A variational formulation of a Multi-Population Mean Field Games with non-local interactions

Luigi De Pascale,* Luca Nenna†

September 11, 2024

We propose a MFG model with quadratic Hamiltonian involving N populations. This results in a system of N Hamilton-Jacobi-Bellman and N Fokker-Planck equations with non-local interactions. As in the classical case we introduce an Eulerian variational formulation which, despite the non convexity of the interaction, still gives a weak solution to the MFG model. The problem can be reformulated in Lagrangian terms and solved numerically by a Sinkhorn-like scheme. We present numerical results based on this approach, these simulations exhibit different behaviours depending on the nature (repulsive or attractive) of the non-local interaction.

Keywords. Mean field games, optimal transport, multi-marginal optimal transport, entropic regularization, convex analysis.
2020 Mathematics Subject Classification. Primary: 49L25 ; Secondary: 49K40, 49N15, 65K10, 91A14.

1. Introduction

The aim of this paper is to establish a variational formulation (both Lagrangian and Eulerian), as well as a suitable numerical methods, for quadratic second order Mean Field Games which involves N different populations interacting through a given non-local functional. Let d be the dimension of the space, then for $i = 1, \dots, N$ we consider

*Dipartimento di Matematica ed Informatica, Universitá di Firenze, Viale Morgagni, 67/a - 50134 Firenze, Italy.
†Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay, ParMA, Inria Saclay, 91405, Orsay, France. email: luca.nenna@universite-paris-saclay.fr

the following system of coupled PDE

$$\boxed{\begin{cases} -\partial_t u^i - \frac{1}{2}\Delta u^i + \frac{1}{2}|\nabla u^i|^2 = \sum_{j\neq i}\int_{\mathbb{R}^d} V^{i,j}(x-y)\rho^j_t \mathrm{d}y, \\ \partial_t \rho^i - \frac{1}{2}\Delta\rho^i + \mathrm{div}(\nabla u^i \rho^i) = 0, \\ \rho^i(0,x) = \rho^i_0(x),\ u^i(T,x) = g^i(x); \end{cases}} \tag{1.1}$$

where $\rho^i(0,x) \in \mathcal{P}_{ac}(\mathbb{R}^d)$ and $g^i \in \mathcal{C}^0(\mathbb{R}^d)$ and $V^{i,j}$ is a given positive and lower semi-continuous potential, the full set of assumptions is at the end of this introduction.
Mean field games involving several populations have been studied recently by [16, 17, 1, 4, 10, 19, 28, 23]; for some more details about the theory of MFG we refer the reader to the seminal work by Lasry and Lions [21], the book [14] or the lecture notes by Cardaliaguet [12]. Notice that the interaction of the populations can be expressed via a more general functional, we will discuss later the extra difficulties, or via some Optimal Transport coupling as done in [4]. We will show that the above system can be seen as the optimality conditions for the following *Eulerian* variational problem

$$\boxed{\inf\left\{\mathcal{J}(\rho^1,v^1;\ldots;\rho^N,v^N)\ \middle|\ \partial_t\rho^i - \frac{1}{2}\Delta\rho^i + \mathrm{div}(\rho^i v^i) = 0,\ \rho^i(0,x) = \rho^i_0, \forall i\right\}}$$

where

$$\mathcal{J}(\rho^1,v^1;\ldots;\rho^N,v^N) := \sum_i \int_{[0,T]\times\mathbb{R}^d} \frac{|v^i|^2}{2}\mathrm{d}\rho^i\mathrm{d}t + \sum_{\substack{1\leq j\leq N\\ j\neq i}} \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} V^{i,j}(x-y)\mathrm{d}\rho^i\otimes\mathrm{d}\rho^j\mathrm{d}t + \int_{\mathbb{R}^d} g^i(x)\mathrm{d}\rho^i(T,x). \tag{1.2}$$

Moreover, we will also relate the above minimization problem with a *Lagrangian* relative entropy minimization problem, that is

$$\boxed{\min\{J(Q^1,\ldots,Q^N)\ :\ e^0_\sharp Q^i = \rho^i_0\}} \tag{1.3}$$

where

$$\begin{aligned} J(Q^1,\ldots,Q^N) &:= \sum_i \mathcal{H}(Q^i|R^i) + \sum_{\substack{1\leq i\leq N\\ j\neq i}} \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} V(x-y)de^t_\sharp Q^i \otimes e^t_\sharp Q^j dt \\ &+ \sum_i \int_{\mathbb{R}^d} g^i(\omega(T))dQ^i, \end{aligned}$$

where $Q^i \in \mathcal{P}(\Omega)$ with $\Omega = \mathcal{C}([0,T];\mathbb{R}^d)$ and and $\mathcal{H}(P \,|\, R)$ is the Boltzmann-Shannon entropy, that is

$$\mathcal{H}(\gamma \,|\, \pi) = \int \rho \log \rho \mathrm{d}\pi, \text{ if } \gamma = \rho\pi.$$

When the reference measure π in the entropy is non indicated, is intended the Lebesgue measure.

We denote the N-uple by $\underline{\rho} := (\rho^1, \dots, \rho^N)$.

Existence and uniqueness of weak solutions for this system will be discussed in two settings, the periodic setting of $\mathbb{T}^d$ and in the full space $\mathbb{R}^d$.

We will make the following assumptions

(A1) $\rho_0^i \in \mathrm{Lip}(\mathbb{T}^d))$ or $\rho_0^i \in \mathrm{Lip}_0(\mathbb{R}^d)$;

(A2) $\mathcal{H}(\rho_0^i) < +\infty$;

(A3) $g^i \in \mathcal{C}^2(\mathbb{R}^d)$,

(A4) $V^{i,j} * \rho \in \mathcal{C}_b(\mathbb{R}^d)$, for all ρ such that $\sqrt{\rho} \in H^1(\mathbb{R}^d)$.

Remark 1.1. Assumption $(A4)$ above is always satisfied if $V^{i,j} \in \mathcal{C}_b(\mathbb{R}^d)$ but also, for example, if $V^{i,j}(x) = \dfrac{1}{|x|^\alpha}$ with α such that $\dfrac{1}{|x|^\alpha} \in L^q_{loc}$ for a $\dfrac{d}{2d-2} \leq q \leq \infty$. So, if $d = 3$, the Coulomb cost is allowed. The condition $\sqrt{\rho} \in H^1(\mathbb{R}^d)$ may look unnatural, however, since ρ is a probability measure, it corresponds to the fact that the Fisher information $I(\rho)$ is finite. Lemma 4.1 of [8] shows that this is always the case in this setting. The relevance of the condition $\sqrt{\rho} \in H^1(\mathbb{R}^d)$ is also given by the fact that such ρ's are the electron densities associated to wave functions in quantum mechanics [22] and, in this respect, the Coulomb type cost discussed above is also relevant (notice that in [21] the authors linked the MFG system with Hartree type equation in which case the Coulomb intercation arises quite naturally).

Also, assumption $(A4)$ is related to Th. 2.6. Choosing to use duality with different functions spaces would, of course, requires different assumptions.

The aim of this work is twofold. On the one hand, we introduce a non-local interaction term, which is not only interesting for the applications (see for instance [16, 17, 1, 4, 10, 19, 28] and the references therein) but also introduces a slight non-convexity. On the other hand, we try to strip naked the structure of the problem so to be as accessible as possible to non-specialists and students.

The paper is organized as follows: Section 2 is devoted to introduce the Eulerian variational formulation (in the same flavor as [21, 13]) as well as the analysis of minimizers and the duality. Notice here that since the interaction term is slightly non-convex we have to introduce and study the linearized functional. In Section 3 we introduce the Lagrangian variational formulation which, as in [7], turns out to correspond to a minimization of a relative entropy. Section 4 is devoted to the time discretization of the two formulations and the $\Gamma-$convergence of the discrete problem to the continuous one. The time discretization, as well as the linearization functional introduced in section 2 are then useful in order to introduce a suitable numerical method in section 5 based on generalized Sinkhorn algorithm.

2. Variational formulation

Let $\mathcal{P}_2 = \left\{ \mu \in \mathcal{P} \mid \mu << \mathrm{Leb}^d, \int |x|^2 \mathrm{d}\mu < +\infty \right\}$ the space of probability measure with finite second moment. We will consider the metric space $(\mathcal{P}_2, \mathcal{W}_2)$ constituted by $\mathcal{P}_2$ equipped with the Wasserstein distance $\mathcal{W}_2$. This is a length metric space (see, for example, [3]) and absolutely continuous curve in this metric space will play a role. Since the functional (1.2) is not convex in the couple (ρ^i, v^i) we have to introduce the momentum variable $m^i = \rho^i v^i$ and re-write the functional in the following convex formulation

$$\mathcal{J}(\rho^1, m^1; \dots; \rho^N, m^N) := \sum_i \int_{[0,T]\times\mathbb{R}^d} \frac{|m^i|^2}{2\rho^i} \mathrm{d}x\mathrm{d}t + \sum_{\substack{1\le j\le N \\ j\neq i}} \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} V^{i,j}(x-y)\mathrm{d}\rho^i \otimes \mathrm{d}\rho^j \mathrm{d}t + \int_{\mathbb{R}^d} g^i(x)\mathrm{d}\rho^i(T,x); \quad (2.1)$$

if $\rho^j \in \mathcal{P}_2$, $m^i << \rho^i$ and $v^i = \frac{\partial m^i}{\partial \rho^i}$, and $+\infty$, otherwise. And $\frac{\partial m^i}{\partial \rho^i}$ denotes the Radon-Nicodym derivative of measures. We will be interested in

$$\inf \left\{ \mathcal{J}(\rho^1, m^1; \dots; \rho^N, m^N) \mid \partial_t \rho^i - \frac{1}{2}\Delta\rho^i + \operatorname{div} m^i = 0, \ \rho^i(0,x) = \rho^i_0, \ \forall i = 1, \cdots, N \right\}. \quad (2.2)$$

Absolutely continuous curves in $\mathcal{P}_2$ are, by now, well characterized. We report here part of the characterization which may be read, for example, in [27] Th. 5.14. For a curve γ in a metric space X we denote by $|\gamma'|$ the metric derivative of γ.

