} \|u - x\|^2 \right\},$$

we obtain that

$$x^{k+1} = \operatorname{prox}_{t^k g}(x^k - t^k \nabla f(x^k)). \quad (5.5)$$

The proximal operator was introduced by Jean-Jacques Moreau in 1965 and its full presentation can be found in [Mor65].

The iteration (5.5) motivates the expression of the algorithm for more general g in the problem

$$\min_{x \in \mathbb{R}^n} F(x) := f(x) + g(x) \quad (5.6)$$

where f and g satisfy the following assumptions:

- $f: \mathbb{R}^n \rightarrow (-\infty, \infty]$ is proper, closed and L_f -smooth and f is defined on $\operatorname{dom}(f)$ which is convex,
- $g: \mathbb{R}^n \rightarrow (-\infty, \infty]$ is proper, closed and convex.

Let us recall the definitions below.

Definition 5.1.2. A function $f: \mathbb{R}^n \rightarrow (-\infty, \infty]$ is proper whether it does not attain $-\infty$ and there exists x such that $f(x) < \infty$.

Definition 5.1.3. A function $f: \mathbb{R}^n \rightarrow (-\infty, \infty]$ is closed if its sublevel sets $\{x; f(x) \leq \alpha\}$ are closed.

Definition 5.1.4. A function $f: \mathbb{R}^n \rightarrow (-\infty, \infty]$ is L_f -smooth over a set $D \subset \mathbb{R}^n$, if it is differentiable over D and for all $x, y \in D$

$$\|\nabla f(x) - \nabla f(y)\| \leq L_f \|x - y\|.$$

One may notice that the convexity hypothesis on g allows the proximal operator in (5.5) to be well-defined. Indeed, the minimum exists and is unique by the strict convexity of the quadratic term.

The step size t^k is usually written as $t^k = \frac{1}{L^k}$ in the *Proximal Gradient method* described below.

The Proximal Gradient Method

Initialisation: $x^0 \in \text{int}(\text{dom}(f))$.

Step for $k \geq 0$:

- take $L^k \geq 0$,
- $x^{k+1} = \text{prox}_{\frac{1}{L^k}g}(x^k - \frac{1}{L^k}\nabla f(x^k))$

In the particular case where $f(x) = \|Ax - b\|^2$ and $g(x) = \lambda\|x\|_1$, this method is known as the *Iterative Shrinkage-Thresholding Algorithm* (ISTA) which is often used in image denoising. I refer to [BT09] and the references therein for an overview of the literature on this algorithm and its applications. By abuse of language, the appellation ISTA describes the Gradient Proximal method for general f and g .

Concerning the rate of convergence of the sequence $(x^k)_k$ defined by the Proximal Gradient method, it is not obvious to prove that x^k converges towards an optimal x^* for the problem (5.6). However, we know that the sequence of values $(F(x^k))_k$ converges towards the minimum $F(x^*)$ in a rate of convergence that is $O(1/k)$.

Theorem 5.1.5. Let $(x^k)_k$ be a sequence defined by the Proximal Gradient method and x^* be a minimizer of F in (5.6). Let us take $L^k = L_f$ for all k . Then for all $k \geq 1$,

$$F(x^k) - F(x^*) \leq \frac{L_f \|x^0 - x^*\|^2}{2k}.$$

Proof. See [BT09, Theorem 3.1]. □

5.1.2 The proximal operator

In some cases, the proximal operator of certain functions is not difficult to compute. Let us see in this subsection some easy examples that we will use.

Definition 5.1.6. Let $f: \mathbb{R}^n \rightarrow (-\infty, \infty]$ be a function. The proximal operator of f is defined such that for all $x \in \mathbb{R}^n$,

$$\text{prox}_f(x) = \underset{u \in \mathbb{R}^n}{\text{argmin}} \left\{ f(u) + \frac{1}{2} \|u - x\|^2 \right\}.$$

The mapping $\text{prox}_f(x)$ is a set that might be empty or with at least one vector. The following theorem ensures that under some conditions on f , the set $\text{prox}_f(x)$ is a singleton.

Theorem 5.1.7. Let $f: \mathbb{R}^n \rightarrow (-\infty, \infty]$ be a proper, closed and convex function. Then $\text{prox}_f(x)$ is a singleton for any $x \in \mathbb{R}^n$.

Proof. See [Bec17, Theorem 6.3]. □

Example 5.1.8. (f is constant) Let $c \in \mathbb{R}$. If $f \equiv c$, then prox_f is the identity. For all $x \in \mathbb{R}^n$,

$$\text{prox}_f(x) = \underset{u \in \mathbb{R}^n}{\text{argmin}} \left\{ c + \frac{1}{2} \|u - x\|^2 \right\} = x.$$

Example 5.1.9. (*f is affine*) Let $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. If $f(x) = \langle a|x \rangle + b$, then prox_f is the translation by $-a$. Indeed, for all $x \in \mathbb{R}^n$,

$$\begin{aligned} \text{prox}_f(x) &= \operatorname{argmin}_{u \in \mathbb{R}^n} \left\{ \langle a|u \rangle + b + \frac{1}{2} \|u - x\|^2 \right\} \\ &= \operatorname{argmin}_{u \in \mathbb{R}^n} \left\{ b + \frac{1}{2} \|u - (x - a)\|^2 + \langle a|x \rangle - \frac{1}{2} \|a\|^2 \right\} \\ &= x - a. \end{aligned}$$

Example 5.1.10. (*f is an indicator*) Let C be a nonempty set of \mathbb{R}^n . To apply Theorem 5.1.7, we need to suppose that C is closed and convex, so that $\text{prox}_f(x)$ is a singleton. Then $\text{prox}_f(x) = P_C(x)$. Indeed, the formulation of the proximal operator is exactly the definition of the projection onto a set:

$$\begin{aligned} \text{prox}_f(x) &= \operatorname{argmin}_{u \in \mathbb{R}^n} \left\{ \delta_C(x) + \frac{1}{2} \|u - x\|^2 \right\} \\ &= \operatorname{argmin}_{u \in C} \|u - x\|^2 = P_C(x). \end{aligned}$$

Example 5.1.11. (*f is the absolute value*) Let $\lambda \in \mathbb{R}$. If $x \in \mathbb{R}$ and $f(x) = \lambda|x|$, then

$$\text{prox}_f(x) = [|x| - \lambda]_+ \operatorname{sgn}(x).$$

In particular, if $\lambda > 0$, then the proximal operator is usually denoted $\mathcal{T}_\lambda(x) := \text{prox}_f(x)$ and

$$\mathcal{T}_\lambda(x) = \begin{cases} x - \lambda, & \text{if } x \geq \lambda, \\ 0, & \text{if } |x| < \lambda, \\ x + \lambda, & \text{if } x \leq -\lambda. \end{cases}$$

This function is called the soft thresholding function.

The case $\lambda < 0$ is less used in the literature. The term $|x| - \lambda$ is always positive so the proximal operator is

$$\text{prox}_f(x) = \begin{cases} x - \lambda, & \text{if } x \geq 0, \\ x + \lambda, & \text{if } x < 0. \end{cases}$$

The last example is treated when f is defined on the one-dimensional set \mathbb{R} . Thanks to the next theorem, it is possible to extend it to \mathbb{R}^n when f is the ℓ^1 -norm such that $f(x_1, \dots, x_n) = \lambda(|x_1| + \dots + |x_n|)$. Indeed, the proximal operator of the sum of functions with separate variables is the product of the proximal operators of each of these functions.

Theorem 5.1.12. Let $f: \mathbb{R}^n \rightarrow (-\infty, \infty]$ be such that

$$f(x_1, \dots, x_n) = \sum_{i=1}^n f_i(x_i), \text{ for all } x_i \in \mathbb{R} \text{ and } i = 1, \dots, n.$$

Then

$$\text{prox}_f(x_1, \dots, x_n) = \text{prox}_{f_1}(x_1) \times \dots \times \text{prox}_{f_n}(x_n).$$

Proof. We have

$$\begin{aligned} \text{prox}_f(x_1, \dots, x_n) &= \operatorname{argmin}_{y_1, \dots, y_n} \sum_{i=1}^n \left[f_i(y_i) + \frac{1}{2} \|y_i - x_i\|^2 \right] \\ &= \prod_{i=1}^n \operatorname{argmin}_{y_i} \left[f_i(y_i) + \frac{1}{2} \|y_i - x_i\|^2 \right] \\ &= \prod_{i=1}^n \text{prox}_{f_i}(x_i). \end{aligned}$$

□

Remark 5.1.13. The notation of the product of two sets $A \times B$ is used in Theorem 5.1.12 in a general setting because $\text{prox}_{f_i}(x_i)$ might not be a singleton. Nonetheless, in what follows we deal with closed and convex functions so that Theorem 5.1.7 holds. Therefore, when $A = \{a\}$ and $B = \{b\}$, we shall denote (a, b) .

Example 5.1.14. (*f is the ℓ^1 -norm*) Let $\lambda \in \mathbb{R}$. If $f(x) = \lambda \|x\|_1 = \lambda(|x_1| + \dots + |x_n|)$, then by Theorem 5.1.12, we have

$$\text{prox}_f(x) = (\mathcal{T}_\lambda(x_1), \dots, \mathcal{T}_\lambda(x_n)) := \mathcal{T}_\lambda(x),$$

where we extend the notation of \mathcal{T}_λ to a vector.

5.1.3 FISTA

The convergence rate of the Proximal Gradient method can be improved from $O(1/k)$ to $O(1/k^2)$ by using an acceleration of the gradient due to Nesterov in 1983 [Nes83]. The method by Nesterov consists in accelerating the Gradient Descent method. The new algorithm introduced for the first time by Beck and Teboulle in 2009 [BT09] called *Fast Iterative Shrinkage-Thresholding Algorithm* (FISTA) mixes the Proximal Gradient method and the accelerated method by Nesterov. The idea consists in applying the usual Proximal Gradient step $x^{k+1} = \text{prox}_{\frac{1}{L^k}g}(y^k - \frac{1}{L^k}\nabla f(y^k))$ to the previous iterative y^k and allowing x^{k+1} to go a little further in the direction $x^{k+1} - x^k$. So the new iteration y^{k+1} is a convex combination of the steps x^{k+1} and x^k with a coefficient chosen according to the Nesterov's method [Nes83].

By taking the same assumptions of f and g as in the previous section, the algorithm is as follows:

<p>FISTA</p> <p>Initialisation: $y^0 = x^0 \in \mathbb{R}^n$, $t_0 = 1$.</p> <p>Step for $k \geq 0$:</p> <ul style="list-style-type: none"> • $L^k > 0$, • $x^{k+1} = \text{prox}_{\frac{1}{L^k}g}(y^k - \frac{1}{L^k}\nabla f(y^k))$, • $t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$, • $y^{k+1} = x^{k+1} + \left(\frac{t_k - 1}{t_{k+1}}\right)(x^{k+1} - x^k)$.

The convergence rate of FISTA is stated in the next theorem.

Theorem 5.1.15. *Let $(x^k)_k$ and $(y^k)_k$ the sequences defined by FISTA and x^* a minimal point for F . Let $L^k = L_f$ for all k . Then for all $k \geq 1$,*

$$F(x^k) - F(x^*) \leq \frac{2L_f \|x^0 - x^*\|^2}{(k+1)^2}.$$

Proof. See [BT09, Theorem 4.4]. □

Remark 5.1.16. *In fact, the proof of this theorem only requires the sequence $(t_k)_k$ to verify the following two conditions:*

- (i) $t_k \geq \frac{k+2}{2}$,
- (ii) $t_{k+1}^2 - t_{k+1} \leq t_k^2$.

The sequence chosen in FISTA such that $t_{k+1} = (1 + \sqrt{1 + 4t_k^2})/2$ actually achieves the equality in condition (ii) and it verifies (i) by induction.

The convergence would also occur when $t_k = (k+2)/2$, since it verifies condition (ii):

$$\begin{aligned} t_{k+1}^2 - t_{k+1} - t_k^2 &= \left(\frac{k+3}{2}\right)^2 - \left(\frac{k+3}{2}\right) - \left(\frac{k+2}{2}\right)^2 \\ &= \frac{(k+3)^2 - 2(k+3) - (k+2)^2}{4} \\ &= \frac{k^2 + 6k + 9 - 2k - 6 - k^2 - 4k - 4}{4} = -\frac{1}{4} \leq 0. \end{aligned}$$

More generally, the previous theorem is also true if

$$t_k = \frac{k+a-1}{a} \text{ for any } a \geq 2. \tag{5.7}$$

In addition, Chambolle and Dossal [CD15] showed that the iterates of FISTA $(x^k)_k$ weakly converges toward a minimizer x^ of F when $(t_k)_k$ is defined by (5.7) for $a \geq 2$, while the convergence rate in FISTA was only known for the values $(F(x^k))_k$.*

In practice, we will use the sequence $(t_k)_k$ defined as in (5.7) with $a = 2$.

5.1.4 Dual Proximal Gradient method

The Dual Proximal Gradient method allows to solve another class of problems that are of the form

$$\min_{x \in \mathbb{R}^n} f(x) + g(\mathcal{A}x) \quad (5.8)$$

where

- $f: \mathbb{R}^n \rightarrow (-\infty, \infty]$ is proper, closed and σ -strongly convex,
- $g: \mathbb{R}^m \rightarrow (-\infty, \infty]$ is proper, closed and convex,
- $\mathcal{A}: \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a linear application.

Definition 5.1.17. A function $f: \mathbb{R}^n \rightarrow (-\infty, +\infty]$ is σ -strongly convex for a given $\sigma > 0$ if $f(\cdot) - \frac{\sigma}{2} \|\cdot\|^2$ is convex.

The algorithm to solve this problem has been introduced by Beck and Teboulle [BT14] in 2014. To write the dual problem of (5.8), we set $z = \mathcal{A}x$ and insert it as a constraint in the problem:

$$\min_{\substack{x \in \mathbb{R}^n, z \in \mathbb{R}^m \\ \mathcal{A}x = z}} f(x) + g(z).$$

For $y \in \mathbb{R}^m$, the Lagrangian of this problem is then

$$\begin{aligned} L(x, z; y) &= f(x) + g(z) - \langle y | \mathcal{A}x - z \rangle \\ &= f(x) + g(z) - \langle \mathcal{A}^T y | x \rangle + \langle y | z \rangle. \end{aligned}$$

If we define the *conjugate* of f and g such as respectively

$$f^*(y) = \max_x \{ \langle y | x \rangle - f(x) \} \text{ and } g^*(y) = \max_z \{ \langle y | z \rangle - g(z) \}, \quad (5.9)$$

then L is also the Lagrangian of the problem

$$\max_y \{ q(y) := -f^*(\mathcal{A}^T y) - g^*(-y) \}. \quad (5.10)$$

By the Fenchel-Rockafellar duality theorem [ET99], we obtain that

$$\inf_{\substack{x \in \mathbb{R}^n, z \in \mathbb{R}^m \\ \mathcal{A}x = z}} \{ f(x) + g(z) \} = \max_y q(y),$$

where the maximum is attained.

Now we focus on the dual problem (5.10) on which we shall use FISTA. Let us denote

$$\mathcal{F}(y) := f^*(\mathcal{A}^T y) \text{ and } \mathcal{G}(y) = g^*(-y), \quad (5.11)$$

so that we consider the problem

$$\min_{y \in \mathbb{R}^m} \mathcal{F}(y) + \mathcal{G}(y). \quad (5.12)$$

In order to apply FISTA, we should check whether \mathcal{F} and \mathcal{G} verify the assumptions from Section 5.1.1.

Lemma 5.1.18. Let \mathcal{F} and \mathcal{G} be functions defined by (5.11) with f and g verifying the hypotheses described at the beginning of this subsection. Then

- $\mathcal{F}: \mathbb{R}^m \rightarrow \mathbb{R}$ is convex and $L_{\mathcal{F}}$ -smooth with $L_{\mathcal{F}} = \|\mathcal{A}\|^2/\sigma$,
- $\mathcal{G}: \mathbb{R}^m \rightarrow \mathbb{R}$ is proper, closed and convex.

Proof. See [Bec17, Lemma 12.3]. □

Consequently, the next algorithm consists in FISTA applied to the dual problem (5.12). It will be referred to as the *dual representation* since the iterations $(y^k)_k$ correspond to the Lagrange multipliers of our initial problem (5.8).

The Fast Dual Proximal Gradient (FDPG) Method - dual representation

Initialisation: $L \geq L_{\mathcal{F}} = \frac{\|\mathcal{A}\|^2}{\sigma}$, $w^0 = y^0 \in \mathbb{R}^m$ $t_0 = 1$.

Step for $k \geq 0$:

- $y^{k+1} = \text{prox}_{\frac{1}{L}\mathcal{G}}(w^k - \frac{1}{L}\nabla\mathcal{F}(w^k)),$
- $t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2},$
- $w^{k+1} = y^{k+1} + \left(\frac{t_k - 1}{t_{k+1}}\right)(y^{k+1} - y^k).$

Remark 5.1.19. *It is always possible to go back to the algorithm without acceleration by taking the constant sequence $t_k = 1$ for all $k \in \mathbb{N}$.*

By Theorem 5.1.15, we obtain a convergence rate of order $O(1/k^2)$ on the sequence $(q(y^k))_k$.

Theorem 5.1.20. *Let $(y^k)_k$ be the sequence defined by the FDPG method (dual representation) and y^* be a minimiser of q . Then for all $k \geq 1$,*

$$q(y^*) - q(y^k) \leq \frac{2L\|y^0 - y^*\|^2}{(k+1)^2}.$$

For the dual representation, the $O(1/k^2)$ convergence is preserved. However, it is not the case for the primal representation which is described below.

Now that we have the dual sequence $(y^k)_k$, we would like to recover a sequence $(x^k)_k$ that approaches the minimizer of (5.8). To do that, we will use the following lemma.

Lemma 5.1.21. *Let \mathcal{F} and \mathcal{G} be functions defined by (5.11). Then for all $y, v \in \mathbb{R}^m$ and $L > 0$, we have*

$$y = \text{prox}_{\frac{1}{L}\mathcal{G}}\left(v - \frac{1}{L}\nabla\mathcal{F}(v)\right) \quad (5.13)$$

if and only if

$$y = v - \frac{1}{L}\mathcal{A}\bar{x} + \frac{1}{L}\text{prox}_{Lg}(\mathcal{A}\bar{x} - Lv), \quad (5.14)$$

$$\text{where } \bar{x} = \text{argmax}_x \{\langle x | \mathcal{A}^T v \rangle - f(x)\}. \quad (5.15)$$

To prove this lemma, we need some properties of the conjugate.

Lemma 5.1.22. *Let $f: \mathbb{R}^n \rightarrow (\infty, \infty]$ and $f^*: \mathbb{R}^m \rightarrow \mathbb{R}$ be its conjugate, then we have for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$:*

$$\langle x | y \rangle = f(x) + f^*(y) \Leftrightarrow y \in \partial f(x),$$

where $\partial f(x)$ is the subdifferential of f such that $\partial f(x) = \{z \in \mathbb{R}^n; f(y) \geq f(x) + \langle z | y - x \rangle \text{ for all } y \in \mathbb{R}^n\}$.

If in addition f is closed, we have

$$\langle x | y \rangle = f(x) + f^*(y) \Leftrightarrow y \in \partial f(x) \Leftrightarrow x \in \partial f^*(y).$$

Proof. See [Bec17, Theorem 4.20]. □

Remark 5.1.23. *If f is differentiable, then the subdifferential of f is a singleton which is its gradient.*

The following lemma is a variant of the Moreau decomposition which claims that $\text{prox}_f(x) + \text{prox}_{f^*}(x) = x$ and also allows to compute the proximal operator of f^* from prox_f .

Lemma 5.1.24. *(extended Moreau decomposition) If $f: \mathbb{R}^n \rightarrow (-\infty, \infty]$ is proper, closed and convex and $\lambda > 0$, then for all $x \in \mathbb{R}^n$,*

$$\text{prox}_{\lambda f}(x) + \lambda \text{prox}_{\lambda^{-1}f^*}(x/\lambda) = x.$$

Proof. See [Bec17, Theorem 4.5]. □

The last property we need is the computation of a proximal operator of a function f composed with $-id$ where id is the identity.

Lemma 5.1.25. *Let $f: \mathbb{R}^n \rightarrow (-\infty, \infty]$ be a proper, closed and convex. We set $g(x) = f(-x)$ for all $x \in \mathbb{R}^n$. Then*

$$\text{prox}_g(x) = -\text{prox}_f(-x).$$

Proof. To show this equality, we use the change of variable $z = -u$:

$$\begin{aligned} \text{prox}_g(x) &= \operatorname{argmin}_u f(-u) + \frac{1}{2}\|u - x\|^2 \\ &= \operatorname{argmin}_{\{z, z=-u\}} f(z) + \frac{1}{2}\|-z - x\|^2 \\ &= -\operatorname{argmin}_z f(z) + \frac{1}{2}\|z + x\|^2 \\ &= -\text{prox}_f(-x). \end{aligned}$$

□

Now, we gather all these lemmas to show the primal representation of the algorithm.

Proof of Lemma 5.1.21. First, we investigate the form of $\nabla \mathcal{F}(v) = \nabla [f^*(\mathcal{A}^T v)]$. Let us define \bar{x} such that

$$\bar{x} = \nabla f^*(\mathcal{A}^T v).$$

Then, the equivalence in Lemma 5.1.22 tells us that \bar{x} is the point at which the maximum in the definition of $f^*(\mathcal{A}v)$ is reached, which gives (5.15)

$$\bar{x} = \operatorname{argmax}_x \{\langle x | \mathcal{A}^T v \rangle - f(x)\}.$$

Then by computing the derivative of an application composed with a linear application we obtain

$$\nabla [f^*(\mathcal{A}^T v)] = \mathcal{A} \nabla f^*(\mathcal{A}^T v) = \mathcal{A} \bar{x}.$$

Consequently, the expression of y becomes

$$y = \operatorname{prox}_{\frac{1}{L}\mathcal{G}} \left(v - \frac{1}{L} \mathcal{A} \bar{x} \right).$$

Since $\mathcal{G}(y) = g^*(-y)$, Lemma 5.1.25 gives

$$y = \operatorname{prox}_{\frac{1}{L}\mathcal{G}} \left(v - \frac{1}{L} \mathcal{A} \bar{x} \right) = -\operatorname{prox}_{\frac{1}{L}g^*} \left(\frac{1}{L} \mathcal{A} \bar{x} - v \right).$$

Next we use the extended Moreau decomposition with $\lambda = 1/L$ applied to g^* (and $g^{**} = g$ since g is proper, closed and convex [Bec17, Theorem 4.8]):

$$\begin{aligned} y &= -\operatorname{prox}_{\frac{1}{L}g^*} \left(\frac{1}{L} \mathcal{A} \bar{x} - v \right) = -\left(\frac{1}{L} \mathcal{A} \bar{x} - v - \frac{1}{L} \operatorname{prox}_{Lg}(\mathcal{A} \bar{x} - Lv) \right) \\ &= v - \frac{1}{L} \mathcal{A} \bar{x} + \frac{1}{L} \operatorname{prox}_{Lg}(\mathcal{A} \bar{x} - Lv), \end{aligned}$$

which is exactly (5.14). □

Thanks to Lemma 5.1.21, the primal representation of the FDPG method is the following.

The Fast Dual Proximal Gradient (FDPG) Method - primal representation

Initialisation: $L \geq L_{\mathcal{F}} = \frac{\|\mathcal{A}\|^2}{\sigma}$, $w^0 = y^0 \in \mathbb{R}^m$ $t_0 = 1$.

Step for $k \geq 0$:

- $u^k = \operatorname{argmax}_u \{\langle u | \mathcal{A}^T w^k \rangle - f(u)\},$
- $y^{k+1} = w^k - \frac{1}{L} \mathcal{A} u^k + \frac{1}{L} \operatorname{prox}_{Lg}(\mathcal{A} u^k - L w^k),$
- $t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2},$
- $w^{k+1} = y^{k+1} + \left(\frac{t_k - 1}{t_{k+1}} \right) (y^{k+1} - y^k).$

This algorithm is a reformulation of the FDPG (dual representation). As mentioned in Theorem 5.1.20, the convergence of the sequence $(q(y^k))_k$ is of order $O(1/k^2)$. By the following theorem, the rate of convergence is demoted to $O(1/k)$. Nevertheless, the convergence occurs on the primal sequence $(x_k)_k$ and not on the values $(f(x^k) + g(\mathcal{A}x^k))_k$. This is very important in those situations where the point found by the algorithm, which should be the optimal one or an approximation of it, is then inserted into new operators and this operation is iterated. An example of this can be found in the case of our mixed variational MFG in Section 5.3, where we insert the FDPG algorithm into the Banach fixed point sequence. Since the convergence is on $(x_k)_k$, this guarantees that we approach the exact solution to our problem.

Theorem 5.1.26. *Let $(y^k)_k$ be the sequence defined by the algorithm FDPG (primal representation) and y^* a minimiser of q . Let $(x^k)_k$ the sequence defined by*

$$x^k = \operatorname{argmax}_x \{ \langle x | \mathcal{A}^T y^k \rangle - f(x) \}$$

and x^* an optimal solution for (5.8). Then for all k ,

$$\|x^k - x^*\| \leq 2\sqrt{\frac{L}{\sigma}} \frac{\|y^0 - y^*\|}{k+1}.$$

Proof. See [BT14, Theorem 4.1]. □

5.2 Numerical simulations

Let $\{t_0, \dots, t_K\}$ and $\{x_0, \dots, x_N\}$ be the regular subdivisions of respectively $[0, T]$ and $[0, S]$ and $h = t_1 - t_0$ the time-step. To approximate the integral in \mathcal{V} , we use the left-rectangle method, hence we will consider solutions $(\rho(t_i, x_j))_{\substack{0 \leq i \leq K-1 \\ 0 \leq j \leq N-1}}$ in the vector space $\mathbb{R}^{K \cdot N}$. With an abuse of notations, we also write $V \in \mathbb{R}^{K \cdot N}$ the vector $(V(t_i, x_j))_{\substack{0 \leq i \leq K-1 \\ 0 \leq j \leq N-1}}$. In the following, the notations $\langle \cdot | \cdot \rangle$ and $\| \cdot \|$ designate respectively the scalar product and the euclidean norm in a vector space of finite dimension.

The problem that will be numerically solved is of the form

$$\min_{\rho \in \mathbb{R}^{K \cdot N}} f(\rho) + g(\mathcal{A}\rho).$$

Here $\rho \in \mathbb{R}^{K \cdot N}$ and $f(\rho) = h \langle V | \rho \rangle + h \frac{\|\rho\|^2}{2}$ is proper, closed and h -strongly convex and $g(\rho_0, \rho_1, \rho_2) = \lambda \|\rho_0\|_1 + \delta_{C_0}(\rho_1) + \delta_{C_1}(\rho_2)$ is proper, closed and convex. The notation $\| \cdot \|_1$ refers to the ℓ_1 -norm in a vector space of finite dimension and δ_{C_i} is the indicator function defined by

$$\delta_{C_i}(\rho) = \begin{cases} 0, & \rho \in C_i, \\ +\infty, & \rho \notin C_i, \end{cases}$$

where $C_0 = (\mathbb{R}_+)^{K \cdot N}$ and $C_1 = \{\rho \in \mathbb{R}^{K \cdot N}; \forall i \in \{0, \dots, K-1\}, \sum_{j=0}^{N-1} \rho(t_i, x_j) \cdot l = 1\}$ are the sets of constraints. The linear transformation \mathcal{A} is $\mathcal{A}(\rho) = (A\rho, \rho, \rho)$, where the matrix A will be detailed for each case.

Note that A should represents the difference between two successive terms in time such that

$$\forall i \in \{0, \dots, K-2\}, \forall j \in \{0, \dots, N-1\}, (A\rho)_{iN+j} = \rho(t_{i+1}, x_j) - \rho(t_i, x_j),$$

so $\|\mathcal{A}\|^2$ is computed as follows

$$\begin{aligned} \|\mathcal{A}\rho\|^2 &= \sum_{i=0}^{K-2} \sum_{j=0}^{N-1} [\rho(t_{i+1}, x_j) - \rho(t_i, x_j)]^2 + 2 \sum_{i=0}^{K-1} \sum_{j=0}^{N-1} \rho(t_i, x_j)^2 \\ &\leq \sum_{i=0}^{K-2} \sum_{j=0}^{N-1} 2\rho(t_{i+1}, x_j)^2 + 2\rho(t_i, x_j)^2 + 2 \sum_{i=0}^{K-1} \sum_{j=0}^{N-1} \rho(t_i, x_j)^2 \\ &\leq 6 \sum_{i=0}^{K-1} \sum_{j=0}^{N-1} \rho(t_i, x_j)^2 = 6\|\rho\|^2. \end{aligned}$$

This explains the choice of $L \geq \|\mathcal{A}\|^2/h$ in Table 5.1 and 5.2.

The algorithm that we use is the primal representation of the FDPG method presented in the previous section.

Since g is a sum of separable functions, its proximal operator is

$$\text{prox}_g(\rho_0, \rho_1, \rho_2) = \left(\text{prox}_{\lambda\|\cdot\|_1}(\rho_0), \text{prox}_{\delta_{C_0}}(\rho_1), \text{prox}_{\delta_{C_1}}(\rho_2) \right).$$

The algorithm is described below.

<p>Initialization: $L \geq \frac{\ A\ ^2}{h}$, $w^0 = y^0 \in (\mathbb{R}^{K \cdot N})^3$, $t_0 = 1$.</p> <p>Step for $k \geq 0$:</p> <ul style="list-style-type: none"> • $u^k = \text{argmax}_u \{ \langle u \mathcal{A}^T w^k \rangle - f(u) \} = \frac{\mathcal{A}^T w^k}{h} - V$ $= \frac{1}{h} (A^T w_0^k + w_1^k + w_2^k) - V$ • $y_0^{k+1} = w_0^k - \frac{1}{L} A u^k + \frac{1}{L} \text{prox}_{L\lambda\ \cdot\ _1}(A u^k - L w_0^k)$ • $y_1^{k+1} = w_1^k - \frac{1}{L} u^k + \frac{1}{L} \text{prox}_{L\delta_{C_0}}(u^k - L w_1^k)$ • $y_2^{k+1} = w_2^k - \frac{1}{L} u^k + \frac{1}{L} \text{prox}_{L\delta_{C_1}}(u^k - L w_2^k)$ • $t^{k+1} = \frac{k+1+a}{a}$ • $w^{k+1} = y^{k+1} + \frac{t^k - 1}{t^{k+1}} (y^{k+1} - y^k)$
--

As discussed in Remark 5.1.16, the parameter t^k is of the form $\frac{k+1+a}{a}$ such that a is a constant greater or equal than 2 and its value is displayed in Table 5.1 and 5.2.

Note that our problem is essentially, up to minor modifications and the presence of extra constraints (unit mass, positivity, Dirichlet conditions), a simplified version of some standard problems in image denoising based on total variation (see, for instance, the classical paper [ROF92]): here, whether the problem is 1D or 2D or higher-dimensional, the main feature is that the total variation is only computed in time.

The following subsections describe different examples of solutions to (5.1) by using this algorithm. For each case, the differences with the description above will be specified.

5.2.1 1D in time and periodic

When we consider the periodic problem (in time, so that the interval $[0, T]$ becomes a circle of length T), and we assume $S = T$ and $V(t, x) = v(t - x)$ for an S -periodic function v , it is possible to reduce the problem to the one dimensional case, namely a problem with one only variable in $[0, T]$ instead of $[0, T] \times [0, S]$. One expects the solution $\rho(t, x) = u(t - x)$ to be transported according to time. The uniqueness of the solution and the symmetry with respect to translations in both time and space (replacing (t, x) with $(t + \delta, x + \delta)$) show that the solution should indeed be of this form. Then, a change of variables $y = t - x$ can be carried out in \mathcal{V} as following:

$$\begin{aligned} \mathcal{V}(\rho) &= \int_0^T \int_0^S \left(\lambda |\dot{u}(t-x)| + v(t-x)u(t-x) + \frac{u(t-x)^2}{2} \right) dx dt \\ &= \int_0^T \int_t^{t-S} \left(\lambda |\dot{u}(y)| + v(y)u(y) + \frac{u(y)^2}{2} \right) dy dt \\ &= T \int_0^S \left(\lambda |\dot{u}(y)| + v(y)u(y) + \frac{u(y)^2}{2} \right) dy. \end{aligned}$$

The problem reduces hence to the search of an S -periodic solution with one variable. By the way, up to multiplicative and additive constants this problem is equivalent to minimizing

$$\int_0^S \left(\lambda |\dot{u}(y)| + \frac{|u(y) + v(y)|^2}{2} \right) dy,$$

which is exactly the problem described as an example at the end of Section 4.1, with $\omega = -v$, except for the constraints $\int_0^S u(y) dy = 1$ and $u \geq 0$.

In the following subsections, we will see that the solution obtained through numerical simulations coincide with u computed as above.

For the numerical simulation shown in Figure 5.1 we take $v(y) = -a_0 \cos(\frac{2\pi}{S}y)$ and the choice of the parameters are shown in Table 5.1.

Let $\{y_0, \dots, y_K\}$ be a regular subdivision of $[0, S]$ and $h := y_1 - y_0$ the step-size of the subdivision. By the left-rectangle method, the problem is approximated by

$$\sum_{i=0}^{K-1} \left(\lambda |u(y_{i+1}) - u(y_i)| + v(y_i)u(y_i)h + \frac{u(y_i)^2}{2}h \right) + \delta_{C_0}(u) + \delta_{C_1}(u).$$

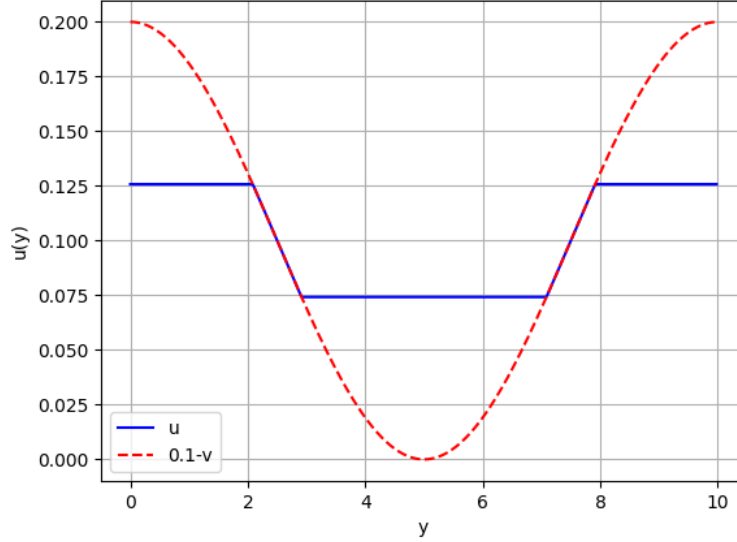


Figure 5.1: The simulation of the solution u to the 1D periodic case with $v(y) = -a_0 \cos(\frac{2\pi}{S}y)$. The parameters are displayed in Table 5.1. The blue solid line corresponds to the solution u , while the red dashed line is the profile of $c - v$ with $c = 1/S$.

Parameter	Value
S	10
K	500
h	0.02
a	200
L	$6/h$
a_0	0.1
λ	0.1

Table 5.1: Parameters for the solution to the problem in Fig. 5.1.

We observe that with this choice of parameters the solution u is strictly positive and “follows” the profile of $c - v$ for a constant c which appears as a Lagrange multiplier for the mass constraint and allows to obtain unit mass. The function u is also a solution of the problem at the end of Section 4.1 with $\omega = c - v$ and its profile, shown in Figure 5.1, is consistent with the explicit description of the solution which we gave.