Theorem 2.1. *Let ρ_t be an absolutely continuous curve in $\mathcal{P}_2(\mathbb{R}^d)$ then for a.e. $t \in [0,T]$, there exists a vector field $v_t \in L^2_{\rho_t}$ such that*

$$\partial_t \rho + \operatorname{div}(v_t \rho_t) = 0, \quad (2.3)$$

and for a.e. t we have $\|v_t\|_{L^2_{\rho_t}} \le |\rho'|$, where $|\rho'|$ denotes the metric derivative of ρ. Conversely, if ρ_t is a curve in $\mathcal{P}_2(\mathbb{R}^d)$ and for a.e. $t \in [0,T]$ there exists $v_t \in L^2_{\rho_t}(\mathbb{R}^d, \mathbb{R}^d)$ such that (2.3) holds, then ρ_t is an absolutely continuous curve in $\mathcal{P}_2$ and for a.e. t we have . $|\rho'| \le \|v_t\|_{L^2_{\rho_t}}$.

If $\{\rho_t\}_{t\in[0,T]}$ is a curve of probability measures which solves

$$\partial_t \rho - \frac{1}{2}\Delta\rho + \operatorname{div}(v\rho) = 0,$$

with

$$\int_0^T \int |v_t|^2 \rho_t dt < +\infty$$

and $\rho(0) = \rho_0$, then by Lemma 4.1 of [8] ρ_t is absolutely continuous since for $w_t := v_t - \frac{1}{2}\nabla \log \rho_t$ it holds

$$\partial_t \rho - \operatorname{div}(w_t \rho_t) = 0,$$

and

$$\int_0^T \int |w_t|^2 \rho_t dx dt = \int_0^T \int |v_t|^2 \rho_t dx dt + \frac{1}{4}\int_0^T I(\rho_t) dt + \mathcal{H}(\rho_T) - \mathcal{H}(\rho_0) < +\infty,$$

where I denotes the Fischer information, that is

$$I(\rho) = 4||\nabla\sqrt{\rho}||^2_{L^2}$$

Actually Lemma 4.1 of [8] is richer and more detailed than we need and, for more, we refer the reader to the original paper. In particular, if $(\rho_1, \ldots, \rho_N)$ is admissible for problem (2.2) then each component ρ_i satisfies the assumptions above and so it is an absolutely continuous curve in $\mathcal{P}_2$.

Existence of at least one minimiser of problem (2.2) will be proved in the next section. It is well known that the first term of $\mathcal{J}$ is not convex in the couple (ρ^i, v^i). However, as we have mentioned above, it can be rewritten as

$$\int_{[0,T]\times\mathbb{R}^d} f_2(\rho^i, m^i) \mathrm{d}x\mathrm{d}t,$$

where

$$f_2(t,z) := \begin{cases} \dfrac{z^2}{t} & \text{if } t > 0, \\ 0 & \text{if } t = 0, \text{and}, \ z = 0. \\ +\infty & \text{otherwise.} \end{cases}$$

and the functional $(\rho^i, m^i) \mapsto \int f_2(\rho^i, m^i)$ is convex and lower semi-continuous (see, for example, Th. 5.18 of [27]).

The second term is only separately convex in the N-uple $(\rho^1, \ldots, \rho^N)$ so, obtaining the optimality conditions goes through a directional linearisation process which will also be used in the following section.

Let $(\overline{\rho}^1, \overline{m}^1; \ldots; \overline{\rho}^N, \overline{m}^N)$ be a minimiser of (2.2), define, for $i = 1, \ldots, N$

$$H^i(t,x) = \sum_{j\neq i} \int (V^{i,j}(x-y) + V^{j,i}(x-y)) \mathrm{d}\overline{\rho}^j_t(y), \tag{2.4}$$

and consider the functionals

$$\mathcal{J}^i(\rho, m) := \int_{[0,T]\times\mathbb{R}^d} \frac{|m|^2}{2\rho} \mathrm{d}x\mathrm{d}t + \int_{[0,T]\times\mathbb{R}^d} H^i \mathrm{d}\rho\mathrm{d}t + k^i + \int_{\mathbb{R}^d} g^i(x)\mathrm{d}\rho^i(T,x),$$

where the constant is given by

$$k^i := \sum_{j\neq i} \int_{[0,T]\times\mathbb{R}^d} \frac{|\overline{m}^j|^2}{2\overline{\rho}^j} \mathrm{d}x\mathrm{d}t + \int_{\mathbb{R}^d} g^j(x)\mathrm{d}\overline{\rho}^j(T,x) + \sum_{\substack{1\le k\le N \\ j\neq k\neq i}} \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} V^{j,k}(x-y)\mathrm{d}\rho^k \otimes \mathrm{d}\rho^j \mathrm{d}t.$$

The following proposition is slightly more than a remark.

Proposition 2.2. *For $i = 1, \dots, N$, the couple $(\overline{\rho}^i, \overline{m}^i)$ is a minimizer, among curves starting at ρ_0^i, of the functional*

$$\mathcal{J}^i(\rho, m),$$

which is, moreover, a convex functional.

Proof. We observe that

$$\mathcal{J}^i(\rho, m) = \mathcal{J}(\overline{\rho}^1, \overline{m}^1; \dots; \rho, m; \dots; \overline{\rho}^N, \overline{m}^N) \ge \mathcal{J}(\overline{\rho}^1, \overline{m}^1; \dots; \overline{\rho}^N, \overline{m}^N) = \mathcal{J}^i(\overline{\rho}^i, \overline{m}^i).$$

The convexity of $\mathcal{J}^i$ is due to the separate convexity of the only term of $\mathcal{J}$ which is not convex. □

We will then make a careful analysis of a functional of the type of $\mathcal{J}^i$.

2.1. Analysis of the minimizers

In this subsection, we will write necessary conditions for the minimization of

$$\mathcal{F}(\rho, m) := \int_{[0,T]\times\mathbb{R}^d} \frac{|m|^2}{\rho} \mathrm{d}\rho dt + \int_{[0,T]\times\mathbb{R}^d} H d\rho dt + \int_{\mathbb{R}^d} g(x) d\rho(T,x),$$

among solutions of

$$\begin{cases} \partial_t \rho - \frac{1}{2}\Delta\rho + \operatorname{div} m = 0; \\ \rho(0,x) = \rho_0(x). \end{cases}$$

As frequently in this settings, the convex analysis, via the Fenchel-Rockafellar theorem, will be the main tool. In the next two lemmas we will write $\mathcal{F}$ as the sum of two convex conjugates. Since (ρ, m) is couple of finite measures, the natural space for duality is $\mathcal{C}([0,T]\times\mathbb{R}^d) \times \mathcal{C}([0,T]\times\mathbb{R}^d;\mathbb{R}^d)$.

Lemma 2.3. *Let*

$$\mathcal{K}_1(\alpha,\beta) := \begin{cases} 0 & \textit{if } \alpha + \dfrac{|\beta|^2}{2} \le H, \\ +\infty, & \textit{otherwise.} \end{cases}$$

Then

$$\mathcal{K}_1^*(\rho,m) = \begin{cases} \displaystyle\int \frac{|m|^2}{\rho} + \int H \mathrm{d}\rho & \textit{if } m << \rho, \\ +\infty & \textit{otherwise.} \end{cases}$$

Proof. This is part of Prop. 5.18 in [27]. □

Lemma 2.4. *Let*

$$\mathcal{K}_2(\alpha,\beta) = \begin{cases} \displaystyle\int \phi(0,x)\rho_0 \mathrm{d}x & \textit{if } \alpha = -\partial_t\phi - \Delta\phi \textit{ and } \beta = -\nabla\phi, \\ +\infty & \textit{otherwise;} \end{cases}$$

for some $\phi \in \mathcal{C}_0^2([0,T]\times\mathbb{R}^d)$*, such that* $\phi(T,x) \ge -g(x)$*. Then*

$$\mathcal{K}_2^*(\rho,m) = \begin{cases} \displaystyle\int g(x)\rho(T,x)\mathrm{d}x & \textit{if } \partial_t\rho - \Delta\rho + \operatorname{div} m = 0,\ \rho(0) = \rho_0 \\ +\infty & \textit{otherwise;} \end{cases}$$

Proof. By the finiteness condition on $\mathcal{K}_2$ we can write

$$\mathcal{K}_2^*(\rho,m) = \sup\{\langle -\partial_t\phi - \Delta\phi, \rho\rangle - \langle\nabla\phi, m\rangle - \int \phi(0,x)\rho_0(x)\mathrm{d}x \mid \phi \in \mathcal{C}_0^2([0,T]\times\mathbb{R}^d),\ \phi(T,x) \ge -g(x).\}$$

Integrating by parts gives

$$\sup\{\langle\phi, \partial_t\rho - \Delta\rho + \operatorname{div} m\rangle - \int \phi_T(x)\rho_T(x)\mathrm{d}x + \int \phi_0(x)\rho(0,x)\mathrm{d}x - \int \phi_0(x)\mathrm{d}\rho_0(x)\}$$

considering the sup over suitable subsets of $\{\phi \in \mathcal{C}_0^2([0,T]\times\mathbb{R}^d),\ \phi(T,x) \ge -g(x)\}$ gives that, unless

$$\partial_t\rho - \Delta\rho + \operatorname{div} m = 0$$

in the sense of distibutions, and

$$\rho(0,x) = \rho_0$$

the sup is $+\infty$. If these last two conditions are satisfied then

$$\mathcal{K}_2^*(\rho,\mu) = \int g(x)\rho_T(x)\mathrm{d}x.$$

□

By Fenchel- Rockafellar duality

Proposition 2.5.

$$\min\{\mathcal{K}_1^*(\rho,m)+\mathcal{K}_2^*(\rho,m)\}=\sup\{-\mathcal{K}_1(\alpha,\beta)-\mathcal{K}_2(-\alpha,-\beta)\}$$

According to the expression of the functionals obtained above

$$\min\{\mathcal{K}_1^*(\rho,m)+\mathcal{K}_2^*(\rho,m)\}=\min\Big\{\int\frac{|m|^2}{\rho}+\int H\mathrm{d}\rho+\int g(x)\rho_T(x)\mathrm{d}x \quad \text{if } m<<\rho \text{ and } \partial_t\rho-\Delta\rho+\operatorname{div} m=0,\Big\}$$

which is our original problem. On the other end the right-hand side is

$$\sup\Big\{\int\phi(0,x)\rho_0(x)\mid\phi\in\mathcal{C}_0^2([0,T]\times\mathbb{R}^d),\quad -\partial_t\phi-\Delta\phi+\frac{|\nabla\phi|^2}{2}\leq H,\ \phi(T,x)\leq g(x).\Big\} \tag{2.5}$$

Concerning the existence of a solution for this last problem, we have

Theorem 2.6. *There exists a solution* ψ *of*

$$\begin{cases}-\partial_t\psi-\Delta\psi+\dfrac{|\nabla\psi|^2}{2}=H(t,x),\\ \psi(T,x)=g(x)\end{cases}$$

$$\int\psi(0,x)\rho_0(x)\mathrm{d}x=(2.5).$$

and such ψ *is a maximizer for problem* (2.5) *above.*

Proof. Setting $u(t,x):=\psi(T-t,x)$ the problem above is transformed in

$$\begin{cases}\partial_t u-\Delta u+\dfrac{|\nabla u|^2}{2}=H(T-t,x),\\ u(0,x)=g(x)\end{cases}$$

Then the Hopf-Cole transform $v(x,t)=e^{-\frac{u}{2}}$ gives the further simplification

$$\begin{cases}\partial_t v-\Delta v=-v\dfrac{1}{2}H(T-t,x),\\ v(0,x)=e^{-\frac{g(x)}{2}}\end{cases}$$

So, setting $h(x,t):=-\dfrac{H(T-t,x)}{2}$ we can look at the solutions of the Heat equation. By assumptions $(A1)$, $(A4)$, Remark 1.1 and Lemma 4.1 of [8], h is continuous and bounded. So that, according also to our assumptions on the functions g^i we can apply Theorem 5.1 or 5.2 of [20] (page 320) which gives a $\mathcal{C}^2$ solution of the problem. The maximum principle implies, then, the maximality. □

We conclude the section by connecting everything back to the system (1.1),

Theorem 2.7. *Let $(\rho^1, m^1, ..., \rho^N, m^N)$ be a minimizer for* (2.2)*, define $v^i = m^i/\rho^i$ and, for all $i = 1, \dots, N$, let u^i be a maximizer for* (2.5) *with $H = H^i$ (defined in equation* (2.4)*), $\rho_0 = \rho_0^i$ and $g = g^i$. Then $v^i = -\nabla u^i$ and $(\rho^1, \dots, \rho^N, u^1, \dots, u^N)$ is a solution of* (1.1)*.*

Proof. By Proposition 2.2, (ρ^i, m^i) minimizes, among curves with the same initial datum, the functional

$$\int_{[0,T]\times\mathbb{R}^d} \frac{|m|^2}{2\rho}\mathrm{d}x\mathrm{d}t + \int_{[0,T]\times\mathbb{R}^d} H^i \mathrm{d}\rho\mathrm{d}t + \int_{\mathbb{R}^d} g^i(x)\mathrm{d}\rho^i(T,x),$$

(k^i is irrelevant in the minimization process). Moreover, by Proposition 2.5 and the explicit expressions of the functionals, immediately following that Proposition,

$$\int_{[0,T]\times\mathbb{R}^d} \frac{|m^i|^2}{2\rho^i}\mathrm{d}x\mathrm{d}t + \int_{[0,T]\times\mathbb{R}^d} H^i \mathrm{d}\rho^i\mathrm{d}t + \int_{\mathbb{R}^d} g^i(x)\mathrm{d}\rho^i(T,x) = \int_{\mathbb{R}^d} u^i(0,x)d\rho_0^i(x) \quad (2.6)$$

Using the inequalities satisfied by u^i

$$\int_{[0,T]\times\mathbb{R}^d} H^i \mathrm{d}\rho^i\mathrm{d}t + \int_{\mathbb{R}^d} g^i(x)\mathrm{d}\rho^i(T,x) \geq \int_{[0,T]\times\mathbb{R}^d} (-\partial_t u^i - \Delta u^i + \frac{|\nabla u^i|^2}{2})\mathrm{d}\rho^i\mathrm{d}t + \int_{\mathbb{R}^d} u^i(T,x)\mathrm{d}\rho^i(T,x). \quad (2.7)$$

Integrating by parts, the right-hand side is equal to

$$\int_{[0,T]\times\mathbb{R}^d} u^i(\partial_t\rho^i - \Delta\rho^i)\mathrm{d}x\mathrm{d}t + \int_{[0,T]\times\mathbb{R}^d} \frac{|\nabla u^i|^2}{2}\mathrm{d}\rho^i\mathrm{d}t + \int_{\mathbb{R}^d} u^i(0,x)\mathrm{d}\rho_0^i,$$

which, in turn, is equal, by the equation satisfied by (ρ^i, m^i), to

$$\int_{[0,T]\times\mathbb{R}^d} u^i(-\operatorname{div} m^i)\mathrm{d}x\mathrm{d}t + \int_{[0,T]\times\mathbb{R}^d} \frac{|\nabla u^i|^2}{2}\mathrm{d}\rho^i\mathrm{d}t + \int_{\mathbb{R}^d} u^i(0,x)\mathrm{d}\rho_0^i.$$

After a last integration by parts we substitute in (2.6) to obtain

$$\int_{[0,T]\times\mathbb{R}^d} \frac{|v^i|^2}{2}\mathrm{d}\rho^i\mathrm{d}t + \int_{[0,T]\times\mathbb{R}^d} v^i\cdot\nabla u^i\mathrm{d}\rho^i\mathrm{d}t + \int_{[0,T]\times\mathbb{R}^d} \frac{|\nabla u^i|^2}{2}\mathrm{d}\rho^i\mathrm{d}t \leq 0,$$

which gives the equality $v^i = -\nabla u^i$. Moreover the equality carry on to the inequality (2.7) giving that u^i is a solution of the desired equation. □

3. The Entropic problem

In this section we focus on a Lagrangian formulation to (2.2) based on a minimization of a relative entropy. In particular by proving the existence of a solution to the Lagrangian formulation we deduce the existence also for the Eulerian one.
Let $\Omega := (\mathcal{C}([0,T],\mathbb{R}^d), \|\cdot\|_\infty)$ and let $R^i \in \mathcal{P}(\Omega)$ be Wiener measures defined by

$$R^i = \int_{\mathbb{R}^d} \rho_0^i \delta_{x+B^i(t)(x)} \mathrm{d}x$$

where the B_i are the classical Brownian motions. For $Q^i \in \mathcal{P}(\Omega)$ consider the following variational problem to which we will refer as Lagrangian

$$\min\{J(Q^1,\dots,Q^N) \ : \ e^0_\sharp Q^i = \rho_0^i\} \tag{3.1}$$

where

$$\begin{aligned} J(Q^1,\dots,Q^N) \quad &:= \quad \sum_i \mathcal{H}(Q^i|R^i) + \sum_{\substack{1\le i\le N\\ j\neq i}} \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} V(x-y)\mathrm{d}e^t_\sharp Q^i \otimes \mathrm{d}e^t_\sharp Q^j \mathrm{d}t \\ &+ \quad \sum_i \int_{\mathbb{R}^d} g^i(x)\mathrm{d}e^T_\sharp Q^i, \end{aligned}$$

where $e^t_\sharp : \mathcal{P}(\Omega) \to \mathcal{P}(\mathbb{R}^d)$ is the evaluation map at time t; moreover, notice that for every $t \in [0,T]$, $e^t_\sharp$ is continuous for the narrow convergences.

Theorem 3.1. *The minimum in* (3.1) *is finite and there exists a minimizer* $(\overline{Q}^1,\dots,\overline{Q}^N)$.

Proof. First observe that choosing $Q^i = R^i$ gives at least one point at which J is finite so that the infimum is finite. Let $\{(Q^1_n,\dots,Q^N_n)\}_{n\in\mathbb{N}}$ be a minimizing sequence, so that $J(Q^1_n,\dots,Q^N_n) \le C$. Since V is bounded from below we have

$$\begin{aligned} \sum_i \mathcal{H}(Q^i_n|R^i) + \boldsymbol{k} + \sum_i \int_{\mathbb{R}^d} g^i(\omega(T))\mathrm{d}Q^i_n + \max_\varphi\{\sum_i \int_{\mathbb{R}^d} \varphi(0,x)\mathrm{d}\rho_0^i - \int_{\mathbb{R}^d} \varphi(0,\omega(0))\mathrm{d}Q^i_n\} \\ \le \sum_i \mathcal{H}(Q^i|R^i) + \sum_{\substack{1\le i,j\le N\\ j\neq i}} \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} V(x-y)\mathrm{d}Q^i_t \otimes e^t_\sharp Q^j \mathrm{d}t \\ + \sum_i \int_{\mathbb{R}^d} g^i(\omega(T))\mathrm{d}Q^i + \max_\varphi\{\sum_i \int_{\mathbb{R}^d} \varphi(0,x)\mathrm{d}\rho_0^i - \int_{\mathbb{R}^d} \varphi(0,\omega(0))\mathrm{d}Q^i\} \\ \le C \end{aligned}$$

Which implies, since the other addenda of the first term have linear growth

$$\mathcal{H}(Q^i_n|R^i) \le C_1, \quad \forall i.$$

This implies (see appendix) that, up to subsequences, there exists Q^i with

$$Q^i_n \overset{*}{\rightharpoonup} Q^i$$

and since all the terms constituting J are lower semicontinuous, $(Q^1, \dots, Q^N)$ is a minimizer.

The first term of J is the entropy which is lower-semicontinuous (see Lemma 9.4.3 of [3] and the appendix), the second and the third terms are given by continuous functionals composed with the continuous operator $Q \mapsto e^t_\sharp Q$ finally the last term is the sup of continuous functionals and l.s.c. as such. □

From now on, we denote $Q^i_t = e^t_\sharp Q^i \ \forall t \in [0,T]$.