Remark 5.2.1. *If $v(y) = -a_0 \cos(\frac{2\pi}{S}y)$, it is possible to compute the critical λ at which the aspect of the solution u switches from Figure 5.1 to the constant solution. The Euler-Lagrange equation of*

$$\min_{\substack{u \in \text{BV}([0, S]; \mathbb{R}); \\ \int_0^S u(y) dy = 1, u \geq 0}} \int_0^S \left(\lambda |\dot{u}(y)| + v(y)u(y) + \frac{u(y)^2}{2} \right) dy$$

is $z' = v + u - c$ on $[0, S]$ where $z(y) \in \partial(\lambda |\cdot|)(u'(y))$ and $c = 1/S$ is the constant due to the mass constraint $\int_0^S u(y) dy = 1$. Assuming that the solution becomes constant when $\varepsilon = \frac{S}{4}$ and extending the

solution periodically on $[-S, 0]$, we integrate the equation $z' = v + u - c$ over $[-\varepsilon, \varepsilon]$:

$$\begin{aligned} -2\lambda &= \int_{-\varepsilon}^{\varepsilon} z'(y) dy = \int_{-\varepsilon}^{\varepsilon} (v(y) + u(y) - c) dy = \int_{-\varepsilon}^{\varepsilon} (v(y) - v(\varepsilon)) dy \\ &= \int_{-\varepsilon}^{\varepsilon} \left(-a_0 \cos\left(\frac{2\pi}{S}y\right) + a_0 \cos\left(\frac{2\pi}{S}\varepsilon\right) \right) dy = -a_0 \frac{S}{\pi} \sin\left(\frac{2\pi}{S}\varepsilon\right) + 2\varepsilon a_0 \cos\left(\frac{2\pi}{S}\varepsilon\right). \end{aligned}$$

By taking $\varepsilon = \frac{S}{4}$, we obtain that $\lambda = \frac{a_0 S}{2\pi}$.

When $\lambda > \frac{a_0 S}{2\pi}$, the solution u is constant equal to c . While the condition $\lambda < \frac{a_0 S}{2\pi}$ gives the solution u as in Figure 5.1. This explains the choice of parameter λ .

5.2.2 1D in time and 1D in space, periodic in time

We consider now the case where we keep two variables, but we assume the time domain to be periodic. This case is slightly simpler to handle from the point of view of the discretization of the time derivative.

By using the subdivisions described at the beginning of Section 5.2, we approximate the integral \mathcal{V} by the left-rectangle method as following:

$$\sum_{i=0}^{K-1} \sum_{j=0}^{N-1} \left(\lambda |\rho(t_{i+1}, x_j) - \rho(t_i, x_j)| l + V(t_i, x_j) \rho(t_i, x_j) h l + \frac{\rho(t_i, x_j)^2}{2} h l \right),$$

where $h := t_1 - t_0$ and $l := x_1 - x_0$ are the step-sizes of each subdivision. We will consider that $\rho(t_0, x_j) = \rho(t_K, x_j)$ for all $j \in \{0, N-1\}$.

With this discretization, we look for a solution in the space $\mathbb{R}^{K \cdot N}$ where ρ is viewed as a vector.

The discretized problem is

$$\min_{\rho \in \mathbb{R}^{K \cdot N}} \sum_{i=0}^{K-1} \sum_{j=0}^{N-1} \left(\lambda |\rho(t_{i+1}, x_j) - \rho(t_i, x_j)| l + V(t_i, x_j) \rho(t_i, x_j) h l + \frac{\rho(t_i, x_j)^2}{2} h l \right) + \delta_{C_0}(\rho) + \delta_{C_1}(\rho). \quad (5.16)$$

To avoid numerical errors due to small values, searching a minimizer of (5.16) is the same as searching a minimizer to the problem divided by l :

$$\min_{\rho \in \mathbb{R}^{K \cdot N}} \sum_{i=0}^{K-1} \sum_{j=0}^{N-1} \left(\lambda |\rho(t_{i+1}, x_j) - \rho(t_i, x_j)| + V(t_i, x_j) \rho(t_i, x_j) h + \frac{\rho(t_i, x_j)^2}{2} h \right) + \delta_{C_0}(\rho) + \delta_{C_1}(\rho).$$

The function $f: \mathbb{R}^{K \cdot N} \rightarrow \mathbb{R}$ defined by

$$f(\rho) = \sum_{i=0}^{K-1} \sum_{j=0}^{N-1} \left(V(t_i, x_j) \rho(t_i, x_j) h + \frac{\rho(t_i, x_j)^2}{2} h \right) = h \langle V | \rho \rangle + h \frac{\|\rho\|^2}{2} \quad (5.17)$$

corresponds to the function f described in Section 5.2.

The linear transformation $\mathcal{A}: \mathbb{R}^{K \cdot N} \rightarrow (\mathbb{R}^{K \cdot N})^3$ is the same as in Section 5.2. The squared matrix A is of order $K \cdot N$ and its coordinates are

$$A_{i,j} = \begin{cases} -1 & \text{if } j = i, \\ 1 & \text{if } j \equiv i + N[K \cdot N], \\ 0 & \text{otherwise,} \end{cases}$$

namely

$$A = \begin{pmatrix} -I_N & I_N & & \\ & \ddots & \ddots & \\ & & -I_N & I_N \\ I_N & & & -I_N \end{pmatrix}$$

where I_N is the identity matrix of order N and blank space corresponds to zeros.

The function $g: (\mathbb{R}^{K \cdot N})^3 \rightarrow \mathbb{R}$ is defined as in Section 5.2:

$$\begin{aligned} g(\rho_0, \rho_1, \rho_2) &= \sum_{i=0}^{K-1} \sum_{j=0}^{N-1} \lambda |\rho_0(t_i, x_j)| + \delta_{C_0}(\rho_1) + \delta_{C_1}(\rho_2) \\ &= \lambda \|\rho_0\|_1 + \delta_{C_0}(\rho_1) + \delta_{C_1}(\rho_2). \end{aligned} \quad (5.18)$$

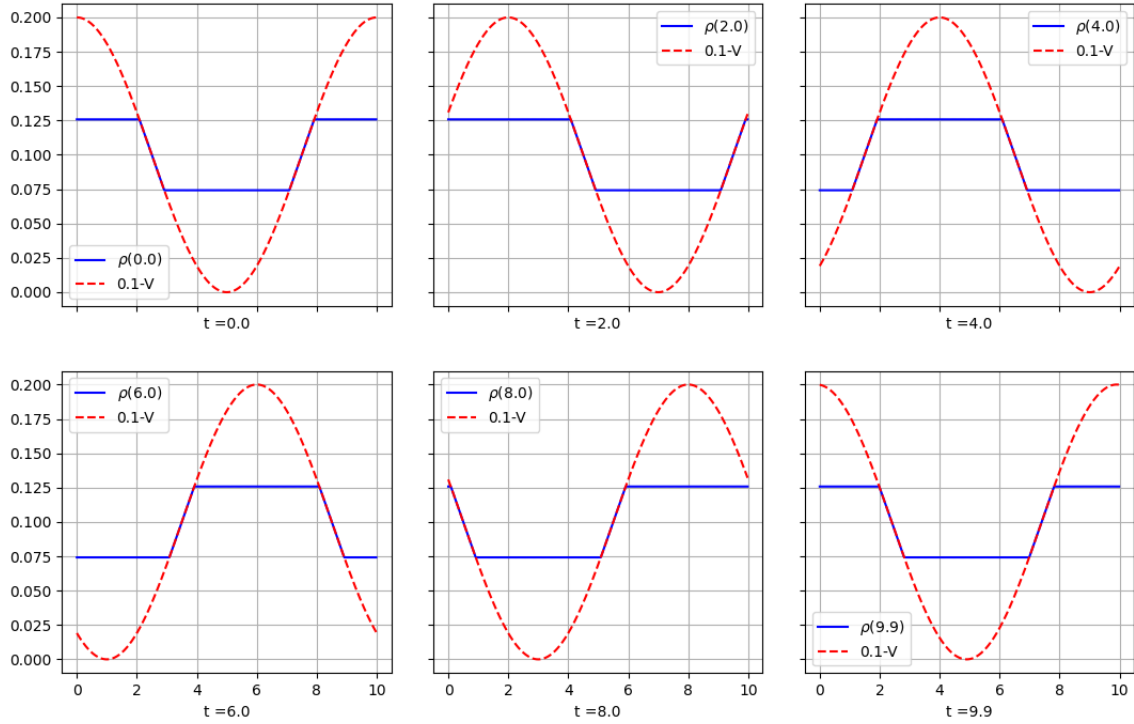


Figure 5.2: Profile of the solution to (5.1) with $V(t, x) = -a_0 \cos(\frac{2\pi}{S}(t - x))$ at times $t = 0, 2, 4, 6, 8, 9.9$. The blue solid line describes the solution ρ at each specified time and the red dashed line is the profile of $0.1 - V$ at each time. The parameters are displayed in Table 5.2.

Parameter	Value
T	10
S	10
K	100
N	500
a	200
h	0.02
L	$6/h$
a_0	0.1
λ	0.1

Table 5.2: Parameters for the numerical simulation of the solution to (5.1).

In Figure 5.2, we consider the same case as in Figure 5.1, but with a dynamic approach. The profile of the solution ρ at time 0 is the same as in Figure 5.1. In Figure 5.1 the problem is viewed on one dimensional space and here we can see that the solution is transported according to time from the left to the right.

The next case (see Figure 5.3) we show is different. Here we impose a Dirichlet boundary condition at time 0 which corresponds here to m_0 being the constant density equal to $1/S$. This is the same as studying the non-periodic problem on $[0, T]$ and imposing two Dirichlet boundary conditions, which are (by chance) equal. In this case, keeping the periodic structure allows to use a simpler form of the matrix A .

Let $m_0 \in \mathbb{R}^N$ a vector verifying the constraints, i.e.,

$$m_0 \in (\mathbb{R}_+)^N \text{ and } \sum_{j=0}^{N-1} m_{0,j} l = 1.$$

Let us define the new set $C_2 = \{\rho \in \mathbb{R}^{K \cdot N}; \forall j \in \{0, \dots, N-1\}, \rho(0, x_j) = m_{0,j}\}$ which encodes the constraint $\rho(0) = m_0$.

The function f is the same as in (5.17). The linear transformation is now $\mathcal{A}\rho = (A\rho, \rho, \rho, \rho)$ and the

function $g: (\mathbb{R}^{K \cdot N})^4 \rightarrow \mathbb{R}$ is

$$\begin{aligned} g(\rho_0, \rho_1, \rho_2, \rho_3) &= \sum_{i=0}^{K-1} \sum_{j=0}^{N-1} \lambda |\rho_0(t_i, x_j)| + \delta_{C_0}(\rho_1) + \delta_{C_1}(\rho_2) + \delta_{C_2}(\rho_3) \\ &= \lambda \|\rho_0\|_1 + \delta_{C_0}(\rho_1) + \delta_{C_1}(\rho_2) + \delta_{C_2}(\rho_3). \end{aligned}$$

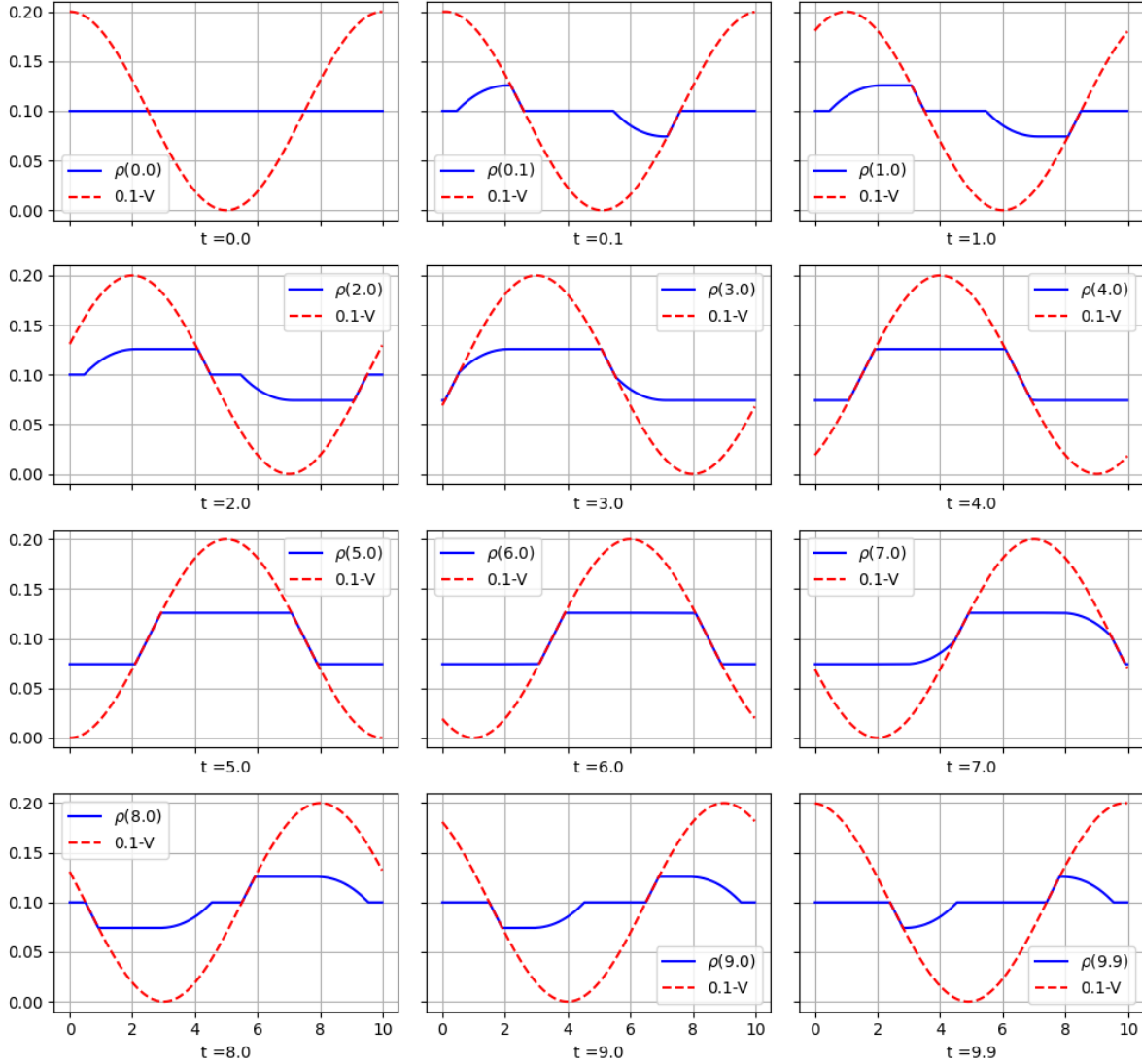


Figure 5.3: Simulation of the solution to (5.1) at times $t = 0, 0.1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 9.9$ with $V(t, x) = -a_0 \cos(\frac{2\pi}{S}(t - x))$ and $m_0(x) = 1/S$ with parameters from Table 5.2 except for parameter L where $L = 7/h$. The blue solid line is the profile of the solution ρ and the red dashed line corresponds to $0.1 - V$ taken at each time t .

In Figure 5.3 we see that the Dirichlet condition is indeed verified at $t = 0$ and then, immediately, the solution jumps (coherently with Remark 4.1.3) from the constant state to $t = 0.1$. One can notice that the profile is different from Figure 5.2 in a way that on the subintervals where the solution does not follow $0.1 - V$, it is not constant in space anymore. However, between times $t = 4$ and $t = 6$, the solution comes back to the profile when there is no boundary condition, namely, it is constant when it does not follow $0.1 - V$. When $t \geq 6$, the profile of the solution varies again on the subintervals where it should be constant.

5.2.3 1D in time, 1D in space and non periodic in time

Unlike Sections 5.2.1 and 5.2.2, we do not impose anymore a periodic time behavior, and the value $\rho(t_K, x_j)$ of the solution $\rho \in \mathbb{R}^{(K+1) \cdot N}$ may be different from $\rho(t_0, x_j)$. The integral in \mathcal{V} is now approxi-

mated by:

$$\sum_{j=0}^{N-1} \left(\sum_{i=0}^{K-1} \lambda |\rho(t_{i+1}, x_j) - \rho(t_i, x_j)| l + \sum_{i=0}^K \left(V(t_i, x_j) \rho(t_i, x_j) h l + \frac{\rho(t_i, x_j)^2}{2} h l \right) \right). \quad (5.19)$$

Remark 5.2.2. *The formula (5.19) is different from (5.16) in the sense that there is the additional term $\sum_{j=0}^{N-1} V(t_K, x_j) \rho(t_K, x_j) h l + \frac{\rho(t_K, x_j)^2}{2} h l$ that depends on t_K which should not appear in the left-rectangle method. However, with this choice of discretization, the function f is strongly convex in $\rho \in \mathbb{R}^{(K+1) \cdot N}$ which allows us to apply the algorithm and the formula (5.19) still converges towards the integral $\int_0^T \int_0^S (\lambda |\dot{\rho}| + V\rho + \frac{\rho^2}{2})$ as $h \rightarrow 0$ and $l \rightarrow 0$.*

The new function f is

$$f(\rho): \mathbb{R}^{(K+1) \cdot N} \longrightarrow \mathbb{R} \quad (5.20)$$

$$\rho \mapsto h \langle V | \rho \rangle + h \frac{\|\rho\|^2}{2} = \sum_{i=0}^K \sum_{j=0}^{N-1} \left(V(t_i, x_j) \rho(t_i, x_j) h + \frac{\rho(t_i, x_j)^2}{2} h \right).$$

The function g is defined as in (5.18) and \mathcal{A} is the same as in Section 5.2 with a different matrix A which is rectangle of dimension $(K \cdot N) \times ((K+1) \cdot N)$ such that

$$A_{i,j} = \begin{cases} -1 & \text{if } i = j, \\ 1 & \text{if } j = i + N, \\ 0 & \text{otherwise,} \end{cases}$$

namely

$$A = \begin{pmatrix} -I_N & I_N & & \\ & \ddots & \ddots & \\ & & -I_N & I_N \end{pmatrix}.$$

Figure 5.4 shows the simulation of the solution ρ to the problem (5.1) with a given V which is not periodic anymore (here $V(t, x) = (t - x)^2$). No Dirichlet boundary conditions nor penalizations ψ_t at $t = 0, T$ are considered. The different times are chosen to show the most significant changes in the profile of ρ . Since $V(t, x)$ is minimal at $x = t$, the general behavior is that the solution is transported from the left to the right. Between times $t = 1.5$ and $t = 8.3$, the solution either follows $c - V$ (where c is a constant to define) or it is constant. Close to the time boundaries, the behavior is different.

The next case that we present (Figure 5.5) involves Dirichlet conditions in time. Let $m_0 \in \mathbb{R}^N$ and $m_T \in \mathbb{R}^N$ be two vectors verifying the boundary conditions, i.e.,

$$m_0 \in (\mathbb{R}_+)^N \text{ and } \sum_{j=0}^{N-1} m_{0,j} l = 1,$$

$$m_T \in (\mathbb{R}_+)^N \text{ and } \sum_{j=0}^{N-1} m_{T,j} l = 1.$$

Let us define the new set of constraints by $C_2 = \{\rho \in \mathbb{R}^{K \cdot N}; \forall j \in \{0, \dots, N-1\}, \rho(0, x_j) = m_{0,j} \text{ and } \rho(T, x_j) = m_{T,j}\}$. The function f is defined as in (5.20), $\mathcal{A}\rho = (A\rho, \rho, \rho, \rho)$ and the function $g: (\mathbb{R}^{K \cdot N})^4 \rightarrow \mathbb{R}$ is

$$\begin{aligned} g(\rho_0, \rho_1, \rho_2, \rho_3) &= \sum_{i=0}^{K-1} \sum_{j=0}^{N-1} \lambda |\rho_0(t_i, x_j)| + \delta_{C_0}(\rho_1) + \delta_{C_1}(\rho_2) + \delta_{C_2}(\rho_3) \\ &= \lambda \|\rho_0\|_1 + \delta_{C_0}(\rho_1) + \delta_{C_1}(\rho_2) + \delta_{C_2}(\rho_3). \end{aligned}$$

Remark 5.2.3. *Instead of adding the constraint $\rho(0) = m_0$ in the problem, one could have discretized the integral with the right-rectangle method in time and directly used that $\rho(0) = m_0$ in the algorithm.*

The profile of Figure 5.5 is quite similar to Figure 5.4 and mainly differs around $t = 0, T$. Again, a jump in time is observed.

The very last example we consider is non-periodic in time and involves both a Dirichlet condition at $t = 0$ and penalization at $t = T$. We consider the penalization

$$\psi_T(\rho(T)) = \int_0^S \Psi_T(x) \rho(T, x) dx.$$

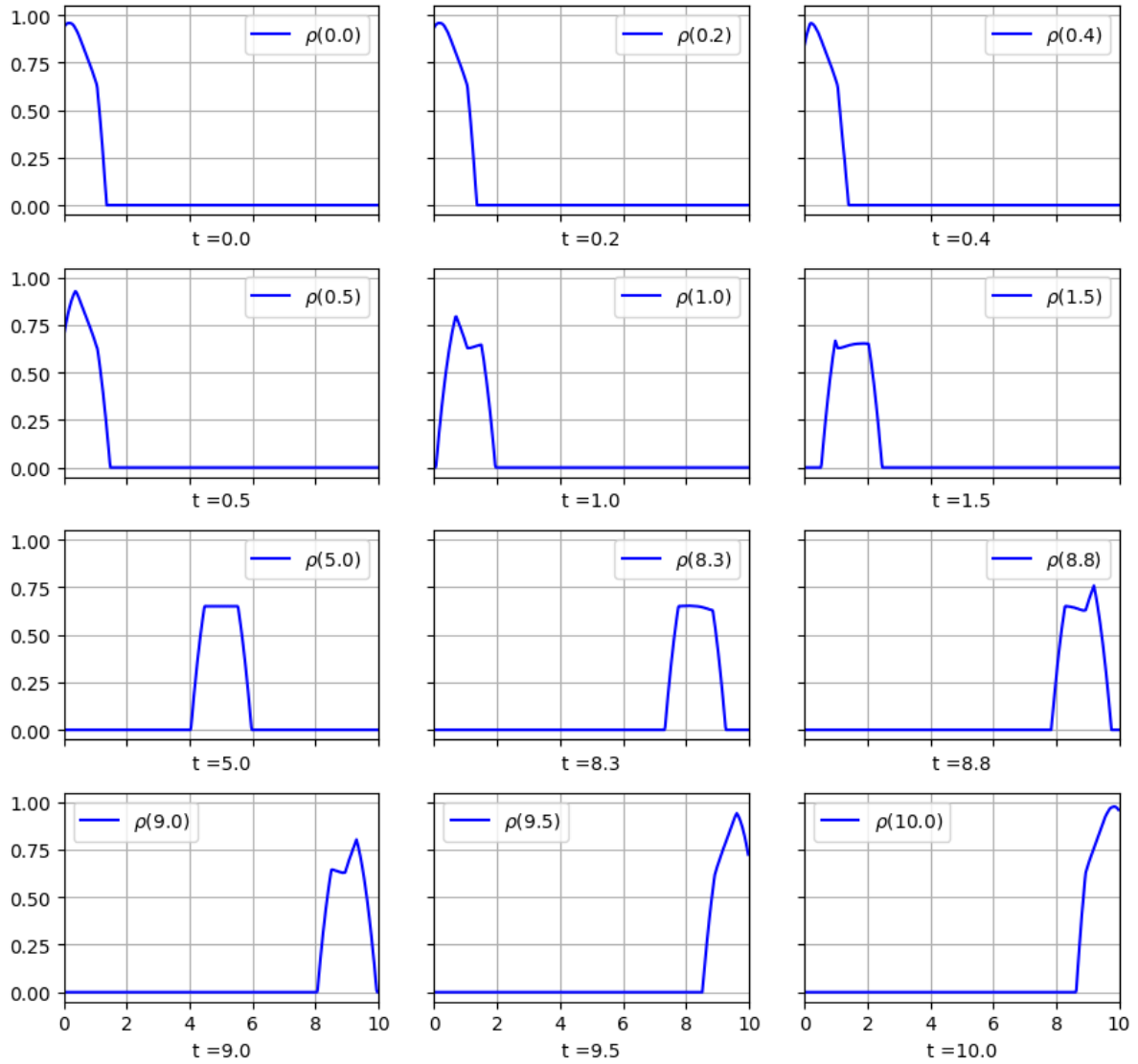


Figure 5.4: The simulation of the solution to (5.1) at times $t = 0, 0.2, 0.4, 0.5, 1, 1.5, 5, 8.3, 8.8, 9, 9.5, 10$ with $V(t, x) = (t - x)^2$ and parameters from Table 5.2. The blue solid line corresponds to the solution ρ at different times.

The left-rectangle method applied to ψ_T gives

$$\psi_T(\rho(T)) = \sum_{j=0}^{N-1} \Psi_T(x_j) \rho(T, x_j) l.$$

We use the discretization from (5.19). The difference is in the definition of V . We shall take \tilde{V} such that

$$\tilde{V} = V + (0, \dots, 0, \frac{\Psi_T}{h}).$$

Again, the profile of Figure 5.6 is similar to Figures 5.4 and 5.5 except for the behavior close to the time boundaries. In Figure 5.6, we have put penalization at times 0 and T such that it costs less to be close to 0 in space, hence the concentration of the mass at time 0 in $x = 0$ and the jump at time T .

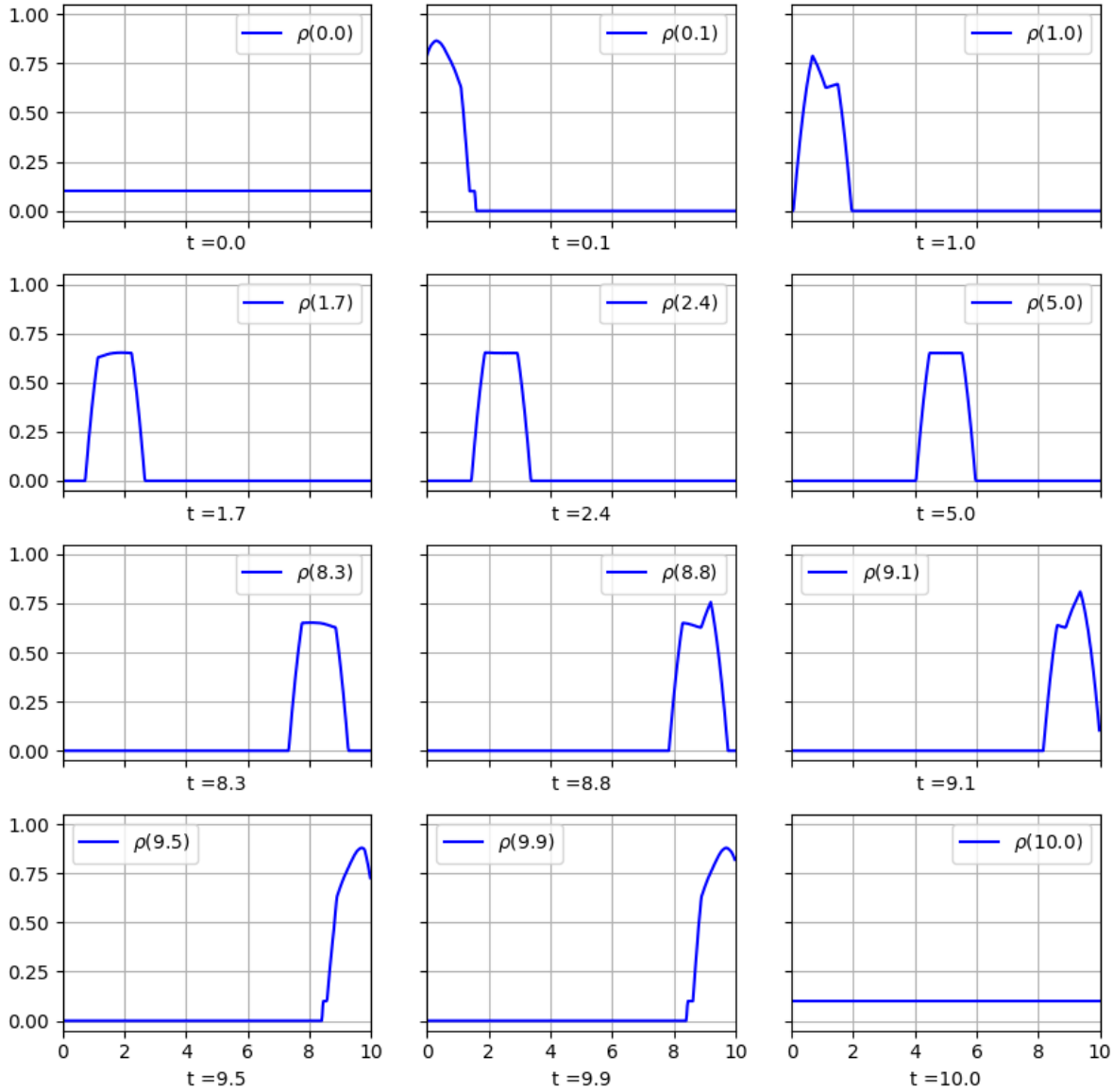


Figure 5.5: The simulation of the solution to (5.1) at times $t = 0, 0.1, 1, 1.7, 2.4, 5, 8.3, 8.8, 9.1, 9.5, 9.9, 10$ with $V(t, x) = (t - x)^2$, $m_0(x) = m_T(x) = 1/S$ and parameters from Table 5.2 except from L which is $L = 7/h$. The blue solid line corresponds to the solution ρ at different times.

5.3 Numerical simulations for the density dependent case

In Section 2.2, we showed that thanks to Banach fixed point theorem there exists a density $\bar{\rho}$ which minimizes

$$\min_{\substack{\rho \in L^2([0, T] \times \Omega) \\ \forall t \in [0, T], \int_{\Omega} \rho(t, x) dx = 1 \\ \rho \geq 0}} \mathcal{V}_{\bar{\rho}}(\rho) = \int_0^T \int_{\Omega} \left(\lambda |\dot{\rho}(t, x)| + V(\bar{\rho})(t, x) \rho(t, x) + c_0 \frac{\rho(t, x)^2}{2} \right) dx dt + \psi_0(\rho(0)) + \psi_T(\rho(T)). \quad (5.21)$$

As in the previous section, we choose the space $\Omega = [0, S]$ to be one-dimensional. This problem differs from the one presented at the beginning of this chapter in (5.1) in a way that $V: L^2 \rightarrow L^2$ is a functional (where the functional space may either be $L^2([0, T] \times \Omega)$ or $L^2(\Omega)$) and we add a factor c_0 to regulate the c_0 -strong convexity of our functional. Since the solution is obtained through a Banach fixed point theorem, one needs the functional $\mathcal{F}: \bar{\rho} \mapsto \operatorname{argmin} \mathcal{V}_{\bar{\rho}}(\cdot)$ to be a contraction mapping and this is achieved whenever $\frac{\operatorname{Lip}(V)}{c_0} < 1$. In addition, this suggests a numerical method to approach $\bar{\rho}$ by computing the sequence $(\mathcal{F}^k(\rho^0))_k$. At each iteration k for a given ρ^k , we use the FDPG method to compute the next iteration $\rho^{k+1} = \mathcal{F}(\rho^k)$ with the initialization at ρ^k . More precisely, we use the algorithm presented at the beginning of Section 5.2 where instead of V , we take $V(\rho^k)$.

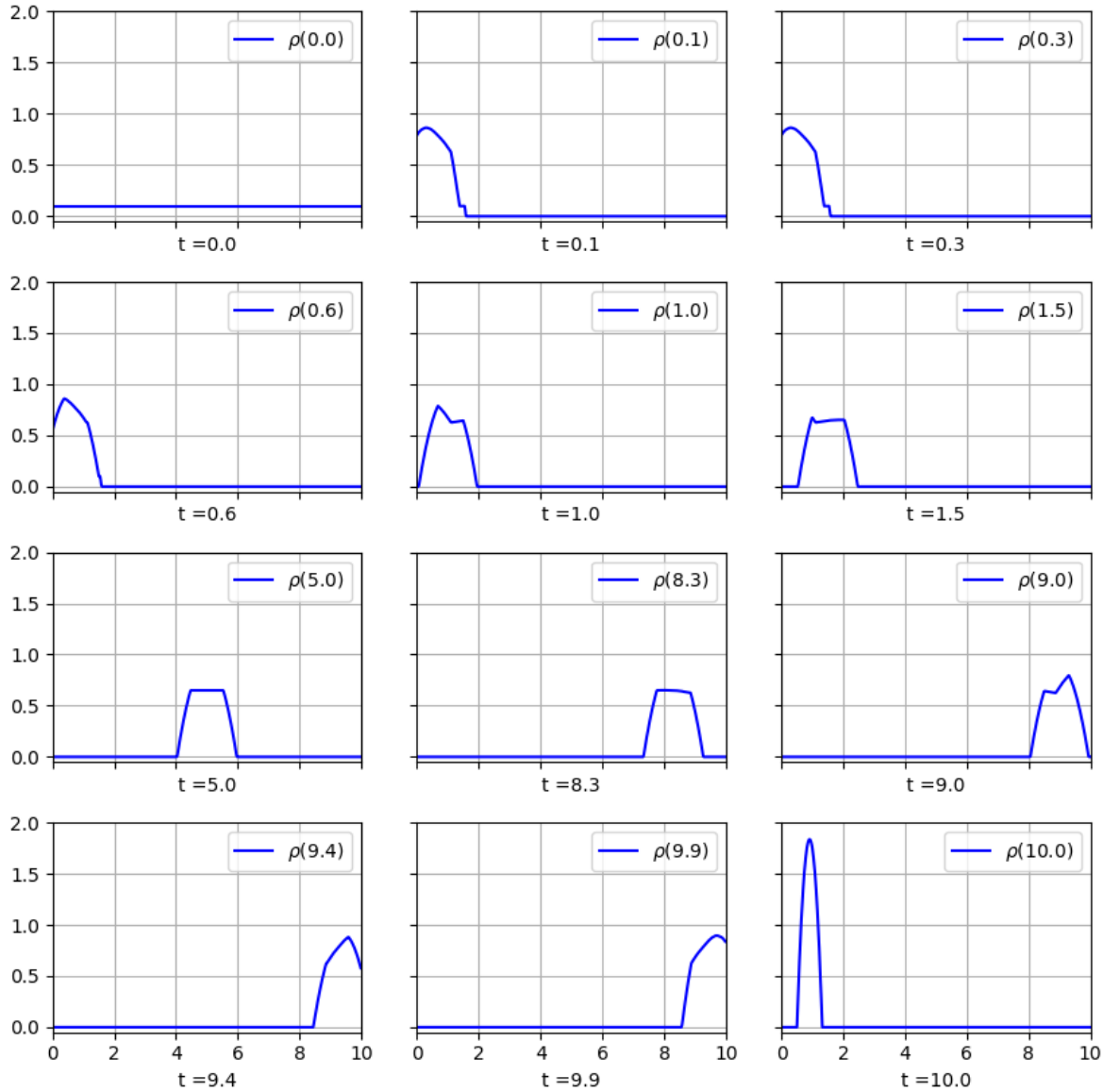


Figure 5.6: The simulation of the solution to (5.1) at times $t = 0, 0.1, 0.3, 0.6, 1, 1.5, 5, 8.3, 9, 9.4, 9.9, 10$ with $V(t, x) = (t - x)^2$, $m_0(x) = 1/S$, $\varphi_T(x) = x^2$ and parameters from Table 5.2 except from L which is $7/h$. The blue solid line corresponds to the solution ρ at different times.

In this section, we present three examples of functions $V(\rho) = V_i(\rho) + V_0$ for $i = 1, 2, 3$. It is important to consider a non-autonomous term V_0 in order to have a more dynamic solution (see the conclusion of Proposition 2.2.7). The function V_0 is defined as

$$V_0(t, x) := \left(\frac{S}{2} + 3 \cos \left(\frac{4\pi}{T} t \right) - x \right)^2, \quad (5.22)$$

so that at time t , it is more interesting to be at position $\frac{S}{2} + 3 \cos \left(\frac{4\pi}{T} t \right)$.