The next theorem from [8] relates the Lagrangian problem (3.1) to the Eulerian formulation of system (1.1) that we discussed in the previous section. This will allow us to obtain the existence of a minimizer $(\overline{\rho}^1, \overline{m}^1 \dots, \overline{\rho}^N, \overline{m}^N)$ for problem (2.2). Since only the first term of the energy in (2.2) depends on m, given a curve $\rho \in \mathcal{C}([0,T]; \mathcal{P}_2, \mathcal{W}_2)$ we define

$$\mathcal{E}(\rho) := \inf_v \left\{ \frac{1}{2} \int |v_t|^2 \mathrm{d}\rho_t \mathrm{d}t \ : \ \partial_t \rho - \frac{1}{2} \Delta \rho + \mathrm{div}(v\rho) = 0 \right\},$$

and then, we decompose the minimization (2.2) in a two steps minimization writing it as

$$\inf \Bigg\{ \sum_i \mathcal{E}(\rho^i) + \sum_{\substack{1 \le j \le N \\ j \ne i}} \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} V^{i,j}(x-y) \mathrm{d}\rho^i \otimes \mathrm{d}\rho^j \mathrm{d}t$$
$$+ \int_{\mathbb{R}^d} g^i(x) \mathrm{d}\rho^i(T,x) \ : \ \rho^i \in \mathcal{C}([0,T]; \mathcal{P}_2, \mathcal{W}_2), \ \rho^i(0,x) = \rho^i_0, \ \forall i = 1, \cdots, N \Bigg\}. \quad (3.2)$$

This last problem only depends on the curve ρ since the role of m (or v) has already been encoded in $\mathcal{E}$. In an analogous way we can decompose also the minimization problem (3.1) as

$$\inf \Bigg\{ \sum_i \mathcal{S}(\rho^i) + \sum_{\substack{1 \le j \le N \\ j \ne i}} \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} V^{i,j}(x-y) \mathrm{d}\rho^i \otimes \mathrm{d}\rho^j \mathrm{d}t$$
$$+ \int_{\mathbb{R}^d} g^i(x) \mathrm{d}\rho^i(T,x) \ : \ \rho^i \in \mathcal{C}([0,T]; \mathcal{P}_2, \mathcal{W}_2), \ \rho^i(0,x) = \rho^i_0, \ \forall i = 1, \cdots, N \Bigg\}, \quad (3.3)$$

where

$$\mathcal{S}(\mu^i) = \inf\{\mathcal{H}(Q^i | R^i) \mid Q^i_t = \mu^i_t, \ \forall t \in [0,T], \} \quad (3.4)$$

so obtaining a problem which depends only on the curve μ^i. Then using the following result of [8] we have the existence of a minimizer of (2.2) from that of a minimizer of (3.1).

Proposition 3.2 (see corollary 4.7 of [8]). *If $\rho_0 \in \mathcal{P}_2$ and $\mathcal{H}(\rho_0) < +\infty$ then*

$$\mathcal{S}(\rho) = \mathcal{E}(\rho) + \mathcal{H}(\rho_0).$$

Theorem 3.3. *There exists minimizers of problems* (3.2) *and* (2.2).

Proof. If $\overline{Q} = (\overline{Q}^1, \ldots, \overline{Q}^N)$ is a minimizer of problem (3.1), the curve $\overline{Q}_t := (\overline{\rho}^1, \ldots, \overline{\rho}^N)$ is a minimizer of problem (3.3). By Proposition 3.2 above, problems (3.2) and (3.3) have the same minimizers (while the minimal values differs by a constant). So $(\overline{\rho}^1, \ldots, \overline{\rho}^N)$ is also a minimizer for problem (3.2). If we choose, then, for each $i \in \{1, \ldots, N\}$, v^i a vector field as in the definition of $\mathcal{E}(\overline{\rho}^i)$ and consider $\overline{m}^i := v^i \overline{\rho}^i$ we have that $(\overline{\rho}^1, \overline{m}^1, \ldots, \overline{\rho}^N, \overline{m}^N)$ is a minimizer for problem (2.2). □

4. Time discretization and Γ-convergence

Before introducing a suitable discretizations we shortly recall the two equivalent problems we studied above. The main player in those problems is a vector curve of probability measures

$$\rho \in (\mathcal{C}([0,T]; (\mathcal{P}_2, \mathcal{W}_2)))^N.$$

Given a positive integer K the discrete version of the ρ above is a N-tuple of $(K+1)$-vectors of probability measures

$$\boldsymbol{\rho}_K \in \mathcal{P}_2^{K+1} \times \cdots \times \mathcal{P}_2^{K+1}.$$

So the i-th components $\boldsymbol{\rho}_0^i, \cdots, \boldsymbol{\rho}_K^i$ of $\boldsymbol{\rho}_K^i$ is a discrete version of the i-th curve. To this $(K+1)-$tuple of probability measures we can associate the piece-wise constant curve

$$\underline{\rho}^i(t) = \boldsymbol{\rho}_j^i, \quad \text{for } t \in \left[\frac{jT}{K}, \frac{(j+1)T}{K}\right).$$

The ambient space for the Eulerian versions of the problems was

$$\mathcal{A}_0 = \{\rho : [0,T] \to (\mathcal{P}_2)^N \mid \text{abs. cont. and s.t. } \rho(0) = \rho_0\}.$$

We may first introduce the discretized space

$$\mathcal{A}_0^K := \{\boldsymbol{\rho}_K \in (\mathcal{P}_2^{(K+1)})^N \ : \ \boldsymbol{\rho}_0^i = \rho_0^i, \ i = 1, \ldots, N\}$$

We say that $\boldsymbol{\rho}_K \to \rho$ as $K \to +\infty$ if for all $i = 1, \ldots, N$

$$\sup_{t \in [0,T]} \mathcal{W}_2(\underline{\rho}^i(t), \rho^i(t)) \to 0.$$

Given N continuous curves of measures $\rho^i \in \mathcal{C}([0,T], (\mathcal{P}_2(\mathbb{R}^d), \mathcal{W}_2))$, that is $\rho^i : t \in [0,T] \mapsto \rho_t^i \in \mathcal{P}_2(\mathbb{R}^d)$, we defined the minimal energy $\mathcal{E}(\rho)$, the minimal entropic cost

$\mathcal{S}(\rho)$, as well as the cost

$$\mathcal{F}(\rho^1, \cdots, \rho^N) = \sum_{\substack{1 \le i,j \le N \\ j \neq i}} \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} V(x-y) \mathrm{d}\rho_t^i \otimes \mathrm{d}\rho_t^j \mathrm{d}t + \sum_i \int_{\mathbb{R}^d} g^i(x) \mathrm{d}\rho_T^i.$$

Thus, the minimization problem (2.2) and (3.1) can now be re-written in the following way

$$\inf\{\sum_{i=1}^N \mathcal{E}(\rho^i) + \mathcal{F}(\rho^1, \cdots, \rho^N) \mid \rho \in \mathcal{A}_0\}, \tag{4.1}$$

and

$$\inf\{\sum_{i=1}^N \mathcal{S}(\rho^i) + \mathcal{F}(\rho^1, \cdots, \rho^N) \mid \rho \in \mathcal{A}_0\}. \tag{4.2}$$

We define the time discretization of (4.1) as

$$\inf\{\sum_{i=1}^N \mathcal{E}^K(\boldsymbol{\rho}^i) + F^K(\boldsymbol{\rho}^1, \cdots, \boldsymbol{\rho}^N) \mid \boldsymbol{\rho}_K \in \mathcal{T}^K\}, \tag{4.3}$$

where

$$\mathcal{E}^K(\boldsymbol{\rho}^i) := \sum_{k=0}^{K-1} \mathcal{E}_{\frac{T}{K}}(\boldsymbol{\rho}_k^i, \boldsymbol{\rho}_{k+1}^i),$$

with

$$\mathcal{E}_{\frac{T}{K}}(\mu, \nu) := \inf\{\frac{1}{2} \int_0^{\frac{T}{K}} \int_{\mathbb{R}^d} |v_t|^2 d\rho_t dt \mid \partial_t \rho - \frac{1}{2} \Delta \rho + \operatorname{div}(\rho v) = 0,\ \rho_0 = \mu,\ \rho_{\frac{T}{K}} = \nu\}.$$

In a similar way one can discretize in time the Lagrangian counterpart

$$\mathcal{S}^K(\boldsymbol{\rho}^i) := \inf\{\mathcal{H}(Q^i|R^i) \mid Q \in \mathcal{P}(\Omega),\ Q^i_{j\frac{T}{K}} = \boldsymbol{\rho}^i_j,\ j = 0, \cdots, K\} \tag{4.4}$$

as well as the interaction term and the final cost

$$\mathcal{F}^K(\boldsymbol{\rho}^1, \cdots, \boldsymbol{\rho}^N) = \frac{T}{K} \sum_{k=1}^{K-1} \sum_{\substack{1 \le i,j \le N \\ j \neq i}} \int_{\mathbb{R}^d \times \mathbb{R}^d} V(x-y) \mathrm{d}\boldsymbol{\rho}_k^i \otimes \mathrm{d}\boldsymbol{\rho}_k^j + \sum_{i=1}^N \int_{\mathbb{R}^d} g^i(x) \mathrm{d}\boldsymbol{\rho}_K^i, \tag{4.5}$$

where, we recall, $\boldsymbol{\rho}^i$ stands now for a vector of measures, that is $\boldsymbol{\rho}^i \in \mathcal{P}_2(\mathbb{R}^d)^{(K+1)}$ which discretize a curve of measures. Notice that (4.4) can be equivalently reformulate as a classical multi-marginal problem; that is for $i = 1, .., N$ we have

$$\mathcal{S}^K(\boldsymbol{\rho}^i) := \inf\{\mathcal{H}(\pi_K^i|R_K^i) \mid \pi_K^i \in \Pi(\boldsymbol{\rho}_0^i, \cdots, \boldsymbol{\rho}_K^i)\}, \tag{4.6}$$

where $\Pi(\boldsymbol{\rho}_0^i, \cdots, \boldsymbol{\rho}_K^i)$ is the set of probability measures on $(\mathbb{R}^d)^{K+1}$ having $\boldsymbol{\rho}_0^i, \cdots, \boldsymbol{\rho}_K^i$ as marginals and

$$R_K^i := (e^0, e^{\frac{T}{K}}, \cdots, e^T)_\sharp R^i.$$

Then the discretized (4.2) takes the form

$$\inf\{\sum_{i=1}^N \mathcal{S}^K(\boldsymbol{\rho}^i) + \mathcal{F}^K(\boldsymbol{\rho}^1, \cdots, \boldsymbol{\rho}^N) \mid (\boldsymbol{\rho}^1, \cdots, \boldsymbol{\rho}^N) \in \mathcal{T}^K\}, \tag{4.7}$$

where

$$\mathcal{T}^K := \{(\boldsymbol{\rho}^1, \cdots, \boldsymbol{\rho}^N) \in ((\mathcal{P}_2(\mathbb{R}^d), \mathcal{W}_2)^{(K+1)})^N \mid \boldsymbol{\rho}_0^i = \rho_0^i, \ \forall i = 1, \cdots, N\}.$$

Theorem 4.1. *As $K \to +\infty$,*

$$\sum_{i=1}^N \mathcal{E}^K(\boldsymbol{\rho}^i) + \mathcal{F}^K(\boldsymbol{\rho}^1, \ldots, \boldsymbol{\rho}^n) \xrightarrow{\Gamma} \sum_{i}^N \mathcal{E}(\rho^i) + \mathcal{F}(\rho^1, \ldots, \rho^N).$$