The discretization is the same as in the previous section.

$$(i) \quad V_1(\rho)(t, x) = \left(\int_{x-\delta}^{x+\delta} \rho(t, y) dy \right)^2.$$

In this example V_1 is a function of a convolution of ρ (here, the convolution with an indicator function). From a modeling point of view, this means that the running cost at a point depends on the mass of ρ in a fixed neighborhood of such a point. Note that, in many cases, using $V_1[\rho] = \eta * \rho$ would be the first variation of a functional (if η is symmetric it is the first variation of $\rho \mapsto \int \int \rho(x) \eta(x - y) \rho(y) dy dx$) but this fails when it is composed with a nonlinear function such as the square here, which makes this term non-variational (it would become variational again if convolved a second time, in the form $\eta * (f'(\eta * \rho))$).

Parameter	Value
T	10
S	10
K	1000
N	1000
a	200
h	0.01
L	$6/(c_0 h)$
λ	0.2
c_0	3

Table 5.3: Parameters for the numerical simulation of the solution in examples (i), (ii) and (iii).

which is the first variation of $\rho \mapsto \int f(\eta * \rho)$. To be more precise, the expression of V_1 has to be modified when $x \pm \delta$ is out of bounds of $[0, S]$ for a given $\delta > 0$, so a more precise definition is the following:

$$V_1(\rho)(t, x) = \begin{cases} \left(\int_0^{x+\delta} \rho(t, y) dy \right)^2, & \text{if } x \in [0, \delta], \\ \left(\int_{x-\delta}^{x+\delta} \rho(t, y) dy \right)^2, & \text{if } x \in]\delta, S - \delta[, \\ \left(\int_{x-\delta}^S \rho(t, y) dy \right)^2, & \text{if } x \in]S - \delta, S]. \end{cases}$$

The aspect of the solution is shown in Figure 5.7.

$$(ii) \quad V_2(\rho)(t, x) = \min \left(\int_0^x \arctan(\rho(t, y)) dy, \int_x^S \arctan(\rho(t, y)) dy \right).$$

This second example models a problem where players want to be ready to escape the domain, exiting either from one side or from the other, but the cost to escape depends on congestion. More precisely, a part of the cost is a congested distance to the boundary in the spirit of the Hughes' model for crowd motion, see [Hug02]. We compute the congestion in terms of a non-linear and bounded function of the density (here, the arctangent), so that the resulting function is Lipschitz in x whatever is ρ . In the present case, $V_2: L^2([0, T] \times \Omega) \rightarrow L^2([0, T] \times \Omega)$ is a 1-Lipschitz operator (because \arctan is 1-Lipschitz), so the parameter $c_0 = 3$ is more than suitable for \mathcal{F} to be a contraction. This example was already mentioned in Section 2.2 and can be seen as a 1D case of the model where $V(\rho) = V_0 + u_\rho$ with u_ρ the viscosity solution to $|\nabla u| = g(\eta * \rho)$ with Dirichlet boundary conditions $u = 0$ on $\partial\Omega$ (here we omit the convolution, or, equivalently, we choose $\eta = \delta_0$).

Compared to the previous example, we keep the same values of the parameters in Table 5.3, but we change the value of λ , taking $\lambda = 1$ (instead of 0.2) in order to observe any effect of an increased cost of jumping.

The profile of the solution is shown in Figure 5.8.

The increased cost of jumping has influenced the solution to be less flexible than the one in Figure 5.7. Starting from $t = 0$, as the density moves from the right to the left, the population is more concentrated near the positions where the minimum of V_0 should approach during a large amount of time, namely $x = 7$ and $x = 3$, whereas the density is less important in places where the minimum of V_0 only passes through, e.g. $x = 5$. This behavior may be caused by the cost of jumping which is higher than the one in Figure 5.7. Indeed, the agents anticipate the position at which it will cost less to live, i.e. near the minimum of V_0 , in order to minimize the number of jumps.

$$(iii) \quad V_3(\rho)(t, x) = \int_0^t \rho(s, x) ds.$$

This example corresponds to the last possibility evoked in Section 2.2 and is also non-variational. As we already pointed out, the function V_3 is 2-Lipschitz in $L^2([0, T] \times \Omega)$ by Hardy's inequality [Har20] reminded below. Therefore, the parameter c_0 is chosen to be strictly bigger than 2. One shall notice that in this case, V_3 at time t does not only depend on ρ_t , but on the whole history. From modeling point of view we consider that agents pay attention to the past experience in estimating the running cost at (t, x) .

Lemma 5.3.1. (Hardy's inequality) *If $f: [0, +\infty[\rightarrow [0, +\infty[$ is a measurable function, then*

$$\int_0^\infty \left(\frac{1}{x} \int_0^x f(t) dt \right)^p dx \leq \left(\frac{p}{p-1} \right)^p \int_0^\infty f(x)^p dx.$$

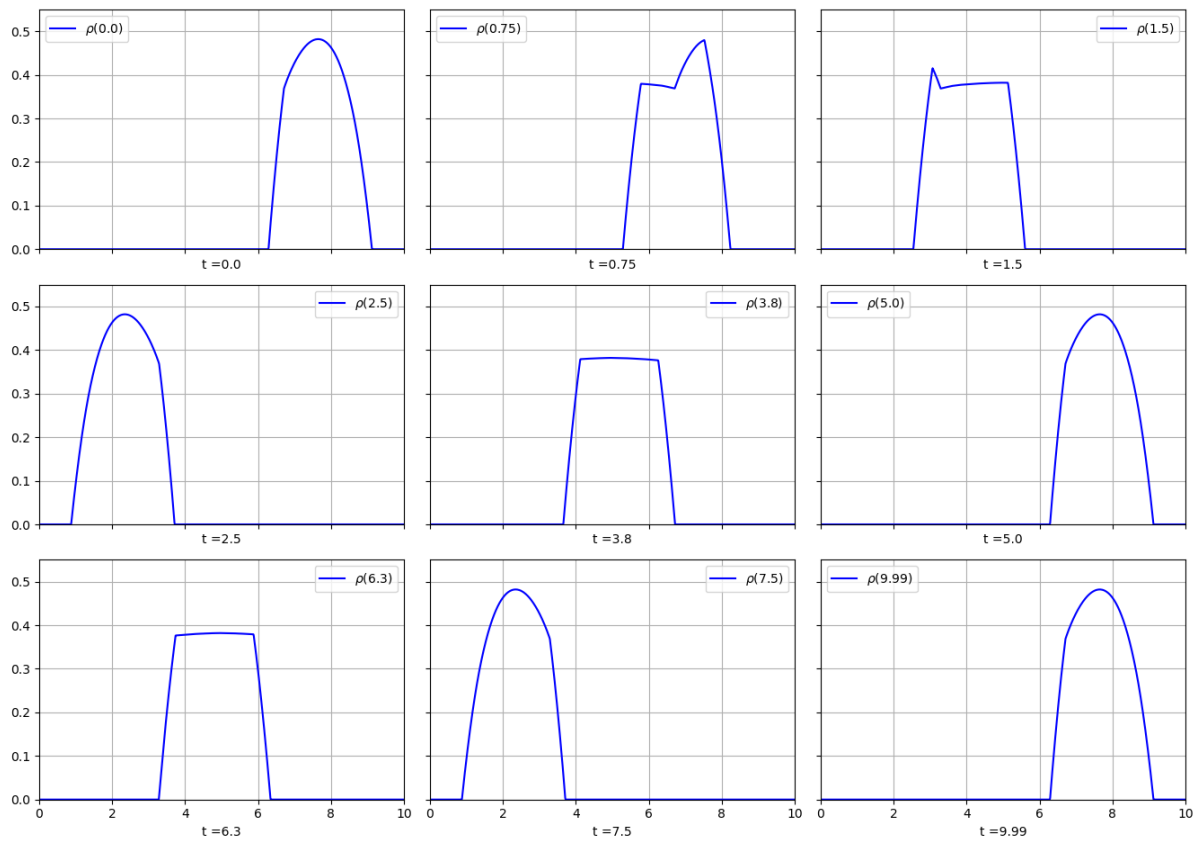


Figure 5.7: Simulation of the solution $\bar{\rho}$ minimizing (5.21) at times 0, 0.75, 1.5, 2.5, 3.8, 5, 6.3, 7.5, 9.99 for $V(\rho) \equiv V_1(\rho) + V_0$ with V_0 defined in (5.22) and parameters from Table 5.3 and $\delta = \frac{1}{2}$.

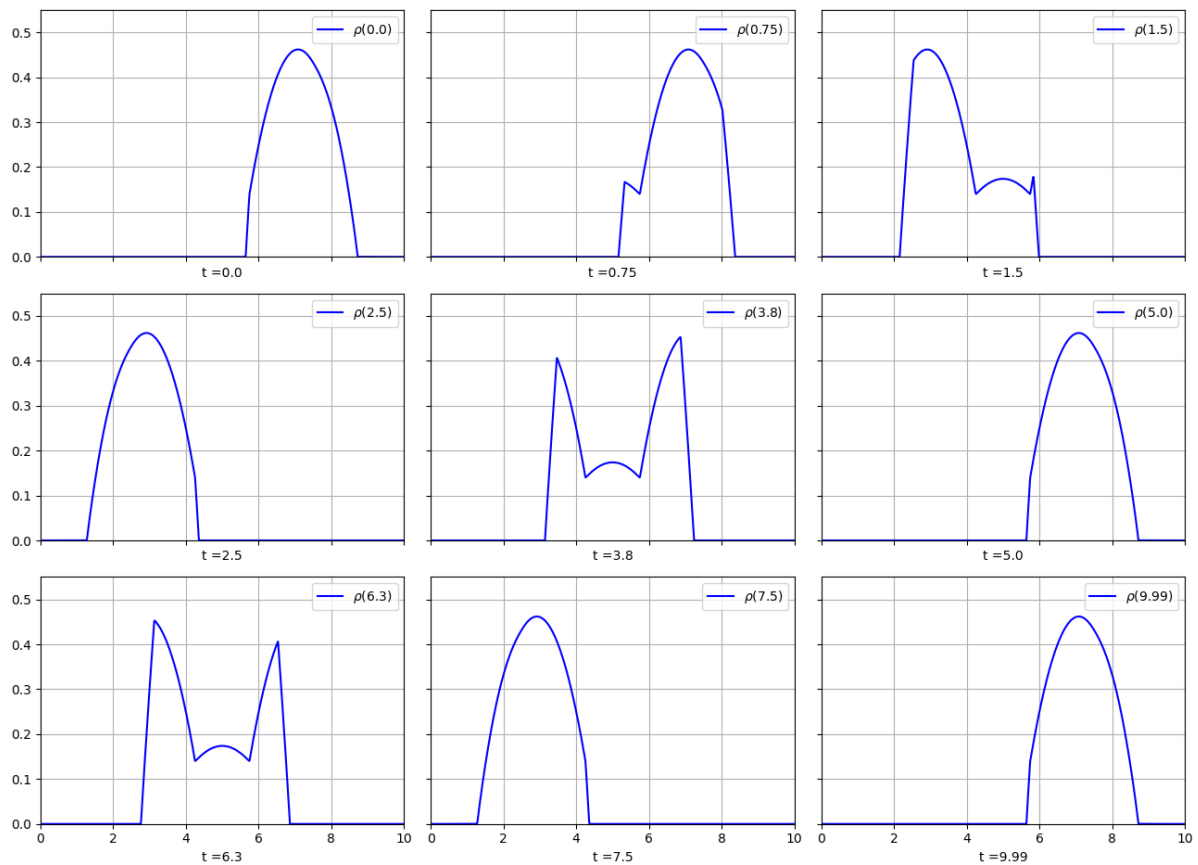


Figure 5.8: Simulation of the solution $\bar{\rho}$ minimizing (5.21) at times 0, 0.75, 1.5, 2.5, 3.8, 5, 6.3, 7.5, 9.99 for $V(\rho) \equiv V_2(\rho) + V_0$ with V_0 defined in (5.22) and parameters from Table 5.3 except for parameter $\lambda = 1$.

Proposition 5.3.2. *The function*

$$\begin{aligned} V: \quad L^2([0, T] \times \Omega) &\rightarrow L^2([0, T] \times \Omega) \\ \rho &\mapsto \left(t \mapsto \int_0^t \rho(s, \cdot) ds + V_0(t, \cdot) \right) \end{aligned}$$

is 2-Lipschitz.

Proof. For all $\rho_1, \rho_2 \in L^2([0, T] \times \Omega)$, we have by Fubini and Hardy's inequality

$$\begin{aligned} \|V(\rho_1) - V(\rho_2)\|_{L^2}^2 &= \int_0^T \int_{\Omega} \left(\frac{1}{t} \int_0^t \rho_1(s, x) - \rho_2(s, x) ds \right)^2 dx dt \\ &\leq 2^2 \int_0^T \int_{\Omega} (\rho_1(t, x) - \rho_2(t, x))^2 dx = 2^2 \|\rho_1 - \rho_2\|_{L^2}^2, \end{aligned}$$

hence V is 2-Lipschitz. □

The profile of the fixed point ρ of \mathcal{F} is displayed in Figure 5.9 at different times t . One can notice first that at each time the density is concentrated near the minimum of $V_0(t)$, i.e. $\frac{\pi}{2} + 3 \cos\left(\frac{4\pi}{T}t\right)$, which moves from the right to the left and vice versa periodically.

Second, the effect of the term $\int_0^t \rho(s, x) ds$ can be observed for example at time $t = 0.75$ where the density is reduced at position for example $x = 7$ and a portion of the density directly go to $x = 6$. In fact, the term $\int_0^t \rho(s, x) ds$ keeps in memory a mean in time of the population who visited position x and a player may not visit a place which was too crowded in the past. In addition, the cost λ for jumping may affect the decision of the players which causes the population to anticipate and go to less crowded places.

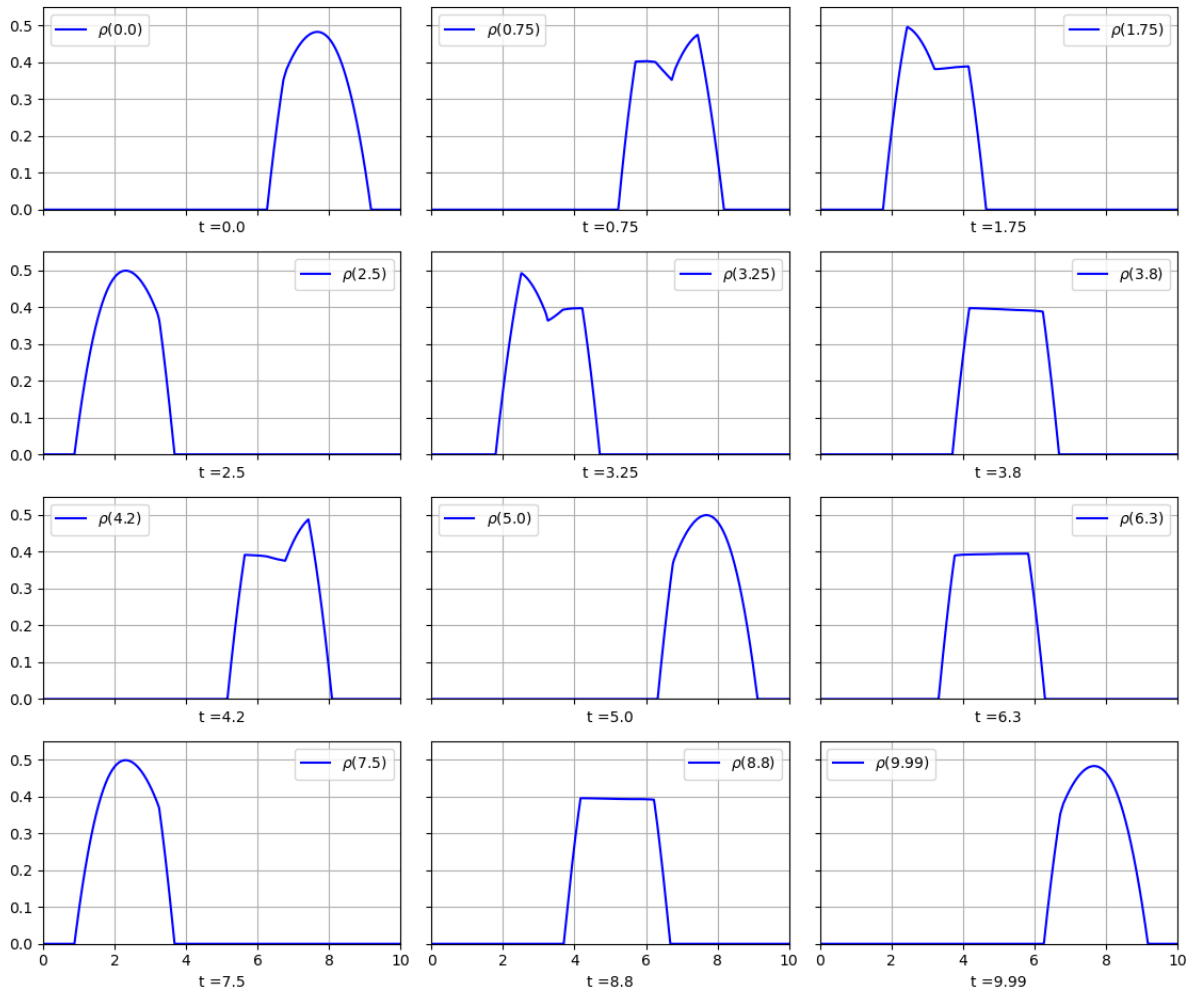


Figure 5.9: Simulation of the solution $\bar{\rho}$ minimizing (5.21) at times $t = 0, 0.75, 1.75, 2.5, 3.25, 3.8, 4.2, 5, 6.3, 7.5, 8.8, 9.99$ for $V(\rho) \equiv V_3(\rho) + V_0$ with V_0 defined in (5.22) and parameters from Table 5.3.