Proof. $\Gamma - \limsup$ **inequality** Let $\rho \in \mathcal{A}_0$ be such that

$$\sum_i \mathcal{E}(\rho^i) + \mathcal{F}(\rho) < +\infty.$$

We consider the discretization $\boldsymbol{\rho}_K$ of ρ given by $\boldsymbol{\rho}_j = \rho(\xi_j^K)$ where the times $\xi_j^K \in \left[j\frac{T}{K}, (j+1)\frac{T}{K}\right)$ for $j = 1, \ldots, K-1$ are chosen according to Remark 4.2 below, $\xi_K^0 = 0$ and $\xi_K^K = T$. Since $t \mapsto \rho_t$ is uniformly continuous, the convergence,

$$\boldsymbol{\rho}_K \to \rho,$$

is verified. We check the convergence of each term of the functional, starting with $\mathcal{F}^K$. Since $t \mapsto \rho_t$ is $\mathcal{W}_2$ continuous, it is $w*$ continuous and since $(x, y) \mapsto V(x-y)$ is lower semi-continuous, the same holds for

$$t \mapsto \sum_{\substack{1 \le i,j \le N \\ i \neq j}} \int V(x-y) \mathrm{d}\rho_t^i(x) \otimes \mathrm{d}\rho_t^j(y).$$

And this last map can be considered as g in Remark 4.2. By the choice of discretizing times

$$\int g^i(x) \mathrm{d}\boldsymbol{\rho}_k^i = \int g^i(x) \mathrm{d}\rho_T^i.$$

Concerning the energy term, it is enough to study the convergence for each i. In the addendum $\mathcal{E}_{\frac{T}{K}}(\boldsymbol{\rho}_k^i, \boldsymbol{\rho}_{k+1}^i)$ we may take $\rho\left(t - k\frac{T}{K}\right)$ and the corresponding optimal vector field v as test function so obtaining

$$\mathcal{E}^K(\boldsymbol{\rho}^i_K) = \sum_{k=0}^{K-1} \mathcal{E}_{\frac{T}{K}}(\boldsymbol{\rho}^i_k, \boldsymbol{\rho}^i_{k+1}) \leq \mathcal{E}(\rho^i),$$

which, passing to the $\limsup$ as $K \to +\infty$ concludes the proof of the $\Gamma - \limsup$ inequality.

$\Gamma - \liminf$ **inequality** Let $\rho \in \mathcal{A}_0$ and let $\boldsymbol{\rho}_K \to \rho$. We have

$$\frac{T}{K}\sum_{k=0}^{K-1} \sum_{\substack{1\leq i,j\leq N\\ i\neq j}} \int V(x-y)\mathrm{d}\boldsymbol{\rho}^i_k \otimes \mathrm{d}\boldsymbol{\rho}^j_k = \int_0^T \sum_{\substack{1\leq i,j\leq N\\ i\neq j}} \int V(x-y)\mathrm{d}\underline{\rho}^i_t \otimes \mathrm{d}\underline{\rho}^j_t \mathrm{d}t.$$

The very definition of convergence $\boldsymbol{\rho}_K \to \rho$ implies that $\underline{\rho}^i_t \otimes \underline{\rho}^j_t \stackrel{*}{\rightharpoonup} \mathrm{d}\rho^i_t \otimes \mathrm{d}\rho^j_t$ for all t, i, j. Since V is lower semi-continuous

$$\mu \otimes \nu \mapsto \int V(x-y)\mathrm{d}\mu \otimes \mathrm{d}\nu,$$

is $w*$ lower semi-continuous, by the lower semi-continuity of the integral

$$\liminf_{K\to+\infty} \int_0^T \sum_{\substack{1\leq i,j\leq N\\ i\neq j}} \int V(x-y)\mathrm{d}\underline{\rho}^i_t \otimes \mathrm{d}\underline{\rho}^j_t \mathrm{d}t.$$

Concerning the last addendum of the functional

$$\int g^i(x)\mathrm{d}\boldsymbol{\rho}^i_K \to \int g^i(x)\mathrm{d}\rho^i,$$

since g^i is continuous and bounded by assumptions. So let us look at the energy term $\mathcal{E}^K$. This term may be studied as in [8] by usual methods in Calculus of Variations. Also in this case it is enough to study $\mathcal{E}^K(\boldsymbol{\rho}^i)$. Since we may assume that the total energy of $\boldsymbol{\rho}_K$ is bounded, the same holds for $\mathcal{E}^K(\boldsymbol{\rho}^i)$. Let $\tilde{\rho}_K$, $\tilde{v}_K$ be an almost minimizer for $\mathcal{E}^K(\boldsymbol{\rho}^i)$, i.e. $\tilde{\rho}_K : [0,T] \to \mathcal{P}_2$ is absolutely continuous, $\tilde{\rho}_K\left(j\frac{T}{K}\right) = \boldsymbol{\rho}^i_j$ for $j = 0, \dots, K$ $\tilde{v}_{K,t} \in L^2_{\tilde{\rho}_t}$,

$$\partial_t \tilde{\rho}_K - \frac{1}{2}\Delta\tilde{\rho}_K + \operatorname{div} \tilde{v}_K \tilde{\rho}_K = 0$$

and

$$C \geq \mathcal{E}^K(\boldsymbol{\rho}^i_K) \geq \int_0^T \int \frac{|\tilde{v}_k|^2}{2}\mathrm{d}\tilde{\rho}_{K,t}\mathrm{d}t - \frac{1}{K}.$$

We estimate the total variation of the measure $\tilde{m}_K := \tilde{v}_K \tilde{\rho}_K$ in $[0,T] \times \mathbb{R}^d$. Using

$$C \geq \int_0^T \int \frac{|\tilde{m}_K|^2}{2\tilde{\rho}_K} \mathrm{d}x\mathrm{d}t,$$

and the Young's inequality we obtain

$$\|\tilde{m}_K\|_{[0,T]\times\mathbb{R}^d} \leq C + \|\tilde{\rho}\|_{[0,T]\times\mathbb{R}^d}.$$

Since $\{\tilde{\rho}_K\}$ is also bounded in total variation we have that, up to subsequences

$$(\tilde{\rho}_K, \tilde{m}_K) \overset{*}{\rightharpoonup} (\rho, m),$$

with (ρ, m) solution of

$$\partial_t \rho - \frac{1}{2}\Delta\rho + \operatorname{div} m\rho = 0.$$

Since $m = v\rho$ for a suitable v and applying the lower semi-continuity Theorem 2.34 of [2], originally from [11], we have,

$$\liminf_{K\to+\infty} \mathcal{E}^K(\boldsymbol{\rho}^i_K) \geq \liminf_{K\to+\infty} \int_0^T \int \frac{|\tilde{v}_k|^2}{2} \mathrm{d}\tilde{\rho}_{K,t} dt =$$

$$\liminf_{K\to+\infty} \int_0^T \int \frac{|\tilde{m}_k|^2}{2\tilde{\rho}_{K,t}} \mathrm{d}x\mathrm{d}t \geq \int_0^T \int \frac{|m|^2}{2\rho_t} \mathrm{d}x\mathrm{d}t \geq \mathcal{E}(\rho^i).$$

□

Remark 4.2. Let $g : [0,T] \to [0,+\infty)$ be a lower-semicontinuous functions and let $K \in \mathbb{N}$. There exists points $\xi^j_K \in \left[j\frac{T}{K}, (j+1)\frac{T}{K}\right)$ such that

$$\frac{T}{K}\sum_{j=1}^{K-1} g(\xi^J_K) \to \int_0^T g(t)\mathrm{d}t.$$

5. Numerical Approximation

We now present a numerical scheme in order to solve the discretized in time problem (4.7). In particular the scheme is based on a variant of the Sinkhorn algorithm, successfully used to solve many variational problems involving optimal transport [18, 7, 5, 9, 25, 26, 15, 24] and it is an adaptation of the scheme introduced in [8, 4].

We recall that (4.7) reads as:

$$\inf\left\{\sum_{i=1}^N \mathcal{S}^K(\boldsymbol{\rho}^i) + \mathcal{F}^K(\boldsymbol{\rho}^i, \cdots, \boldsymbol{\rho}^i) \mid (\boldsymbol{\rho}^i, \cdots, \boldsymbol{\rho}^i) \in \mathcal{T}^K\right\}$$

where $\mathcal{S}^K$ is itself defined by (4.6) which is an entropy minimization with multi-marginal constraints.

Denoting $P^k : (\mathbb{R}^d)^{K+1} \to (\mathbb{R}^d)$ the $k-$th canonical projection we can obviously rewrite (4.7) as an optimization problem over plans π^i_N only:

$$\inf \Bigg\{ \sum_{i=1}^N \mathcal{H}(\pi^i_K|R^i_K) + i_{\rho^i_0}(P^0_\# \pi^i_k) + G(P^K_\# \pi^i_K) \\ + \mathcal{F}^K(\boldsymbol{P_\# \pi^1_K}, \cdots, \boldsymbol{P_\# \pi^N_K}) \ : \ (\pi^i_K)^N_{i=1} \in (\mathcal{P}((\mathbb{R}^d)^{K+1}))^N \Bigg\}, \quad (5.1)$$

where $\boldsymbol{P_\# \pi^i_K} = (P^0_\# \pi^i_K, \cdots, P^K_\# \pi^i_K)$ and

$$i_{\rho_0}(\rho) = \begin{cases} 0 \text{ if } \rho = \rho_0 \\ +\infty \text{ otherwise} \end{cases}$$

is the indicator function in the convex analysis sense and is used to enforce the initial condition. We recall that for all $i = 1, \cdots, N$, the static reference measure R^i_K in the relative entropy term is defined as follows

$$R^i_K := (e^0, e^{\frac{T}{K}}, \cdots, e^T)_\sharp R^i.$$

Moreover, since we are considering the reversible Wiener measure, it turns out that R^i_K can be decomposed by using the heat kernel as

$$R^i_K := \left(\prod_{k=1}^K H_{\frac{T}{K}}(x_k - x_{k-1}) \right) dx_0, \cdots dx_K,$$

where

$$H_t(z) := \frac{1}{(2\pi t)^d} \exp\left(- \frac{|z|^2}{2t} \right), \ t > 0, \ z \in \mathbb{R}^d,$$

and $|\cdot|$ denotes the standard euclidean norm. We also need a discretization in space, for instance we use a M grid points to discretize $\mathbb{R}^d$, then π^i_K and R^i_K become tensors in $\mathbb{R}^{MN}$. For sake of simplicity we will keep the continuous in space notation, but form now on integral must be understood as finite sums and $x_0, \cdots, x_K$ as M vectors. Notice that thanks to the euclidean norm, the heat kernel $H_t(z)$ can be decomposed as a product along the dimension of the one dimensional kernel, that is

$$H_t(z) = \prod_{j=1}^d h_t(z_j),$$

where $h_t(z_j)$ is the heat kernel in dimension one. This implies that, instead of storing a matrix $H_t \in \mathbb{R}^{M \times M}$, one can just store d small matrices belonging to $\mathbb{R}^{\sqrt[d]{M} \times \sqrt[d]{M}}$. One can now try to generalize the algorithm introduced in [15] and its multi-marginal variant [7] to the multi-population case in the same flavour as [4]. However, since the interaction term between populations is non-convex, it happens that we are out of the domain of application of Sinkhorn algorithm. A way to overcome this difficulty is through a semi-implicit approach in order to treat the interaction term, that is at step $n+1$ the $i-$th plan $\pi_K^{i,(n+1)}$ is computed as the optimal solution of a linearized problem obtained by injecting the $j-$th, with $j \neq i$, plans $\pi_K^{j,(n)}$ computed at the previous step: for all $i = 1, \cdots, N$

$$\pi_K^{i,(n+1)} := \operatorname{argmin}_{\pi_K^i \in \mathcal{P}((\mathbb{R}^d)^{K+1})} \Big\{ \mathcal{H}(\pi_K^i | R_K^i) + i_{\rho_0^i}(P_\#^0 \pi_k^i) + G(P_\#^K \pi_K^i) + \mathcal{F}_i^K(\boldsymbol{P_\# \pi_K^i}) \Big\}, \quad (5.2)$$

where

$$\mathcal{F}_i^K(\boldsymbol{P_\# \pi_K^i}) := \mathcal{F}^K(\boldsymbol{P_\# \pi_K^{1,(n+1)}}, \cdots, \boldsymbol{P_\# \pi_K^i}, \cdots, \boldsymbol{P_\# \pi_K^{N,(n)}})$$

We now have to solve N finite-dimensional strictly convex minimization problems. Then, for each problem strong duality holds and (5.2) cab be re-written as follows

$$\sup_{(u_0^i, \cdots, u_K^i)} -F_0^*(-u_0^i) - G^*(-u_K^i) - \mathcal{F}_i^*(-u_1^i, \cdots, -u_{K-1}^i) - \int \Big(\exp(\oplus_{j=0}^K u_j^i) - 1 \Big) R_K^i, \quad (5.3)$$

where with a slight abuse of notation F_i^* denotes the sum of the Fenchel-Legendre transform of each term in $\mathcal{F}_i^K$. Denoting by $\pi_K^{i,(n+1)}$ and $u_j^{i,(n+1)}$ the optimal solution to (5.2) and (5.3), respectively, it follows that the unique solution to (5.2) has the form

$$\pi_K^{i,(n+1)}(\boldsymbol{x}) := \Big(\otimes_{k=0}^K e^{u_k^{i,(n+1)}(x_k)} \Big) R_K^i(\boldsymbol{x}),$$

where $\boldsymbol{x} = (x_0, \cdots, x_K)$.

Remark 5.1 (Structure of the optimal solution). By definition of the linearized term $\mathcal{F}_i^K$ it follows that the $\mathcal{F}_i^*$ is just a sum of indicator function in the convex analysis sense; this implies that for all $k = 1, \cdots, K-1$

$$u_k^{i,(n+1)}(x_k) = \sum_{j \neq i} \int V^{i,j}(x_k - y_k) \rho_k^{j,(n)}(y_k) \mathrm{d}y_k,$$

where $\rho_k^{j,(n)}(y_k) := P_{k,\sharp} \pi_K^{j,(n)}$ is the marginal of the solution computed at the previous step. In the same way if the final cost if of the form $G(\rho) = \int g \mathrm{d}\rho$ then $u_K^{i,\star}(x_K) = g^i(x_K)$. For sake of clarity we consider always these kinds of functional, even if the algorithm can be defined with more complex functional (we refer the reader to [7, 4] for some examples with 1 or 2 populations; the extension to N populations is then straightforward).

Notice that thanks to the remark above the generalised Sinkhorn algorithm takes now the following form.

Algorithm 1 Multi-population Sinkhorn

Require: $u_k^{i,(0)} = 0$

while $\sum_{i=1}^{N} ||\rho_0^{i,(n)} - \rho_0^i|| <$tol **do**

  for $i = 1 : N$ **do**

    for $k = 0 : K$ **do**

      if $k = 0$ **then**

        $u_0^{i,(n+1)} = \log(\rho_0^i) - \log\left(\int \left(\exp\left(\oplus_{j=1}^{K} u_j^{i,(n)}\right)\right) R_K^i\right)$

      else if $k \neq 0, K$ **then**

        $u_K^{i,(n+1)} = \sum_{j \neq i} \int V^{i,j}(x_k - y_k)\rho_k^{j,(n)}(y_k)dy_k$

      else if $k = K$ **then**

        $u_K^{i,(n+1)} = g^i$

      end if

    end for

  end for

end while

In the following numerical results we take a space $M \times M$ discretization of $[0,1]^2$ with $M = 100$ and a time discretization of $[0,1]$ with $K = 32$ time step. Let us, firstly, consider the 2 densities case: we have always considered the same initial data

$$\rho_0^1 = \exp(-50(x^1 - .2)^2 - 50(y^1 - .5)^2), \quad \rho_0^2 = \exp(-50(x^2 - .8)^2 - 50(y^2 - .5)^2)$$

and the same final costs

$$g^1 = 50((x^1 - 0.8)^2 + (y^1 - 0.45)^2), \quad g^2 = 50((x^2 - 0.2)^2 + (y^2 - 0.5)^2),$$

such that we expect the two Gaussians to switch position. As for the interaction potential we have considered in Figure 5.1 a strong repulsion given by $V(x,y) = 120\chi_{||x-y||<0.2}(x,y)$ and in Figure 5.2 a truncated Coulomb repulsion $V(x,y) = \min\left(1000, \frac{1}{||x-y||}\right)$. Notice in both cases the effect of entropic term which oblige the densities to spread but at the same time the effect of the repulsive interaction forbid them to touch each other (the distance between them depends on the kind of the repulsion term).

It is now straightforward to extend the theory and the numerical method to a slightly more general model with a viscosity parameter ε. The Mean Field Game system (1.1)

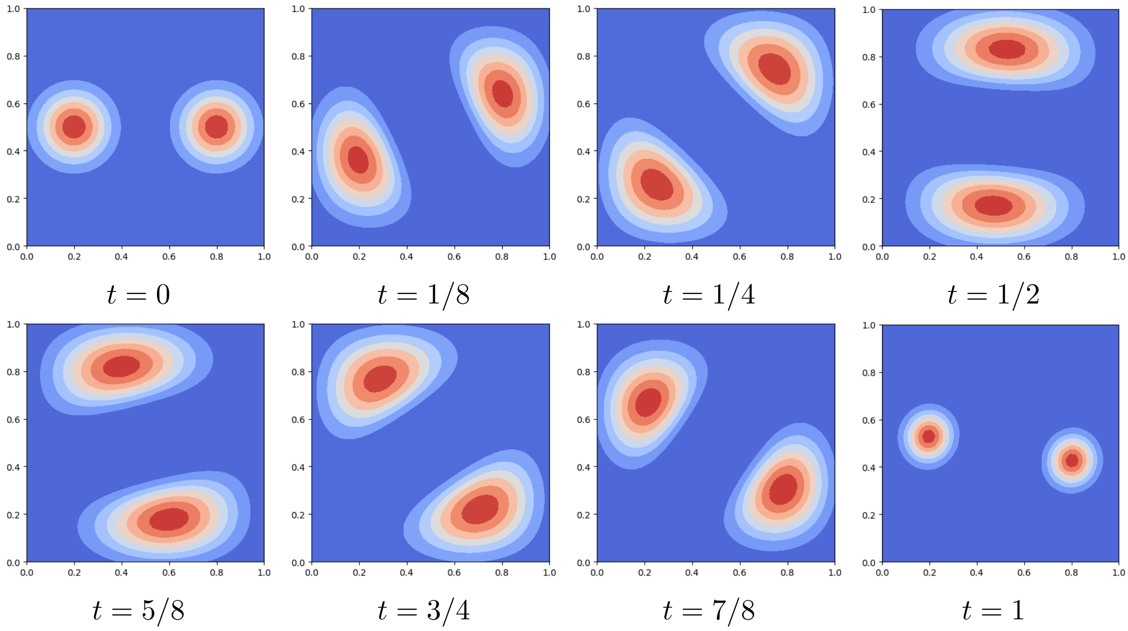

Figure 5.1: Support of ρ^1 and ρ^2 for $V(x,y) = 120\chi_{||x-y||<0.2}(x,y)$.