Chapter 6

Multi-species Modeling

In this chapter, we present two models inspired by the MFG problem we already studied in the form

$$\min_{\substack{\rho \in L^2, \rho \geq 0 \\ \forall t, \int_{\Omega} \rho(t) dx = 1}} \mathcal{V}_{\bar{\rho}}(\rho) := \int_0^T \int_{\Omega} \lambda |\dot{\rho}| + V(\bar{\rho})\rho + f(\rho) dxdt + \psi_0(\rho(0)) + \psi_T(\rho(T)). \quad (6.1)$$

The first example in Section 6.1 is a model of two populations interacting with each other through a congestion term. Numerical simulations has been carried out in this context.

The second example in Section 6.2 is a presentation of a model of two populations (firms and workers) inspired by a model by Carlier and Ekeland [CE07]. Three possibilities of numerical resolutions of the problem are presented.

6.1 A two-population model

Mean Field Games with two populations have been studied for instance by Achdou, Bardi and Cirant in [ABC17] in the context of urban settlements motivated by the segregation model by Schelling [Sch71]. Amongst other contributions in this article, they show existence of a solution to a system of four PDEs associated with the two-population MFG and present numerical simulations in the static and evolutive cases. The individual cost is the one inspired by Schelling's model which is a function of the proportion of the neighborhood's population whose the player belongs to. In our example, we shall not use the proportion of the population, but a quadratic term that depends on the other population.

Here, we consider two different populations ρ_1 and ρ_2 such that the equilibrium is a couple of measures $(\bar{\rho}_1, \bar{\rho}_2)$ which is a fixed point of

$$\mathcal{G}: L^2([0, T] \times \Omega) \times L^2([0, T] \times \Omega) \rightarrow L^2([0, T] \times \Omega) \times L^2([0, T] \times \Omega) \quad (6.2)$$

$$(\bar{\rho}_1, \bar{\rho}_2) \mapsto \left(\begin{array}{cc} \operatorname{argmin}_{\substack{\rho_1 \geq 0 \\ \forall t, \int \rho_1(t) = \frac{1}{2}}} \mathcal{V}_{\bar{\rho}_2}^1(\rho_1); & \operatorname{argmin}_{\substack{\rho_2 \geq 0 \\ \forall t, \int \rho_2(t) = \frac{1}{2}}} \mathcal{V}_{\bar{\rho}_1}^2(\rho_2) \end{array} \right)$$

where

$$\mathcal{V}_{\bar{\rho}}^1(\rho) := \int_0^T \int_{\Omega} \lambda |\dot{\rho}| + V_1(\bar{\rho})\rho + f(\rho) dxdt + \psi_0(\rho(0)) + \psi_T(\rho(T))$$

and

$$\mathcal{V}_{\bar{\rho}}^2(\rho) := \int_0^T \int_{\Omega} \lambda |\dot{\rho}| + V_2(\bar{\rho})\rho + f(\rho) dxdt + \psi_0(\rho(0)) + \psi_T(\rho(T))$$

with

$$V_1(\rho_2) = \alpha \rho_2^2 + V_{0,1} \text{ and } V_2(\rho_1) = \beta \rho_1^2 + V_{0,2}.$$

The penalizations ψ_0 and ψ_T and the coefficient λ can be different between $\mathcal{V}_{\bar{\rho}}^1$ and $\mathcal{V}_{\bar{\rho}}^2$.

Once V_1 and V_2 are computed according to some measures $\bar{\rho}_1$ and $\bar{\rho}_2$, the optimization problems in (6.2) become separate and it is possible compute the optimal ρ_1 and ρ_2 separately. This problem is inspired from Section 2.2 so that under suitable conditions on V_1 and V_2 , the fixed point to \mathcal{G} exists, since the product of two complete spaces is complete.

Because of Proposition 2.2.7, V_1 and V_2 should include a part with explicit dependence on t in order to obtain a dynamic solution instead of a constant in time solution, hence the presence of the terms $V_{0,1}$

and $V_{0,2}$. In this section, we define

$$V_{0,1}(t, x) = \left(\frac{S}{2} + 3 \cos \left(\frac{2\pi}{T} t \right) - x \right)^2 \quad \text{and} \quad V_{0,2}(t, x) = \left(\frac{S}{2} - 3 \cos \left(\frac{2\pi}{T} t \right) - x \right)^2, \quad (6.3)$$

such that at time t , it is more interesting to live near position $\frac{S}{2} + 3 \cos \left(\frac{2\pi}{T} t \right)$ for the first group ρ_1 and near $\frac{S}{2} - 3 \cos \left(\frac{2\pi}{T} t \right)$ for the second group ρ_2 .

One may notice that the preference of ρ_1 moves from the right to the left between positions $x = \frac{S}{2} + 3 = 8$ and $x = \frac{S}{2} - 3 = 2$ and after it moves from $x = 2$ to $x = 8$. The preference of ρ_2 is similar, but it starts from $x = 2$. At times $t = 2.5, 7.5$, the preferences are the same for both populations. However, due to the term V_1 and V_2 , both population are less willing to share the same position, so we expect that they will repel each other. As a comparison, we simulate the profile of a single-population called ρ_0 which solves (6.1) with the same preference as ρ_1 (i.e. $V(\rho) = V_{0,1}$) and with mass constraint $\int \rho_t dx = \frac{1}{2}$, but without the term $\alpha \rho_2^2$ to suppress the interaction with another population.

The simulations are performed with the method presented in Section 5.3 with parameters from Table 6.1. We also choose $f(\rho) = c_0 \frac{\rho^2}{2}$.

The profile of the equilibrium (ρ_1, ρ_2) is displayed in Figure 6.1. The blue solid line (or the density that starts near position $x = 8$) corresponds to the density ρ_1 while the red solid line (or the density that starts near position $x = 2$) is the density of ρ_2 . The green dashed line represents the minimizer ρ_0 of \mathcal{V} when $V(\bar{\rho}) = V_{0,1}$, namely V does not depend on the density.

Similarly to Figures 5.7, 5.8 and 5.9, at each time t the densities ρ_1 and ρ_2 concentrate near the minimum of $V_{0,1}(t)$ and $V_{0,2}(t)$ respectively. When the two populations are far from each other, their behavior does not vary from single population's ρ_0 . Indeed, at times $t = 0, 5, 9.98$, the profile of ρ_1 is combined with the profile of ρ_0 . At other times in Figure 6.1, the dynamic of ρ_1 differs from ρ_0 as soon as the density gets closer to the second population. For example, at time $t = 2.4$ and at position $x = 5$, the population ρ_1 is reduced compared to ρ_0 . Also, the cost of jumping may cause the population to directly relocate in places less concentrated by the other population for example at time $t = 1.5$ near position $x = 6$ or at time $t = 3.3$ near position $x = 4$.

Parameter	Value
T	10
S	10
K	500
N	500
a	200
h	0.02
L	$6/h$
λ	0.5
c_0	10
α	10
β	10

Table 6.1: Parameters for the numerical simulation of the fixed point of (6.2).

6.2 Land Use Between Firms and Workers

The last example in this thesis is inspired from the article by Carlier and Ekeland [CE07] in which they describe the equilibrium of the structure of the city by using optimal transport theory. We only develop the modeling here. The simulations are left for further studies.

We consider two populations in this model: firms and workers. The revenue of an individual living at x will be denoted $\varphi(x)$ and the wage paid by the firms located at y will be $\psi(y)$. If $c(x, y)$ is the commuting cost of the worker living at x who go to work at y , then the worker would like to maximize his income according to the wage and the transportation cost:

$$\varphi(x) = \sup_y \psi(y) - c(x, y). \quad (6.4)$$

Similarly, firms seek to minimize the wage they pay which corresponds to the problem:

$$\psi(y) = \inf_x \varphi(x) + c(x, y). \quad (6.5)$$

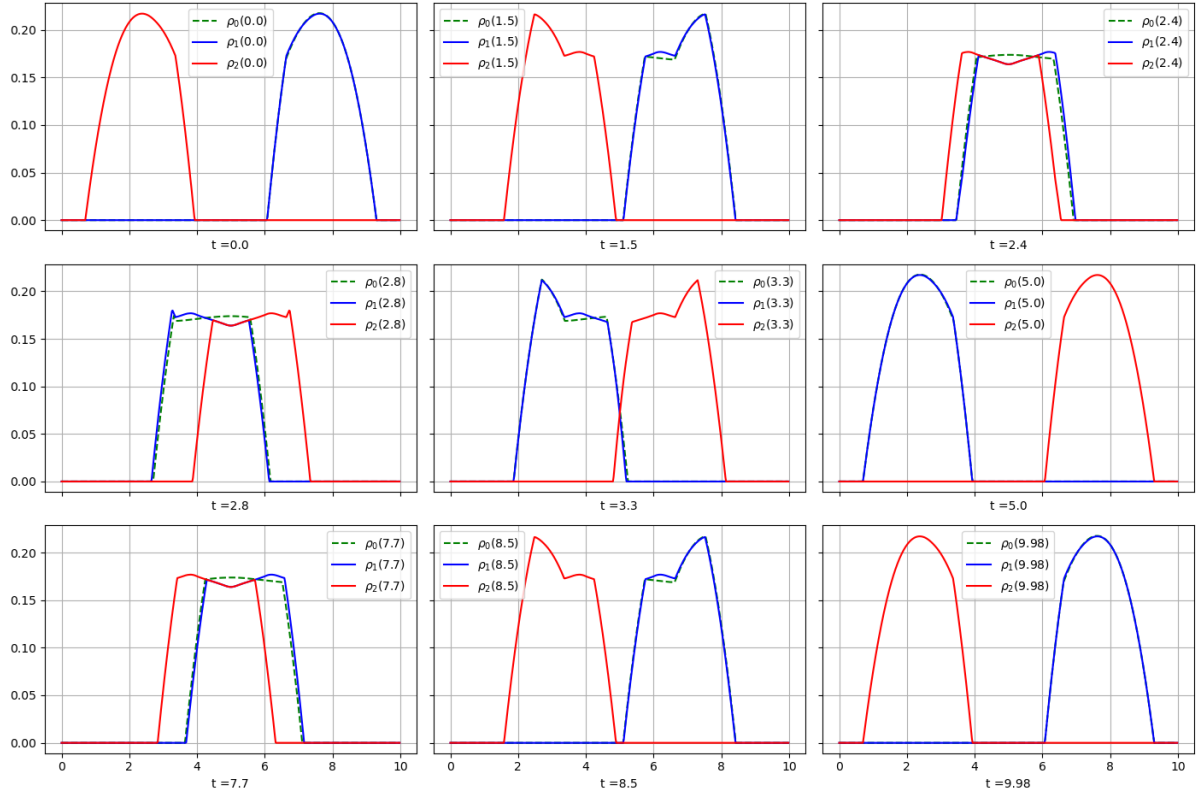


Figure 6.1: Simulation of the fixed point to (6.2) at times $t = 0, 1.5, 2.4, 2.8, 3.3, 5, 7.7, 8.5, 9.98$ for $V_1(\rho_2) = \alpha\rho_2^2 + V_{0,1}$ and $V_2(\rho_1) = \beta\rho_1^2 + V_{0,2}$ with $V_{0,i}$ defined in (6.3) and parameters from Table 6.1. The blue solid line is the profile of the population ρ_1 , the red solid line to the population ρ_2 and the green dashed line to the solution ρ_0 when $V(\rho) = V_{0,1}$.

If φ and ψ verify (6.4) and (6.5), then they are said to be *conjugate* to each other.

Let μ designate the density of the workers and ν the density of the firms. For a given cost c , Carlier and Ekeland show the existence of an equilibrium $(\mu, \nu, \varphi, \psi, \gamma)$ such that φ and ψ are conjugate to each other for the commuting cost c and γ is a probability measure on $\Omega \times \Omega$ with marginals μ and ν which verifies

$$\psi(y) - \varphi(x) = c(x, y), \quad \gamma - a.e.$$

Thanks to optimal transport theory, we know (see [Vil03]) that in particular this is equivalent to γ solving the Monge-Kantorovitch problem

$$\inf_{\gamma \in \Pi(\mu, \nu)} \int_{\Omega \times \Omega} c(x, y) d\gamma(x, y) \quad (\mathcal{M})$$

and (φ, ψ) solving

$$\sup_{\varphi, \psi} \left\{ \int_{\bar{\Omega}} \psi d\nu - \int_{\bar{\Omega}} \varphi d\mu ; \psi(y) - \varphi(x) \leq c(x, y), \forall (x, y) \in \bar{\Omega} \times \bar{\Omega} \right\}. \quad (\mathcal{D})$$

In addition, these two problems are dual to each other in the sense that $(\mathcal{M}) = (\mathcal{D})$. Under some conditions, the solution to (\mathcal{D}) is unique up to an additive constant which is added to ψ and φ due to the condition $\psi(y) - \varphi(x) \leq c(x, y)$. We will call φ and ψ the *Kantorovitch potentials*.

In our model, we consider that each worker tries to minimize

$$\min_{\gamma(0)=x_0} 2\lambda_1 S(\gamma) + \int_0^T [-\varphi_{\mu, \nu}(\gamma(t)) + V_{0,1}(t, \gamma(t)) + f'(\mu)(t, \gamma(t))] dt + \phi_T(\gamma(T))$$

and each firm would like to minimize

$$\min_{\gamma(0)=x_0} 2\lambda_2 S(\gamma) + \int_0^T [\psi_{\mu, \nu}(\gamma(t)) + V_{0,2}(t, \gamma(t)) + f'(\nu)(t, \gamma(t))] dt + \phi_T(\gamma(T)).$$

The parameters λ_1 and λ_2 describe the cost of moving which shall not necessarily be the same for workers and firms.

In this way, the equilibrium (μ, ν) would be the solution to the problem

$$\begin{aligned} \min_{\substack{\mu, \nu \geq 0 \\ \int \mu_t = \int \nu_t = \frac{1}{2}}} \int_0^T \int_{\Omega} \lambda_1 |\dot{\mu}| + \lambda_2 |\dot{\nu}| + V_{0,1}\mu + V_{0,2}\nu + f(\nu) + f(\mu) dx dt + \int_0^T W_c(\mu_t, \nu_t) dt \\ + \Phi_0(\mu(0)) + \Phi_0(\nu(0)) + \Phi_T(\mu(T)) + \Phi_T(\nu(T)) \end{aligned} \quad (6.6)$$

where $W_c(\mu, \nu) = \inf_{\gamma \in \Pi(\mu, \nu)} \int_{\bar{\Omega} \times \bar{\Omega}} c(x, y) d\gamma(x, y)$ by using the fact that (\mathcal{M}) and (\mathcal{D}) are dual to each other. Indeed, The Kantorovitch potentials $-\varphi_{\mu, \nu}$ and $\psi_{\mu, \nu}$ are also the first variations of the transport cost W_c (see [San15, Section 7.2.2]).

We will present three different approaches to solve numerically (6.6). In all these cases, it will be more convenient to consider the regularized Wasserstein distance such that

$$\begin{aligned} W_{c, \varepsilon}(\mu, \nu) &= \inf_{\gamma \in \Pi(\mu, \nu)} \int_{\Omega \times \Omega} c(x, y) d\gamma(x, y) - \varepsilon \int \gamma(\log \gamma - 1) \\ &= \sup_{\varphi, \psi} \int_{\Omega} \psi(y) d\nu(y) - \int_{\Omega} \varphi(x) d\mu(x) - \varepsilon \int_{\Omega \times \Omega} e^{\frac{\psi(y) - \varphi(x) - c(x, y)}{\varepsilon}} d(\mu \otimes \nu)(x, y), \end{aligned} \quad (6.7)$$

because it is easier to compute the Kantorovitch potentials thanks to the Sinkhorn algorithm [Cut13]. In addition, this formulation eliminates the constraint $\psi(y) - \varphi(x) \leq c(x, y)$ compared to (\mathcal{D}) .

(i) *The mixed approach.*

In this first approach, we set aside the variational formulation (6.6) of the problem and see (μ, ν) as the solution of a mixed variational MFG in the sense that it is the fixed point of the application \mathcal{G} defined by (6.2) in the two-population example with

$$\begin{aligned} \mathcal{V}_{\bar{\mu}, \bar{\nu}}^1(\mu) &:= \int_0^T \int_{\Omega} \lambda_1 |\dot{\mu}| - \varphi_{\bar{\mu}, \bar{\nu}} \mu + f(\mu) dx dt + \Phi_0(\mu(0)) + \Phi_T(\mu(T)) \\ \text{and } \mathcal{V}_{\bar{\mu}, \bar{\nu}}^2(\nu) &:= \int_0^T \int_{\Omega} \lambda_2 |\dot{\nu}| + \psi_{\bar{\mu}, \bar{\nu}} \nu + f(\nu) dx dt + \Phi_0(\nu(0)) + \Phi_T(\nu(T)). \end{aligned}$$

The numerical simulations were carried out in this example such that at iteration k , from (μ^k, ν^k) we compute φ_{μ^k, ν^k} and ψ_{μ^k, ν^k} thanks to the Sinkhorn algorithm [Cut13] and then we apply the FDPG algorithm to obtain $(\mu^{k+1}, \nu^{k+1}) = \mathcal{G}(\mu^k, \nu^k)$. We also defined $f(\rho) = c_0 \frac{\rho^2}{2}$ as usual.

However, some tentative numerical computations did not converge even for very large c_0 (e.g. $c_0 = 10000$) which was expected to ensure \mathcal{G} to be a contraction. This raises the question of the dependency of the Kantorovitch potentials $\varphi_{\bar{\mu}, \bar{\nu}}$ and $\psi_{\bar{\mu}, \bar{\nu}}$ according to the densities $\bar{\mu}$ and $\bar{\nu}$. If the dependency were Lipschitz continuous, then we would be able to provide a parameter c_0 so that the sequence $(\mathcal{G}^n(\mu^0, \nu^0))_n$ converges towards a solution. On the other hand, the Kantorovitch potentials are known to depend continuously on the measures, but the fact that this dependency is not Lipschitz continuous is an open question (which, by the way, could be explored numerically).

To our knowledge, we need to add extra regularization of the transport cost to obtain some regular dependency. For example, the entropic regularization described in (6.7) adds some robustness of the solution according to the data (see [Cut13]). I refer to [PC19] for a review of the literature on the regularization of optimal transport. They themselves propose an equivalent definition of the 2-Wasserstein distance which is more robust.

(ii) *The variational approach through the FDPG method.*

Differently from the previous approach, the fixed point strategy may not be necessary if one directly applies the FDPG method to the variational formulation. We consider (μ, ν) to be a minimizer

$$\begin{aligned} \min_{\substack{\mu, \nu \geq 0 \\ \int \mu_t = \int \nu_t = \frac{1}{2}}} \int_0^T \int_{\Omega} \lambda_1 |\dot{\mu}| + \lambda_2 |\dot{\nu}| + V_{0,1}\mu + V_{0,2}\nu + f(\nu) + f(\mu) dx dt + \int_0^T W_{c, \varepsilon}(\mu_t, \nu_t) dt \\ + \Phi_0(\mu(0)) + \Phi_0(\nu(0)) + \Phi_T(\mu(T)) + \Phi_T(\nu(T)). \end{aligned} \quad (6.8)$$

To explain the method, we take $f(\rho) = \frac{\rho^2}{2}$ and $\Phi_0 = \Phi_T = 0$. One can view the problem (6.8) as

$$\min_{\mu, \nu \in \mathbb{R}^{K \cdot N}} h(\mu, \nu) + g(\mathcal{A}(\mu, \nu)), \quad (6.9)$$

where

$$h(\mu, \nu) = \int_0^T \int_{\Omega} V_{0,1}\mu + V_{0,2}\nu + \frac{\mu^2}{2} + \frac{\nu^2}{2} dxdt + \int_0^T W_{c,\varepsilon}(\mu_t, \nu_t) dt$$

and $g(\rho_0, \rho_1, \rho_2) = \lambda \|\rho_0\|_1 + \delta_{C_0}(\rho_1) + \delta_{C_1}(\rho_2)$

where C_0 and C_1 are defined in Section 5.2.

To use the FDPG method, we need to compute the conjugate h^* in the sense of the definition (5.9). Let us write $h(\mu, \nu) = h_1(\mu, \nu) + h_2(\mu, \nu)$ such that

$$h_1(\mu, \nu) = \int_0^T \int_{\Omega} V_{0,1}\mu + V_{0,2}\nu + \frac{\mu^2}{2} + \frac{\nu^2}{2} dxdt \text{ and } h_2(\mu, \nu) = \int_0^T W_{c,\varepsilon}(\mu_t, \nu_t) dt.$$

The property described below states that the conjugate of a sum is the *infimum convolution* of the conjugate functions.

Proposition 6.2.1 (Conjugate of the sum). *Let $h_1: \mathbb{R}^d \rightarrow (-\infty, \infty]$ be a proper convex function and $h_2: \mathbb{R}^d \rightarrow \mathbb{R}$ be a convex function. Then*

$$(h_1 + h_2)^*(x) = \min_{u+v=x} \{h_1^*(u) + h_2^*(v)\}.$$

Proof. See [Bec17, Theorem 4.17]. □

Consequently, the conjugate of h is the following:

$$h^*(\eta^1, \eta^2) = (h_1 + h_2)^*(\eta_1, \eta_2) = \min_{\substack{u_1+v_1=\eta_1 \\ u_2+v_2=\eta_2}} \{h_1^*(u_1, u_2) + h_2^*(v_1, v_2)\}.$$

An explicit formula for h_1^* gives

$$\begin{aligned} h_1^*(u_1, u_2) &= \sup_{\mu, \nu} \int_0^T \int_{\Omega} u_1\mu + u_2\nu - V_{0,1}\mu - V_{0,2}\nu - \frac{\mu^2}{2} - \frac{\nu^2}{2} dxdt \\ &= \int_0^T \int_{\Omega} \frac{(u_1 - V_{0,1})^2}{2} + \frac{(u_2 - V_{0,2})^2}{2} dxdt, \end{aligned}$$

which yields

$$\begin{aligned} h^*(\eta^1, \eta^2) &= \min_{\substack{u_1+v_1=\eta_1 \\ u_2+v_2=\eta_2}} \int_0^T \int_{\Omega} \frac{(u_1 - V_{0,1})^2}{2} + \frac{(u_2 - V_{0,2})^2}{2} + \sup_{\mu, \nu} \int_0^T \int_{\Omega} v_1\mu + v_2\nu \\ &\quad - \int_0^T \left[\sup_{\varphi, \psi} \int_{\Omega} \psi \nu_t - \varphi \mu_t - \varepsilon \int_{\Omega \times \Omega} e^{\frac{\psi(y) - \varphi(x) - c(x,y)}{\varepsilon}} d(\mu_t \otimes \nu_t)(x, y) \right] dt. \end{aligned}$$

The latter may be challenging to compute.

Another approach to apply the FDPG method would be, instead of seeing $W_{c,\varepsilon}$ as a smooth convex function, we can view it as a non-smooth convex function so that it can be included in $g(\mathcal{A}(\mu, \nu))$ in (6.9). Now the functions h and g are

$$h(\mu, \nu) = \int_0^T \int_{\Omega} V_{0,1}\mu + V_{0,2}\nu + \frac{\mu^2}{2} + \frac{\nu^2}{2} dxdt$$

and $g(\rho_0, \rho_1, \rho_2, \rho_3, \rho_4) = \lambda \|\rho_0\|_1 + \delta_{C_0}(\rho_1) + \delta_{C_1}(\rho_2) + W_{c,\varepsilon}(\rho_3, \rho_4)$.

The difficulty here is that g is not the sum of functions with separate variables due to the Wasserstein term. A way to compute the proximal operator of $h(\mu, \nu) := W_{c,\varepsilon}(\mu, \nu)$ is

$$\begin{aligned} \text{prox}_h(u_1, u_2) &= \text{argmin}_{\mu, \nu} W_{c,\varepsilon}(\mu, \nu) + \|\mu - u_1\|^2 + \|\nu - u_2\|^2 \\ &= \text{argmin}_{\mu, \nu} \left[\max_{\varphi, \psi} \int_{\Omega} \psi d\nu - \int_{\Omega} \varphi d\mu - \varepsilon \int_{\Omega \times \Omega} e^{\frac{\psi(y) - \varphi(x) - c(x,y)}{\varepsilon}} d(\mu \otimes \nu)(x, y) \right] \\ &\quad + \|\mu - u_1\|^2 + \|\nu - u_2\|^2. \end{aligned}$$

This formulation already shows that μ and ν are linked through the tensor product and the exponential.

The problem in $\text{prox}_h(u_1, u_2)$ can also be viewed as the conjugate of a certain function:

$$\begin{aligned}
& \min_{\mu, \nu} \left[\max_{\varphi, \psi} \int_{\Omega} \psi d\nu - \int_{\Omega} \varphi d\mu - \varepsilon \int_{\Omega \times \Omega} e^{\frac{\varphi(x) + \psi(y) - c(x, y)}{\varepsilon}} d(\mu \otimes \nu)(x, y) \right] + \|\mu - u_1\|^2 + \|\nu - u_2\|^2 \\
&= \max_{\mu, \nu} \left[\min_{\varphi, \psi} - \int_{\Omega} \psi d\nu + \int_{\Omega} \varphi d\mu + \varepsilon \int_{\Omega \times \Omega} e^{\frac{\psi(y) - \varphi(x) - c(x, y)}{\varepsilon}} d(\mu \otimes \nu)(x, y) \right] - \|\mu - u_1\|^2 - \|\nu - u_2\|^2 \\
&= \max_{\mu, \nu} 2 \langle \mu | u_1 \rangle + 2 \langle \nu | u_2 \rangle - \|\mu\|^2 - \|\nu\|^2 + \left[\min_{\varphi, \psi} - \int_{\Omega} \psi d\nu + \int_{\Omega} \varphi d\mu + \varepsilon \int_{\Omega \times \Omega} e^{\frac{\psi(y) - \varphi(x) - c(x, y)}{\varepsilon}} d(\mu \otimes \nu)(x, y) \right] \\
&= \max_{\mu, \nu} 2 \langle \mu | u_1 \rangle + 2 \langle \nu | u_2 \rangle - \|\mu\|^2 - \|\nu\|^2 + \left[\max_{\varphi, \psi} \int_{\Omega} \psi d\nu - \int_{\Omega} \varphi d\mu - \varepsilon \int_{\Omega \times \Omega} e^{\frac{\psi(y) - \varphi(x) - c(x, y)}{\varepsilon}} d(\mu \otimes \nu)(x, y) \right] \\
&:= H^*(u_1, u_2),
\end{aligned}$$

where $H(\mu, \nu) = \left(\|\mu\|^2 + \|\nu\|^2 + \left[\max_{\varphi, \psi} \int_{\Omega} \psi d\nu - \int_{\Omega} \varphi d\mu - \varepsilon \int_{\Omega \times \Omega} \exp\left(\frac{\psi(y) - \varphi(x) - c(x, y)}{\varepsilon}\right) d(\mu \otimes \nu)(x, y) \right] \right) / 2$.

(iii) *The variational approach through the Uzawa algorithm.*

To introduce this method, we reformulate the problem (6.6) in a min-max form. Again, we take $f(\rho) = \frac{\rho^2}{2}$ and $\Phi_0 = \Phi_T = 0$. We have

$$\begin{aligned}
& \min_{\substack{\mu, \nu \geq 0 \\ \int \mu_t = \int \nu_t = \frac{1}{2}}} \int_0^T \int_{\Omega} \lambda_1 |\dot{\mu}| + \lambda_2 |\dot{\nu}| + V_{0,1}\mu + V_{0,2}\nu + \frac{\mu^2}{2} + \frac{\nu^2}{2} dxdt + \int_0^T W_{c,\varepsilon}(\mu_t, \nu_t) dt \\
&= \min_{\substack{\mu, \nu \geq 0 \\ \int \mu_t = \int \nu_t = \frac{1}{2}}} \int_0^T \int_{\Omega} \lambda_1 |\dot{\mu}| + \lambda_2 |\dot{\nu}| + V_{0,1}\mu + V_{0,2}\nu + \frac{\mu^2}{2} + \frac{\nu^2}{2} dxdt \\
&\quad + \int_0^T \left[\sup_{\varphi_t, \psi_t} \int_{\Omega} \psi_t(y) d\nu_t(y) - \int_{\Omega} \varphi_t(x) d\mu_t(x) - \varepsilon \int_{\Omega \times \Omega} e^{\frac{\psi_t(y) - \varphi_t(x) - c(x, y)}{\varepsilon}} d(\mu_t \otimes \nu_t)(x, y) \right] dt \\
&= \min_{\substack{\mu, \nu \geq 0 \\ \int \mu_t = \int \nu_t = \frac{1}{2}}} \sup_{\varphi, \psi} \int_0^T \int_{\Omega} \lambda_1 |\dot{\mu}| + \lambda_2 |\dot{\nu}| + V_{0,1}\mu + V_{0,2}\nu + \frac{\mu^2}{2} + \frac{\nu^2}{2} dxdt \\
&\quad + \int_0^T \left[\int_{\Omega} \psi_t(y) d\nu_t(y) - \int_{\Omega} \varphi_t(x) d\mu_t(x) - \varepsilon \int_{\Omega \times \Omega} e^{\frac{\psi_t(y) - \varphi_t(x) - c(x, y)}{\varepsilon}} d(\mu_t \otimes \nu_t)(x, y) \right] dt \quad (6.10)
\end{aligned}$$

where $\varphi: [0, T] \rightarrow \text{Lip}(\Omega)$ and $\psi: [0, T] \rightarrow \text{Lip}(\Omega)$ have to be understood as functions which associate t with the Kantorovitch potentials respectively φ_t and ψ_t obtained from the densities μ_t and ν_t .

If there exists an equilibrium $(\mu, \nu, \varphi, \psi)$ which is a saddle-point of a Lagrangian associated to (6.10), then a possible idea would be to apply Uzawa's algorithm [Uza89]. Even if this method has not been explored yet in this thesis by lack of time, we observe that we have all ingredients to do it. In addition, this method is advantageous, because it does not require to use the Sinkhorn algorithm to compute φ_t and ψ_t . Indeed, let us denote the problem (6.10) by

$$\inf_{\mu, \nu} \sup_{\varphi, \psi} \mathcal{L}(\mu, \nu, \varphi, \psi).$$

At iteration k , if (φ^k, ψ^k) is given and (μ^k, ν^k) is such that

$$\mathcal{L}(\mu^k, \nu^k, \varphi^k, \psi^k) = \inf_{\mu, \nu} \mathcal{L}(\mu, \nu, \varphi^k, \psi^k),$$

(which exactly requires an optimization method as in Chapter 5), then the next iteration is computed through the gradient method

$$(\varphi^{k+1}, \psi^{k+1}) = (\varphi^k, \psi^k) + t^k \nabla_{\varphi, \psi} \mathcal{L}(\mu^k, \nu^k, \varphi^k, \psi^k)$$

for some sequence $(t^k)_k$ and where

$$\nabla_{\varphi, \psi} \mathcal{L}(\mu, \nu, \varphi, \psi)(t, x) = \begin{pmatrix} \mu_t(x) \left(1 + \int_{\Omega} e^{\frac{\psi_t(y) - \varphi_t(x) - c(x, y)}{\varepsilon}} d\nu_t(y) \right) \\ \nu_t(x) \left(1 - \int_{\Omega} e^{\frac{\psi_t(x) - \varphi_t(y) - c(y, x)}{\varepsilon}} d\mu_t(y) \right) \end{pmatrix}.$$

Next, (μ^{k+1}, ν^{k+1}) is obtained by

$$\mathcal{L}(\mu^{k+1}, \nu^{k+1}, \varphi^{k+1}, \psi^{k+1}) = \inf_{\mu, \nu} \mathcal{L}(\mu, \nu, \varphi^{k+1}, \psi^{k+1}).$$

Bibliography

- [ABC17] Yves Achdou, Martino Bardi, and Marco Cirant. “Mean field games models of segregation”. In: *Mathematical Models and Methods in Applied Sciences* 27.01 (Jan. 2017). Publisher: World Scientific Publishing Co., pp. 75–113.
- [AL16] Yves Achdou and Mathieu Laurière. “Mean Field Type Control with Congestion (II): An Augmented Lagrangian Method”. en. In: *Applied Mathematics & Optimization* 74.3 (Dec. 2016), pp. 535–578.
- [AF09] Luigi Ambrosio and Alessio Figalli. “Geodesics in the Space of Measure-Preserving Maps and Plans”. en. In: *Archive for Rational Mechanics and Analysis* 194.2 (Nov. 2009), pp. 421–462.
- [AGS05] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient Flows*. en. Lectures in Mathematics ETH Zürich. Basel: Birkhäuser-Verlag, 2005.
- [And17] Roman Andreev. “Preconditioning the Augmented Lagrangian Method for Instationary Mean Field Games with Diffusion”. In: *SIAM Journal on Scientific Computing* 39.6 (Jan. 2017). Publisher: Society for Industrial and Applied Mathematics, A2763–A2783.
- [BFG22] Tigran Bakaryan, Rita Ferreira, and Diogo Gomes. “A potential approach for planning mean-field games in one dimension”. en. In: *Communications on Pure and Applied Analysis* 21.6 (May 2022). Publisher: Communications on Pure and Applied Analysis, pp. 2147–2187.
- [BC97] Martino Bardi and Italo Capuzzo-Dolcetta. *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*. en. Boston, MA: Birkhäuser, 1997.
- [Bar85] Guy Barles. “Deterministic Impulse Control Problems”. In: *SIAM Journal on Control and Optimization* 23.3 (May 1985). Publisher: Society for Industrial and Applied Mathematics, pp. 419–432.
- [Bec17] Amir Beck. *First-Order Methods in Optimization*. MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, Oct. 2017.
- [BT09] Amir Beck and Marc Teboulle. “A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems”. In: *SIAM Journal on Imaging Sciences* 2.1 (Jan. 2009). Publisher: Society for Industrial and Applied Mathematics, pp. 183–202.
- [BT14] Amir Beck and Marc Teboulle. “A fast dual proximal gradient algorithm for convex minimization and applications”. In: *Operations Research Letters* 42.1 (Jan. 2014), pp. 1–6.
- [BB00] Jean-David Benamou and Yann Brenier. “A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem”. en. In: *Numerische Mathematik* 84.3 (Jan. 2000), pp. 375–393.
- [BC15] Jean-David Benamou and Guillaume Carlier. “Augmented Lagrangian Methods for Transport Optimization, Mean Field Games and Degenerate Elliptic Equations”. en. In: *Journal of Optimization Theory and Applications* 167.1 (Oct. 2015), pp. 1–26.
- [BCS17] Jean-David Benamou, Guillaume Carlier, and Filippo Santambrogio. “Variational Mean Field Games”. In: *Active Particles, Volume 1*. Springer, 2017, pp. 141–171.
- [BFM82] Alain Bensoussan, Jens Frehse, and Umberto Mosco. “A stochastic impulse control problem with quadratic growth Hamiltonian and the corresponding quasi variational inequality.” und. In: *Journal für die reine und angewandte Mathematik* 331 (1982), pp. 124–145.