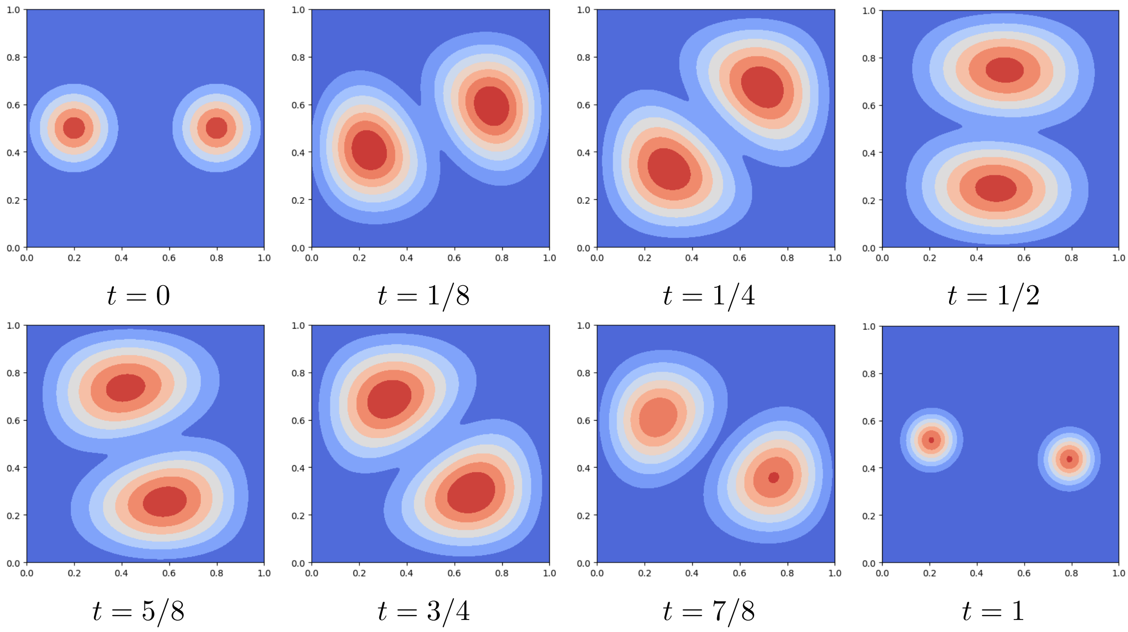

Figure 5.2: Support of ρ^1 and ρ^2 for $V(x,y) = \min\left(1000, \dfrac{1}{||x-y||}\right)$.

now takes the form :

$$
\begin{cases}
-\partial_t u^i - \dfrac{\varepsilon}{2}\Delta u^i + \dfrac{1}{2}|\nabla u^i|^2 = \displaystyle\sum_{j\neq i}\int_{\mathbb{R}^d} V^{i,j}(x-y)\rho_t^j \mathrm{d}y, \\
\partial_t \rho^i - \dfrac{\varepsilon}{2}\Delta \rho^i + \mathrm{div}(\nabla u^i \rho^i) = 0, \\
\rho^i(0,x) = \rho_0^i(x),\ u^i(T,x) = g^i(x);
\end{cases}
\tag{5.4}
$$

The Lagrangian formulation (3.1) we have proposed becomes

$$
\min\{J_\varepsilon(Q^1,\dots,Q^N)\ :\ e_\sharp^0 Q^i = \rho_0^i\} \tag{5.5}
$$

where

$$
\begin{aligned}
J_\varepsilon(Q^1,\dots,Q^N) \quad &:= \quad \sum_i \mathcal{H}(Q^i|R_\varepsilon^i) + \sum_{\substack{1\leq i\leq N\\ j\neq i}} \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} V(x-y)\mathrm{d}e_\sharp^t Q^i \otimes \mathrm{d}e_\sharp^t Q^j \mathrm{d}t \\
&+ \quad \sum_i \int_{\mathbb{R}^d} g^i(x)\mathrm{d}e_\sharp^T Q^i.
\end{aligned}
$$

where R_ε^i are the reversible Wiener measure induced by a Brownian motion with variance ε. Notice that one could choose different ε for each population. In particular, if we discretize the problem in time, we have that the reference measure $R_{\varepsilon,K}^i$ can be still decomposed by using the heat kernel $H_{\varepsilon t}(z)$.

Notice that we can still use the algorithm we have introduced above, but the performance, in terms of iterations to converge, will be affected by small values of ε. At least formally, when the viscosity is small, (5.5) is an approximation of the following Lagrangian formulation of first-order variational mean-field games (see [6] for the one population case)

$$
\min\{\mathcal{K}(Q^1,\dots,Q^N)\ :\ e_\sharp^0 Q^i = \rho_0^i\} \tag{5.6}
$$

where

$$
\begin{aligned}
\mathcal{K}(Q^1,\dots,Q^N) \quad &:= \quad \sum_i K(Q_i) + \sum_{\substack{1\leq i\leq N\\ j\neq i}} \int_0^T \int_{\mathbb{R}^d\times\mathbb{R}^d} V(x-y)\mathrm{d}e_\sharp^t Q^i \otimes \mathrm{d}e_\sharp^t Q^j \mathrm{d}t \\
&+ \quad \sum_i \int_{\mathbb{R}^d} g^i(x)\mathrm{d}e_\sharp^T Q^i.
\end{aligned}
$$

where

$$
K(Q) := \frac{1}{2}\int_\Omega \int_0^T |\dot{\omega}(t)|^2 \mathrm{d}t\mathrm{d}Q(\omega). \tag{5.7}
$$

This also implies that we can use the Sinkhorn algorithm, with small ε, in order to approximate the solution to first-order MFGs. In Figures 5.3 and 5.4 we have considered the same data as above but with $\varepsilon = 0.005$; notice now the effect of a weaker diffusion term which prevents the densities from spreading. Finally, we consider a 3 populations

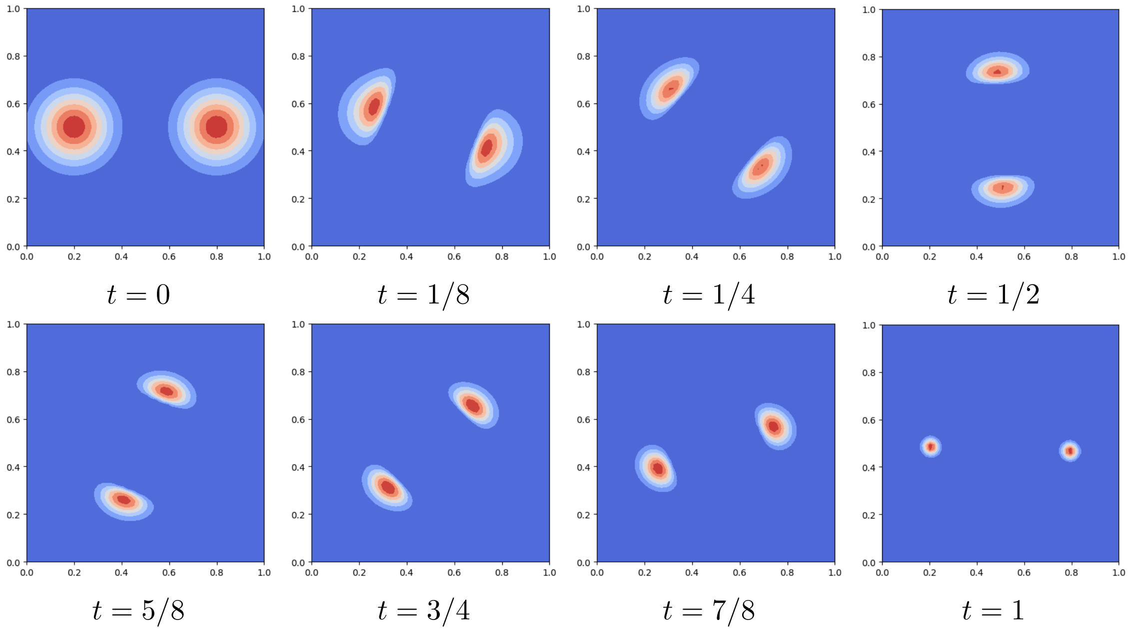

Figure 5.3: Support of ρ^1 and ρ^2 for $\varepsilon = .005$ and $V(x, y) = 120\chi_{||x-y||<0.2}(x, y)$.

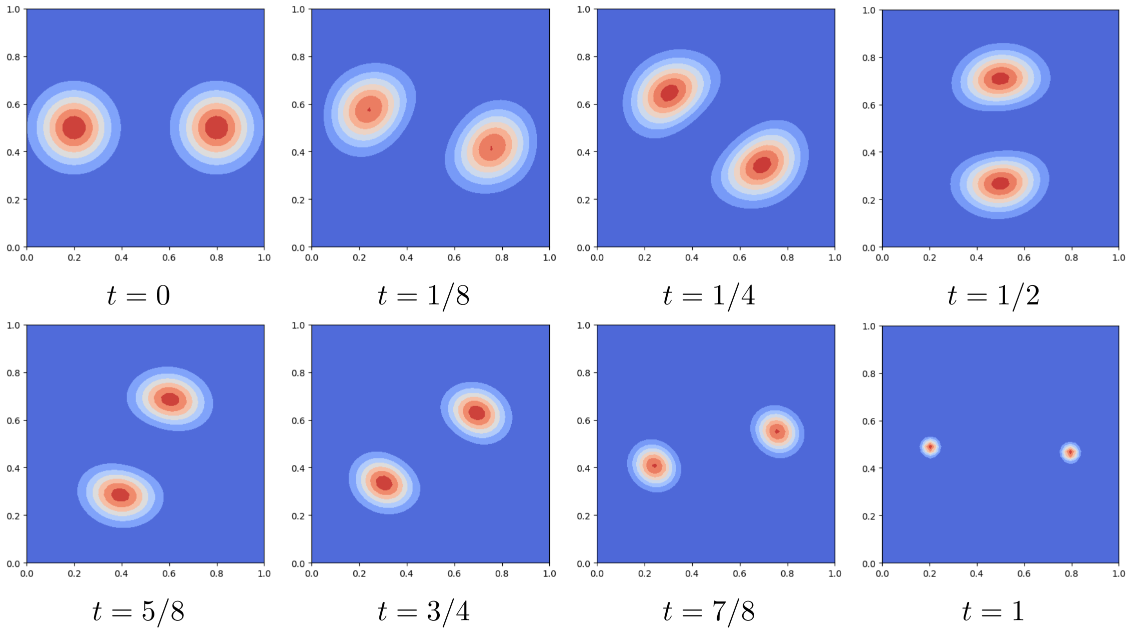

Figure 5.4: Support of ρ^1 and ρ^2 for $\varepsilon = .005$ and $V(x, y) = \min\left(1000, \dfrac{1}{||x-y||}\right)$.

example with initial data

$$\rho_0^1 = \exp(-50(x^1 - .2)^2 - 50(y^1 - .5)^2),\ \rho_0^2 = \exp(-50(x^2 - .8)^2 - 50(y^2 - .5)^2),$$

$$\rho_0^3 = \exp(-80(x^3 - .5)^2 - 80(y^3 - .1)^2),$$

and final costs

$$g^1 = 50((x^1-0.8)^2+(y^1-0.5)^2),\ g^2 = 50((x^2-0.5)^2+(y^2-0.1)^2),\ g^3 = 50((x^2-0.2)^2+(y^2-0.5)^2)$$

which induces a rotation of the populations. In Figure 5.5 we plot the evolution of support of the 3 densities considering as interaction term the strong repulsion we have taken above.

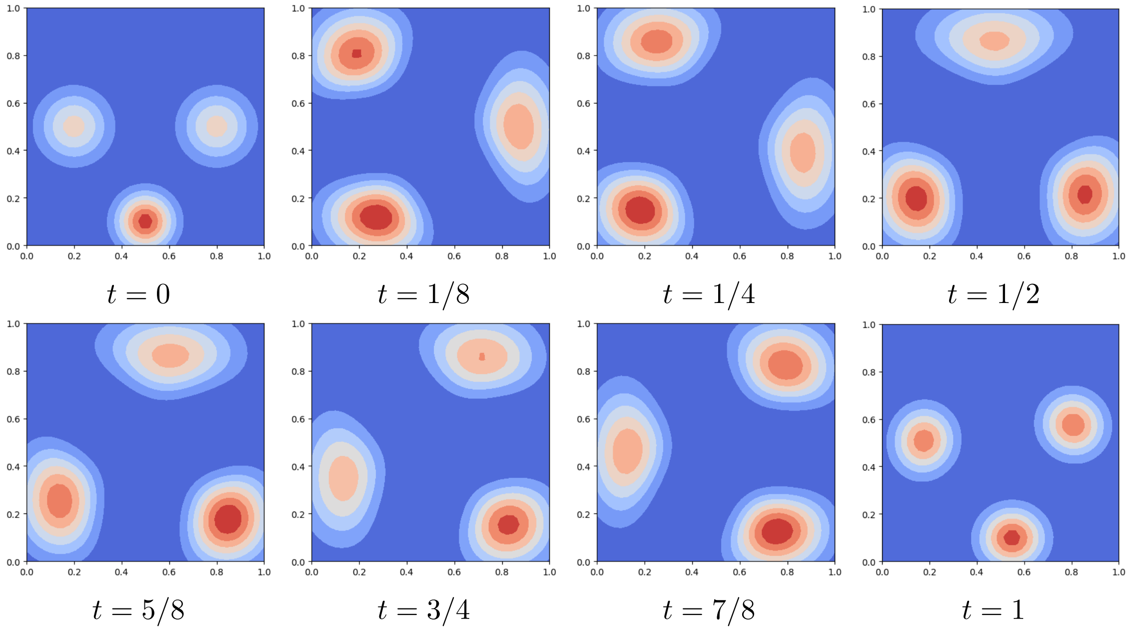

Figure 5.5: Support of ρ^1, ρ^2 and ρ^3 for $V(x, y) = 120\chi_{||x-y||<0.2}(x, y)$.

A. Entropy and the De la Vallée Poussin Theorem

The existence of minimizers for J relies on the compactness of minimizing sequences that follows from the super-linearity of the entropy functional. This is a classical fact which we shortly report here for the reader's convenience.

Let μ and σ be two probability measures on a metric space X. We say that μ is absolutely continuous with respect to σ if there exists a function $f \in L^1_\sigma$ such that

$$\mu = f(x)\sigma,$$

and in this case we write $\mu << \sigma$ and we use the classical notation $f = \dfrac{\mathrm{d}\mu}{\mathrm{d}\sigma}$.

The relative entropy of μ with respect to σ is defined as

$$\mathcal{H}(\mu|\sigma) = \begin{cases} \displaystyle\int_X \frac{\mathrm{d}\mu}{\mathrm{d}\sigma} \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\sigma}\right)\mathrm{d}\sigma & \text{if } \mu << \sigma, \\ +\infty & \text{otherwise.} \end{cases}$$

Following [3] (Example 9.3.6) one can introduce the function

$$H(s) = \begin{cases} s(\log s - 1) + 1 & \text{if } s > 0, \\ 1 & \text{if } s = 0, \\ +\infty & \text{if } s < 0, \end{cases}$$

which is nonnegative, lower semi-continuous, strictly convex and super-linear at $+\infty$. Then it holds

$$\mathcal{H}(\mu|\sigma) = \int H\left(\frac{\mathrm{d}\mu}{\mathrm{d}\sigma}\right)\mathrm{d}\sigma;$$

and

$$\mathcal{H}(\mu|\sigma) = 0 \leftrightarrow \mu = \sigma.$$

This way to rewrite the relative entropy is handy to make a connection with the De la Vallée Poussin theorem. In fact, let $\{\mu_i\}_{i\in I} \subset \mathcal{P}(X)$ be such that $\mathcal{H}(\mu_i|\sigma) < C$ then the family $\left\{\frac{\mathrm{d}\mu_i}{\mathrm{d}\sigma}\right\}_{i\in I} \subset L^1_\sigma(X)$ is weakly compact thanks to the the theorem

Theorem A.1 (De la Vallée Poussin)**.** *Let $\{f_i\}_{i\in I} \subset L^1_\sigma$ then the following are equivalent*

- *the functions $\{f_i\}_{i\in I}$ are uniformly integrable (and then weakly compact in L^1_σ by the Dunford-Pettis theorem,*
- *there exists a function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ non-decreasing and such that*

$$\lim_{t\to+\infty} \frac{\varphi(t)}{t} = +\infty,$$

such that $\int_X \varphi(|f_i|)d\sigma < C$.

A second advantage of writing the entropy using the function H is the lower semi-continuity with respect to the weak L^1_σ-convergence of the densities, in the sense that if

$$f_n \overset{w-L^1_\sigma}{\rightharpoonup} f,$$

then

$$\liminf_{n\to\infty} \int_X H(f_n)d\sigma \geq \int_X H(f)d\sigma.$$

Acknowledgment:

The first author acknowledges the support of GNAMPA-INDAM, of the PRIN (Progetto di ricerca di rilevante interesse nazionale) 2022J4FYNJ, *Variational methods for stationary and evolution problems with singularities and interfaces* and of *"Fondi di ricerca di ateneo, ex 60 %"* of the University of Firenze. The second author is partially on academic leave at Inria (team Matherials) for the year 2023-2024 and acknowledges the hospitality of this institution during this period. His work benefited from the support of the FMJH Program PGMO, from H-Code, Université Paris-Saclay and from the ANR project GOTA (ANR-23-CE46-0001).

Both author warmly thanks the hospitality of *equipe MoKaPlan of the INRIA Paris*, where this paper has been written.

References

[1] Yves Achdou, Pierre Cardaliaguet, François Delarue, Alessio Porretta, Filippo Santambrogio, Pierre Cardaliaguet, and Alessio Porretta. An introduction to mean field game theory. *Mean Field Games: Cetraro, Italy 2019*, pages 1–158, 2020.

[2] L Ambrosio, N Fusco, and D Pallara. Functions of bounded variation and free continuity problems, 2000.

[3] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Lectures in Mathematics. Birkhäuser Boston, 2nd edition, 2004.

[4] César Barilla, Guillaume Carlier, and Jean-Michel Lasry. A mean field game model for the evolution of cities. *Journal of Dynamics & Games*, 8(3):299, 2021.

[5] J.-D. Benamou, G. Carlier, and L. Nenna. A numerical method to solve multi-marginal optimal transport problems with coulomb cost. In *Splitting Methods in Communication, Imaging, Science, and Engineering*, pages 577–601. Springer International Publishing, 2016.

[6] J.-D. Benamou, G. Carlier, and F. Santambrogio. Variational mean field games. In *Active Particles, Volume 1*, pages 141–171. Springer, 2017.

[7] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.*, 37(2):A1111–A1138, 2015.

[8] Jean-David Benamou, Guillaume Carlier, Simone Di Marino, and Luca Nenna. An entropy minimization approach to second-order variational mean-field games. *Mathematical Models and Methods in Applied Sciences*, 29(08):1553–1583, 2019.

[9] Jean-David Benamou, Guillaume Carlier, and Luca Nenna. Generalized incompressible flows, multi-marginal transport and sinkhorn algorithm. *to appear on Numerische Mathematik*, 2018.

[10] Alain Bensoussan, Tao Huang, and Mathieu Lauriere. Mean field control and mean field game models with several populations. *arXiv preprint arXiv:1810.00783*, 2018.

[11] Giuseppe Buttazzo and Lorenzo Freddi. Functionals defined on measures and applications to non equi-uniformly elliptic problems. *Annali di matematica pura ed applicata*, 159(1):133–149, 1991.

[12] Pierre Cardaliaguet. A short course on mean field games. *Unpublished lecture notes*, 2018.

[13] Pierre Cardaliaguet, P Jameson Graber, Alessio Porretta, and Daniela Tonon. Second order mean field games with degenerate diffusion and local coupling. *Nonlinear Differential Equations and Applications NoDEA*, 22(5):1287–1317, 2015.

[14] René Carmona, François Delarue, et al. *Probabilistic theory of mean field games with applications I-II.* Springer, 2018.

[15] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. Scaling algorithms for unbalanced transport problems. Technical report, http://arxiv.org/abs/1607.05816, 2016.

[16] Marco Cirant. Multi-population mean field games systems with neumann boundary conditions. *Journal de Mathématiques Pures et Appliquées*, 103(5):1294–1315, 2015.

[17] Marco Cirant and Gianmaria Verzini. Bifurcation and segregation in quadratic two-populations mean field games systems. *ESAIM: Control, Optimisation and Calculus of Variations*, 23(3):1145–1177, 2017.

[18] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems (NIPS) 26*, pages 2292–2300, 2013.

[19] Gokce Dayanikli and Mathieu Lauriere. Multi-population mean field games with multiple major players: Application to carbon emission regulations. *arXiv preprint arXiv:2309.16477*, 2023.

[20] Olga Aleksandrovna Ladyzhenskaia, Vsevolod Alekseevich Solonnikov, and Nina N Ural'tseva. *Linear and quasi-linear equations of parabolic type*, volume 23. American Mathematical Soc., 1988.

[21] Jean-Michel Lasry and Pierre-Louis Lions. Mean field games. *Japanese journal of mathematics*, 2(1):229–260, 2007.

[22] Elliott H Lieb. Density functionals for coulomb systems. *International Journal of Quantum Chemistry*, 24(3):243–277, 1983.

[23] Alpár Richárd Mészáros and Francisco J Silva. On the variational formulation of some stationary second-order mean field games systems. *SIAM Journal on Mathematical Analysis*, 50(1):1255–1277, 2018.

[24] Luca Nenna. *Numerical methods for multi-marginal optimal transportation.* PhD thesis, PSL Research University, 2016.

[25] Gabriel Peyré. Entropic approximation of Wasserstein gradient flows. *SIAM J. Imaging Sci.*, 8(4):2323–2351, 2015.

[26] Gabriel Peyré and Marco Cuturi. Computational optimal transport. Technical report, 2017.

[27] Filippo Santambrogio. *Optimal transport for applied mathematicians.* Progress in Nonlinear Differential Equations and their Applications, 87. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.

[28] Li-Hsien Sun. Mean field games with heterogeneous groups: Application to banking systems. *Journal of Optimization Theory and Applications*, 192(1):130–167, 2022.