- [BL75] Alain Bensoussan and Jacques-Louis Lions. “Nouvelles Methodes en Contrôle Impulsionnel”. fr. In: *Applied Mathematics and Optimization* 1.4 (Dec. 1975), pp. 289–312.
- [BL77] Alain Bensoussan and Jacques-Louis Lions. “Optimal impulse and continuous control: Method of nonlinear quasi-variational inequalities”. English. In: *Proceedings of the Steklov Institute of Mathematics* 134 (1977), pp. 7–25.
- [Ber18] Charles Bertucci. “Contributions à la théorie des jeux à champ moyen”. Thèse de doctorat. Paris Sciences et Lettres (ComUE), Dec. 2018.
- [Bil99] Patrick Billingsley. *Convergence of Probability Measures 2e*. Anglais. New York: Wiley, Blackwell, 1999.
- [Bla77] Austin Blaquière. “Differential games with piece-wise continuous trajectories”. en. In: *Differential Games and Applications*. Berlin, Heidelberg: Springer, 1977, pp. 34–69.
- [Bla85] Austin Blaquière. “Impulsive optimal control with finite or infinite time horizon”. en. In: *Journal of Optimization Theory and Applications* 46.4 (Aug. 1985), pp. 431–439.
- [Bre03] Yann Brenier. “Extended Monge-Kantorovich Theory”. en. In: *Optimal Transportation and Applications: Lectures given at the C.I.M.E. Summer School, held in Martina Franca, Italy, September 2-8, 2001*. Lecture Notes in Mathematics. Berlin, Heidelberg: Springer, 2003, pp. 91–121.
- [BKS18] Luis M. Briceño-Arias, Dante Kalise, and Francisco José Silva. “Proximal Methods for Stationary Mean Field Games with Local Couplings”. In: *SIAM Journal on Control and Optimization* 56.2 (Jan. 2018). Publisher: Society for Industrial and Applied Mathematics, pp. 801–836.
- [BJO09] Giuseppe Buttazzo, Chloé Jimenez, and Edouard Oudet. “An Optimization Problem for Mass Transportation with Congested Dynamics”. In: *SIAM Journal on Control and Optimization* 48 (June 2009).
- [CC18] Piermarco Cannarsa and Rossana Capuani. “Existence and Uniqueness for Mean Field Games with State Constraints”. en. In: *PDE Models for Multi-Agent Phenomena*. Cham: Springer International Publishing, 2018, pp. 49–71.
- [Car12] Pierre Cardaliaguet. *Notes on Mean Field Games*. 2012.
- [Car15] Pierre Cardaliaguet. “Weak Solutions for First Order Mean Field Games with Local Coupling”. en. In: *Analysis and Geometry in Control Theory and its Applications*. Springer INdAM Series. Cham: Springer International Publishing, 2015, pp. 111–158.
- [CDLL19] Pierre Cardaliaguet, François Delarue, Jean-Michel Lasry, and Pierre-Louis Lions. *The Master Equation and the Convergence Problem in Mean Field Games: (AMS-201)*. Vol. 2. Princeton University Press, 2019.
- [CG15] Pierre Cardaliaguet and P. Jameson Graber. “Mean field games systems of first order”. en. In: *ESAIM: Control, Optimisation and Calculus of Variations* 21.3 (July 2015). Number: 3 Publisher: EDP Sciences, pp. 690–722.
- [CGPT15] Pierre Cardaliaguet, P. Jameson Graber, Alessio Porretta, and Daniela Tonon. “Second order mean field games with degenerate diffusion and local coupling”. en. In: *Nonlinear Differential Equations and Applications NoDEA* 22.5 (Oct. 2015), pp. 1287–1317.
- [CMS16] Pierre Cardaliaguet, Alpár R. Mészáros, and Filippo Santambrogio. “First Order Mean Field Games with Density Constraints: Pressure Equals Price”. In: *SIAM Journal on Control and Optimization* 54.5 (Jan. 2016). Publisher: Society for Industrial and Applied Mathematics, pp. 2672–2709.
- [Car07] Guillaume Carlier. *Notes de cours: Programmation dynamique*. 2007.
- [CE07] Guillaume Carlier and Ivar Ekeland. “Equilibrium structure of a bidimensional asymmetric city”. In: *Nonlinear Analysis: Real World Applications* 8.3 (July 2007), pp. 725–748.
- [CJS08] Guillaume Carlier, Chloé Jimenez, and Filippo Santambrogio. “Optimal Transportation with Traffic Congestion and Wardrop Equilibria”. In: *SIAM Journal on Control and Optimization* 47.3 (Jan. 2008). Publisher: Society for Industrial and Applied Mathematics, pp. 1330–1350.

- [CD18] René Carmona and François Delarue. *Probabilistic Theory of Mean Field Games with Applications II*. Vol. 84. Probability Theory and Stochastic Modelling. Cham: Springer International Publishing, 2018.
- [CV77] Charles Castaing and Michel Valadier. *Convex Analysis and Measurable Multifunctions*. Vol. 580. Lecture Notes in Mathematics. Berlin, Heidelberg: Springer, 1977.
- [CFL09] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. “Statistical physics of social dynamics”. In: *Reviews of Modern Physics* 81.2 (May 2009). Publisher: American Physical Society, pp. 591–646.
- [CD15] Antonin Chambolle and Charles H. Dossal. “On the convergence of the iterates of "FISTA"”. en. In: *Journal of Optimization Theory and Applications* Volume 166.Issue 3 (Aug. 2015), p. 25.
- [Cia07] Philippe G. Ciarlet. *Introduction à l’analyse numérique matricielle et à l’optimisation*. fr. Dunod, Apr. 2007.
- [CN18] Marco Cirant and Levon Nurbekyan. “The variational structure and time-periodic solutions for mean-field games systems”. In: *Minimax Theory and its Applications* 3 (Apr. 2018), pp. 227–260.
- [Cut13] Marco Cuturi. “Sinkhorn Distances: Lightspeed Computation of Optimal Transportation Distances”. In: *Advances in Neural Information Processing Systems* 26 (June 2013).
- [Dac07] Bernard Dacorogna. *Direct Methods in the Calculus of Variations*. en. Vol. 78. Applied Mathematical Sciences. New York, NY: Springer, 2007.
- [DS23] Annette Dumas and Filippo Santambrogio. “Optimal trajectories in L^1 and under L^1 penalizations”. In: *Comptes Rendus Mathématiques (Accepted Paper)* (2023).
- [DS24] Annette Dumas and Filippo Santambrogio. “Deterministic Mean Field Games with Jumps and Mixed Variational Structure”. In: *(Submitted paper)* (2024).
- [ET99] Ivar Ekeland and Roger Témam. *Convex Analysis and Variational Problems / SIAM Publications Library*. en. Society for Industrial and Applied Mathematics, 1999.
- [EBB10] Naïma El Farouq, Guy Barles, and Pierre Bernhard. “Deterministic Minimax Impulse Control”. en. In: *Applied Mathematics and Optimization* 61.3 (June 2010), pp. 353–378.
- [FR75] Wendell Fleming and Raymond Rishel. *Deterministic and Stochastic Optimal Control*. New York, NY: Springer, 1975.
- [FG83] Michel Fortin and Roland Glowinski. *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*. en. North-Holland: Amsterdam, 1983.
- [FM82] Jens Frehse and Umberto Mosco. “Irregular obstacles and quasi-variational inequalities of stochastic impulse control”. eng. In: *Annali della Scuola Normale Superiore di Pisa - Classe di Scienze* 9.1 (1982). Publisher: Scuola normale superiore, pp. 105–157.
- [GM18] P. Jameson Graber and Alpár R. Mészáros. “Sobolev regularity for first order mean field games”. In: *Annales de l’Institut Henri Poincaré C, Analyse non linéaire* 35.6 (Sept. 2018), pp. 1557–1576.
- [GMST19] P. Jameson Graber, Alpár R. Mészáros, Francisco José Silva, and Daniela Tonon. “The planning problem in mean field games as regularized mass transport”. en. In: *Calculus of Variations and Partial Differential Equations* 58.3 (June 2019), p. 115.
- [Gué09] Olivier Guéant. “Théorie des jeux à champ moyen et applications économiques : second sujet : Taux d’escompte et développement durable”. Thèse de doctorat. Paris 9, 2009.
- [Har20] Godfrey Harold Hardy. “Note on a theorem of Hilbert”. en. In: *Mathematische Zeitschrift* 6.3 (Sept. 1920), pp. 314–317.
- [HLP88] Godfrey Harold Hardy, John Edensor Littlewood, and George Pólya. *Inequalities*. en. Cambridge University Press, Feb. 1988.
- [HMC06] Minyi Huang, Roland P. Malhamé, and Peter E. Caines. “Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence princi-

- ple”. In: *Communications in Information & Systems* 6.3 (Jan. 2006). Publisher: International Press of Boston, pp. 221–252.
- [Hug02] Roger L. Hughes. “A continuum theory for the flow of pedestrians”. In: *Transportation Research Part B: Methodological* 36.6 (July 2002), pp. 507–535.
- [Kak41] Shizuo Kakutani. “A generalization of Brouwer’s fixed point theorem”. In: *Duke Mathematical Journal* 8.3 (Sept. 1941). Publisher: Duke University Press, pp. 457–459.
- [LL06a] Jean-Michel Lasry and Pierre-Louis Lions. “Jeux à champ moyen. I - Le cas stationnaire”. In: *Comptes Rendus Mathématique* 343.9 (Nov. 2006), pp. 619–625.
- [LL06b] Jean-Michel Lasry and Pierre-Louis Lions. “Jeux à champ moyen. II - Horizon fini et contrôle optimal”. In: *Comptes Rendus Mathématique* 343.10 (Nov. 2006), pp. 679–684.
- [LL07] Jean-Michel Lasry and Pierre-Louis Lions. “Mean field games”. en. In: *Japanese Journal of Mathematics* 2.1 (Mar. 2007), pp. 229–260.
- [LS18] Hugo Lavenant and Filippo Santambrogio. “Optimal density evolution with congestion: L^∞ bounds via flow interchange techniques and applications to variational Mean Field Games”. In: *Communications in Partial Differential Equations* 43.12 (Dec. 2018), pp. 1761–1802.
- [LS19] Hugo Lavenant and Filippo Santambrogio. “New estimates on the regularity of the pressure in density-constrained Mean Field Games”. In: *Journal of the London Mathematical Society* 100.2 (Oct. 2019). arXiv:1811.09147 [math], pp. 644–667.
- [Loj88] Stanislaw Lojasiewicz. *An Introduction to the Theory of Real Functions*. en. Wiley, Aug. 1988.
- [MS19] Guilherme Mazanti and Filippo Santambrogio. “Minimal-time mean field games”. In: *Mathematical Models and Methods in Applied Sciences* 29.08 (July 2019). Publisher: World Scientific Publishing Co., pp. 1413–1464.
- [Men81] José-Luis Menaldi. “A singular perturbation result for variational and quasi-variational inequalities”. In: *Nonlinear Analysis: Theory, Methods & Applications* 5.4 (Jan. 1981), pp. 381–400.
- [Men82] José-Luis Menaldi. “Le Problème de contrôle impulsif optimal déterministe et l’inéquation quasi-variationnelle du premier ordre associée”. fr. In: *Applied Mathematics and Optimization* 8.1 (Jan. 1982), pp. 223–243.
- [Mor65] Jean-Jacques Moreau. “Proximité et dualité dans un espace hilbertien”. fr. In: *Bulletin de la Société Mathématique de France* 93 (1965), pp. 273–299.
- [Mor87] Jean-Jacques Moreau. “Bounded Variation in Time”. In: *Topics in Nonsmooth Mechanics*. Birkhäuser Verlag, 1987.
- [Mor89] Jean-Jacques Moreau. “Jump functions of a real interval to a Banach space”. en. In: *Annales de la faculté des sciences de Toulouse Mathématiques* S10 (1989), pp. 77–91.
- [Muñ22] Sebastian Muñoz. “Classical and weak solutions to local first-order mean field games through elliptic regularity”. en. In: *Annales de l’Institut Henri Poincaré C, Analyse non linéaire* 39.1 (Feb. 2022), pp. 1–39.
- [Nas51] John Nash. “Non-Cooperative Games”. In: *Annals of Mathematics* 54.2 (1951). Publisher: Annals of Mathematics, pp. 286–295.
- [Nes83] Yurii Nesterov. “A method of solving a convex programming problem with convergence rate $O(1/k^{*2})$ ”. Russian. In: *Doklady Akademii Nauk SSSR* 269.3 (1983), p. 543.
- [OPS19] Carlo Orrieri, Alessio Porretta, and Giuseppe Savaré. “A variational approach to the mean field planning problem”. In: *Journal of Functional Analysis* 277.6 (Sept. 2019), pp. 1868–1957.
- [PC19] François-Pierre Paty and Marco Cuturi. “Subspace Robust Wasserstein Distances”. en. In: *Proceedings of the 36th International Conference on Machine Learning*. ISSN: 2640-3498. PMLR, May 2019, pp. 5072–5081.

- [Per85] Benoît Perthame. “Some remarks on quasi-variational inequalities and the associated impulsive control problem”. eng. In: *Annales de l’I.H.P. Analyse non linéaire* 2.3 (1985). Publisher: Gauthier-Villars, pp. 237–260.
- [Por23] Alessio Porretta. “Regularizing effects of the entropy functional in optimal transport and planning problems”. In: *Journal of Functional Analysis* 284.3 (Feb. 2023), p. 109759.
- [PS17] Adam Proskinski and Filippo Santambrogio. “Global-in-time regularity via duality for congestion-penalized Mean Field Games”. In: *Stochastics* 89.6-7 (Oct. 2017), pp. 923–942.
- [RZ88] Ryszard Rempala and Jerzy Zabczyk. “On the maximum principle for deterministic impulse control problems”. en. In: *Journal of Optimization Theory and Applications* 59.2 (Nov. 1988), pp. 281–288.
- [ROF92] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. “Nonlinear total variation based noise removal algorithms”. In: *Physica D: Nonlinear Phenomena* 60.1 (Nov. 1992), pp. 259–268.
- [Sad22] Saeed Sadeghi Arjmand. “Mean field games with free final time”. Thèse de doctorat. Institut polytechnique de Paris, Dec. 2022.
- [San15] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians - Calculus of Variations, PDEs and Modeling*. en. Birkhäuser Cham, 2015.
- [San20] Filippo Santambrogio. “Lecture notes on Variational Mean Field Games”. en. In: (2020).
- [San23] Filippo Santambrogio. *A Course in the Calculus of Variations: Optimization, Regularity, and Modeling*. en. Springer Nature, Dec. 2023.
- [SS21] Filippo Santambrogio and Woojoo Shim. “A Cucker–Smale Inspired Deterministic Mean Field Game with Velocity Interactions”. In: *SIAM Journal on Control and Optimization* 59.6 (Jan. 2021). Publisher: Society for Industrial and Applied Mathematics, pp. 4155–4187.
- [Sar22] Clément Sarrazin. “Lagrangian Discretization of Variational Mean Field Games”. In: *SIAM Journal on Control and Optimization* 60.3 (June 2022). Publisher: Society for Industrial and Applied Mathematics, pp. 1365–1392.
- [Sch71] Thomas C. Schelling. “Dynamic models of segregation”. In: *The Journal of Mathematical Sociology* 1.2 (July 1971), pp. 143–186.
- [Uza89] Hirofumi Uzawa. “Iterative methods for concave programming”. In: *Preference, Production and Capital: Selected Papers of Hirofumi Uzawa*. Cambridge: Cambridge University Press, 1989, pp. 135–148.
- [Vil03] Cédric Villani. *Topics in Optimal Transportation*. en. American Mathematical Soc., 2003.
- [War52] John Glen Wardrop. “Some theoretical aspects of road traffic research”. In: (Jan. 1952), pp. 325–378.

Deterministic Mean Field Games with Jumps

Abstract: This thesis focuses on the study of deterministic Mean Field Games with jumps. The motivation for this problem is the modeling of the dynamics of inhabitants in a city where agents move from one place to another. Thus, the strategies of the players are piecewise constant curves on which they optimize a cost that depends on the number of jumps, a term dependent on the density in a variational way, a term dependent on the density in a non-variational way, and a final time cost.

The manuscript begins with an introduction to Mean Field Game theory and presents a new existence result of a fixed point for the mixed problem (a problem with both a variational and a non-variational term), thanks to Kakutani's theorem.

We then provide an Eulerian formulation of the previous problem, which is in Lagrangian form, thanks to a property of optimal transport when the cost is the trivial distance. The existence of a fixed point in the Eulerian formulation can even be achieved using Banach's fixed-point theorem.

We then show that the fixed point obtained as an optimizer of a variational problem is actually a Nash equilibrium, which requires using regularity results of the solution.

Regularity is studied through the Eulerian formulation which reveals an L^1 norm in the density velocity. We obtain a Lipschitz regularity in time with values in $L^2(\Omega)$. Under Dirichlet conditions in time, the solution has the same spatial regularity as the data.

In a numerical section, the solution is approximated using non-smooth optimization algorithms (fast dual proximal gradient method). The last chapter presents two applications of the deterministic mean field games model with jumps in the case of multiple populations.

Keywords: Mean Field Games; Deterministic; Jumps; Calculus of Variations; Regularities; Numerical Simulations; Non-smooth Optimization; Proximal Gradient.

Jeux à champ moyen déterministes avec sauts

Résumé: Cette thèse porte sur l'étude des jeux à champ moyen déterministes avec sauts. La motivation de ce problème est la modélisation de la dynamique des habitants dans une ville où les agents déménagent d'un endroit à un autre. Ainsi, les stratégies des joueurs sont des courbes constantes par morceaux sur lesquelles ils optimisent un coût dépendant du nombre de sauts, d'un terme dépendant de la densité de façon variationnelle, d'un terme dépendant de la densité de façon non-variationnelle et d'un coût en temps final.

Le manuscrit commence par une introduction à la théorie des jeux à champs moyen et présente un résultat nouveau d'existence d'un point fixe pour le problème mixte (problème avec un terme variationnel et un autre non-variationnel), grâce au théorème de Kakutani.

Nous donnons ensuite une formulation Eulérienne du problème précédent qui est sous forme Lagrangienne, grâce à une propriété du transport optimal lorsque le coût est la distance triviale. L'existence d'un point fixe dans la formulation Eulérienne peut même se faire par le théorème de point fixe de Banach.

Nous montrons ensuite que le point fixe obtenu comme un optimiseur d'un problème variationnel est en fait un équilibre de Nash, ce qui demande à utiliser des résultats de régularité de la solution.

La régularité est étudiée grâce à la formulation Eulérienne qui fait apparaître une norme L^1 dans la vitesse de la densité. Nous obtenons une régularité Lipschitz en temps à valeurs dans $L^2(\Omega)$. Sous des conditions de Dirichlet en temps, la solution a la même régularité en espace que les données.

Dans une partie numérique, la solution est approximée en utilisant des algorithmes d'optimisation non-lisses (méthode de gradient proximal dual rapide). Le dernier chapitre est une présentation de deux applications du modèle de jeux à champ moyen déterministes avec sauts dans le cas à plusieurs populations.

Mots clés: Jeux à champ moyen; Déterministe; Sauts; Calcul des variations; Régularités; Simulations numériques; Optimisation non-lisse; Gradient proximal.

Image en couverture : Vue sur Tokyo depuis la Tokyo Skytree.

